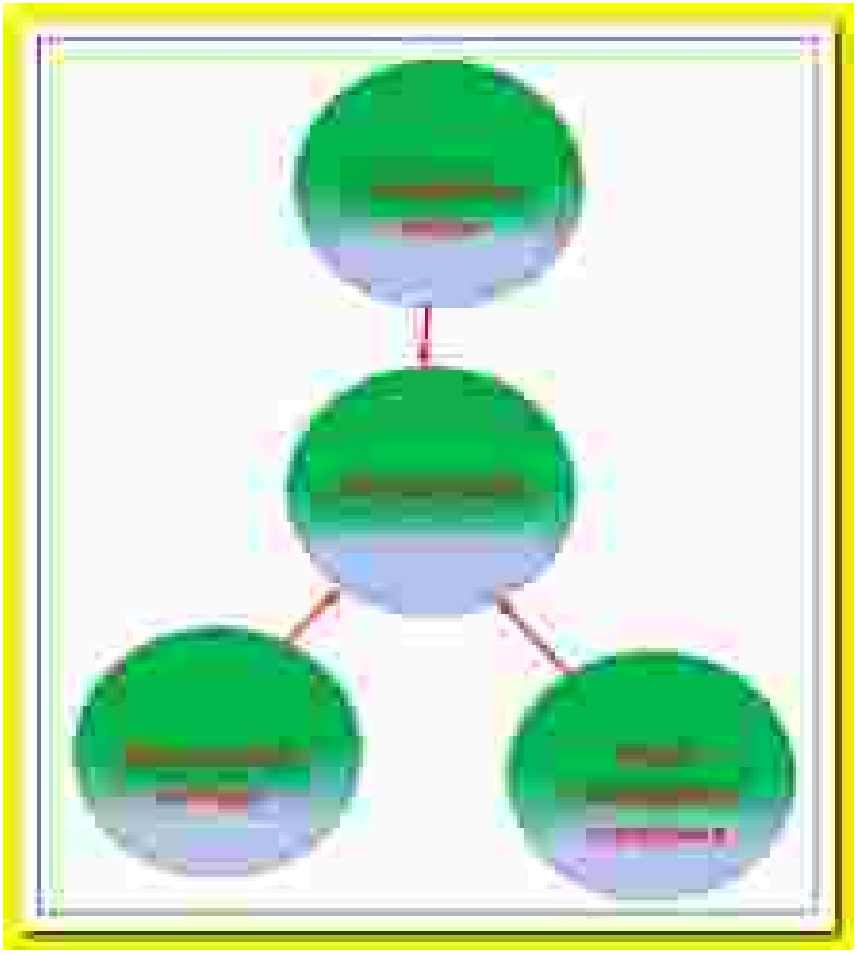
І.М. Пістунов,
О.Ю. Приходченко

ЕКОНОМЕТРИКА
з розрахунками на Excel



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ДНІПРОВСЬКА ПОЛІТЕХНІКА»

І.М. Пістунов, О.Ю. Приходченко

ЕКОНОМЕТРИКА.
З РОЗРАХУНКАМИ НА EXCEL

Навчальний посібник

Дніпро
НТУ «ДП»
2024

П34

Рекомендовано вченою радою університету як навчальний посібник з дисципліни „Економетрика” для студентів очної та заочної форм навчання в циклі професійної підготовки бакалавра за напрямами підготовки 051 Економіка та 071, 072, 073 Менеджмент (протокол № __ від __.__.2023 р).

Рецензенти:

Н.К. Васильєва, д-з. екон. наук, проф., завідувач кафедри інформаційних систем і технологій Дніпровського державного аграрно-економічного університету

К.Ф. Ковальчук, д-р екон. наук, проф., завідувач кафедри фінансів (Дніпропетровська національна металургійна академія України).

Пістунов І.М., Приходченко О.Ю.

П34 Економетрика. З розрахунками на Excel: Навч. Посібн. Дніпро: НТУ «ДП», 2024. 221 с. Режим доступу: http://pistunovi.inf.ua/EkMe.pdf (дата звернення: 17.01.2024). – Назва з екрану.

Подано теорію та приклади розв’язання задач з розрахунку коефіцієнтів для лінійних та нелінійних економіко-математичних моделей. Надано основні прийоми уникнення мультиколінеарності для багатофакторних моделей, розрахунки якості апроксимації та прогнозування.

Кожен розділ містить теорію, приклади розв’язання та індивідуальні завдання для закріплення отриманих знань. Наведено приклади застосування основних положень теорії ймовірності та математичної статистики в економіці. Розділи закінчуються методичними вказівками по розрахунках на комп’ютері за наведеними формулами.

Призначено для студентів вищих навчальних закладів і може бути корисним для викладачів, які застосовують теорію ймовірностей та математичну статистику у власних курсах.

Посібник базується на літературних джерелах вітчизняних та зарубіжних авторів, комп’ютерній програмі Excel та на досвіді викладання дисципліни «Економетрика» в Національному технічному університеті «Дніпровська політехніка»

УДК 519.21:330:004.67(075.8)

ЗМІСТ

ВСТУП

Вчені економісти вже близько ста років тому прийшли до висновку, що на відміну від технічних наук, розрахунки в яких базуються на відомих формулах з фізики, в ринковій економіці зв’язок різних факторів описати складно. Тому, для формульного визначення зв’язку економічних факторів почали використовувати статистичні дослідження, дані яких використовувалися для знайдення коефіцієнтів економіко-математичних моделей.

Назва «економетрика» вперше було введено в 1926 р норвезьким економістом і статистом Рагнаром Фришем (Frisch). Тому, можна сказати, що економетрія - порівняно молода галузь науки. 29 грудня 1930 за ініціативою І.Фішера (1867-1947), Р.Фріша (1895-1973), Я.Тінбергена (1903-1995), Й.Шумпетера (1883-1950), О.Андерсона (1887-1960) та інших вчених на засіданні Американської асоціації розвитку науки (США, Клівленд, штат Огайо) було створено економетричне суспільство, на якому норвезький учений Р.Фріш дав новій науці назву «економетрика». Створене Економетричне товариство позначило себе так: «Міжнародне товариство для розвитку економетричної теорії і її зв'язок зі статистикою і математикою».

Внаслідок активної діяльності економістів, були розроблені необхідні методи, які дозволяють розрахувати точність створеної моделі, якість прогнозування, методи розрахунку складних економічних процесів.

Визначення економетрики: «Економетрика – це не те ж саме, що економічна статистика. Вона не ідентична і тому, що ми називаємо економічної теорією, хоча значна частина цієї теорії носить кількісний характер. Економетрика не є синонімом додатків математики до економіки. Як показує досвід, кожна з трьох відправних точок – статистика, економічна теорія та математика – необхідна, але не достатня умова для розуміння кількісних співвідношень в сучасному економічному житті. Це єдність всіх трьох складових. І це єдність утворює економетрику».

Для економістів економетрика є однією з базових фундаментальних дисциплін тому, що фінансова діяльність у суспільстві з вільною економікою не дозволяє мати єдину формулу для економічних явищ. Навіть однотипні явища для різних країн змінюють, якщо не вид моделі, то її коефіцієнти. Тому, економісти повинні уміти складати такі формули, чому і вчить цей посібник.

У цьому посібнику ви знайдете не тільки теорію, але й багато задач, для яких подані і розв'язання, що дозволить при необхідності застосовувати їх при розв'язанні контрольних робіт і на практиці, для реальних умов. Окрім них, наведено індивідуальні завдання, які кожен студент має виконати протягом вивчення предмету. Кожне завдання має по декілька груп задач. Із кожної групи студент вибирає одну задачу по останній цифрі номера своєї залікової книжки, а номер варіанта числових значень вибирається з таблиць за номером студента в журналі студентської групи.

В.1. Порядок засвоєння матеріалу

В кінці кожного розділу подані індивідуальні завдання, які студенти виконують під час практичних занять та у вільний від навчання час.

Кожне завдання має на меті поглибити розуміння щодо основних принципів та законів економетрики, та розкриє можливості застосування цих положень на практиці.

Завдання треба здавати у письмовому вигляді оформити таким чином :

Економетрика Індивідуальне завдання № 1 Варіант задач №0, варіант числових даних № 17 Виконав: студент групи ЕК-00-1 Петренко Семен

Навчальним планом з дисципліни "Економетрика" передбачено виконання лабораторних робіт. Перед розв'язуванням задач необхідно вивчити відповід-

ний розділ теоретичного матеріалу. При виконанні лабораторної роботи студент повинний дотримувати таких правил:

1. Титульна сторінка роботи оформлюється за зразком. Або файл Excel'я має бути з ім'ям, де вказано прізвище студента та номер лабораторної роботи.

2. Розв'язування кожного завдання треба починати з наведення його повної умови.

3. Рішення завдань необхідно супроводжувати поясненнями, графіками та посиланнями на відповідні теоретичні поняття та формули.

4. Якщо лабораторна робота після перевірки не зарахована, треба виправити помилки згідно з зауваженнями викладача. Це необхідно робити у кінці роботи (або в окремому зошиті), написавши спочатку титул "Робота над помилками". Вносити зміни до тексту вже перевіреної роботи категорично забороняється. Доопрацьована лабораторна робота надається для повторної перевірки разом з першим варіантом.

5. Студент, що не виконав лабораторні роботи, до іспиту не допускається.

6. Для всіх лабораторних робіт студент робить заготовку у вигляді таблиці, показаної нижче.

7. Всі розрахунки повинні мати висновки щодо отриманих результатів

8. Кожна колонка таблиці заповнюється за допомогою функції RANDBETWEEN (СЛУЧМЕЖДУ) за формулами, вказаними у таблиці, наведеної нижче.

№ п/п	X1	X2	X3	X4	Y1	Y2
1	12n;28n	(12n;28n)*1.45	(12n;28n)/1.2	(12n;28n)*3.7	(12n;28n)/4.5	(12n;28n)*7.3
……						
25						

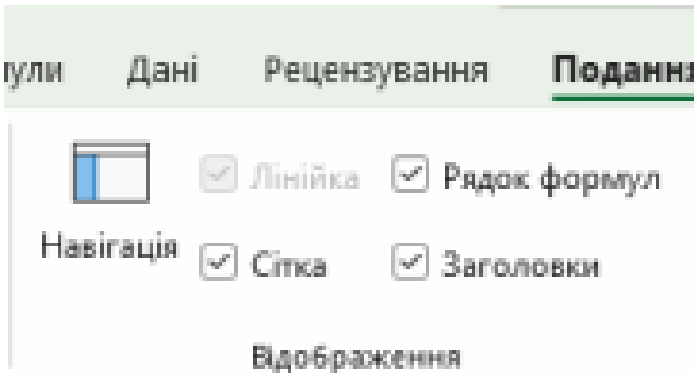
де n – номер студента за списком групи.

9. Всі отримані числа потрібно відмітити мишкою і натиснути Ctrl + C.

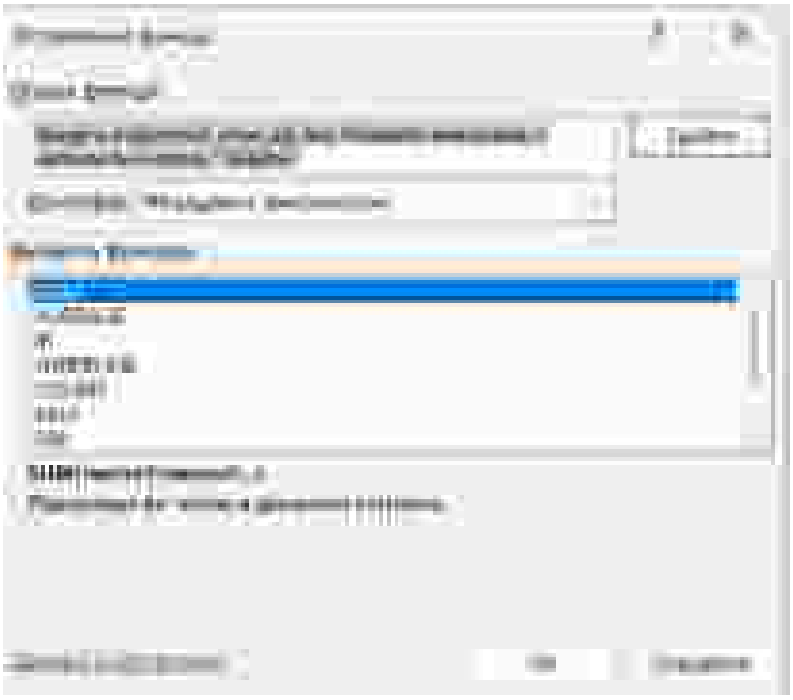
10. Обрати пункт меню «Основне – Вставити – Вставити значення». Тепер ваші числові ряди будуть незмінними при виконанні лабораторних робіт.

В.2. Налаштування електронних таблиць Excel

Електронні таблиці Excel дозволяють автоматизувати всі розрахунки ймовірності та статистичні розрахунки.. Для цього потрібно звернутися до функцій, які можна викликати за допомогою кнопки ***fx*** або через меню “Подання”. Потрібно відмітити галочками пункти Сітка, Рядок формул та Заголовки.



Тоді стане доступним кнопка формули, яка знаходиться поруч і рядком формул.

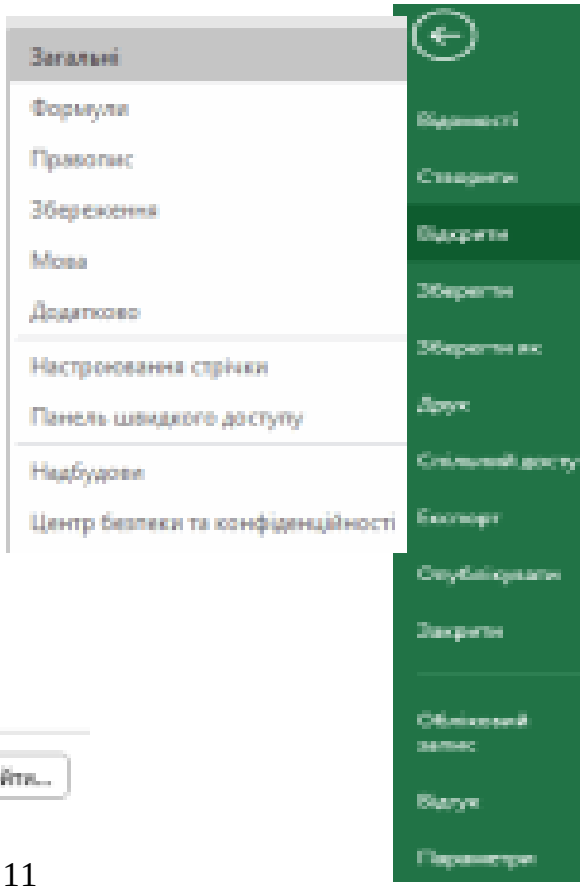


Вікно, що відчиниться, містить перелік усіх функцій Excel. З цього переліку ми можемо обрати потрібну функцію, яка має вигляд вікна з клітинками, у які треба або вписати потрібне число або вказати адресу клітинки чи ряду клітинок, що містять числа, потрібні для

розрахунків. Усі функції можуть бути вставлені у формули як звичайна адреса клітинки, оскільки вони повертають значення у вигляді числа. Або логічної змінної.

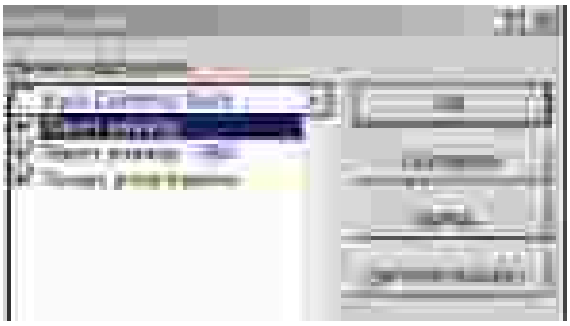
Розгалужена система надання довідок дозволяє швидко вибрати потрібну функцію та розібратися з її параметрами. Бо кожна з них супроводжується короткою анотацією, а при натисканні кнопки зі знаком питання, з'являється її розширений опис.

Перед початком роботи потрібно налаштувати надбудови Microsoft Excel.



Для цього потрібно вибрати пункт «Файл-Параметри». В меню, що з'явиться, вибрати пункт «Надбудови». Далі, біля віконці «Надбудови Excel», клацнути кнопку «Перейти».

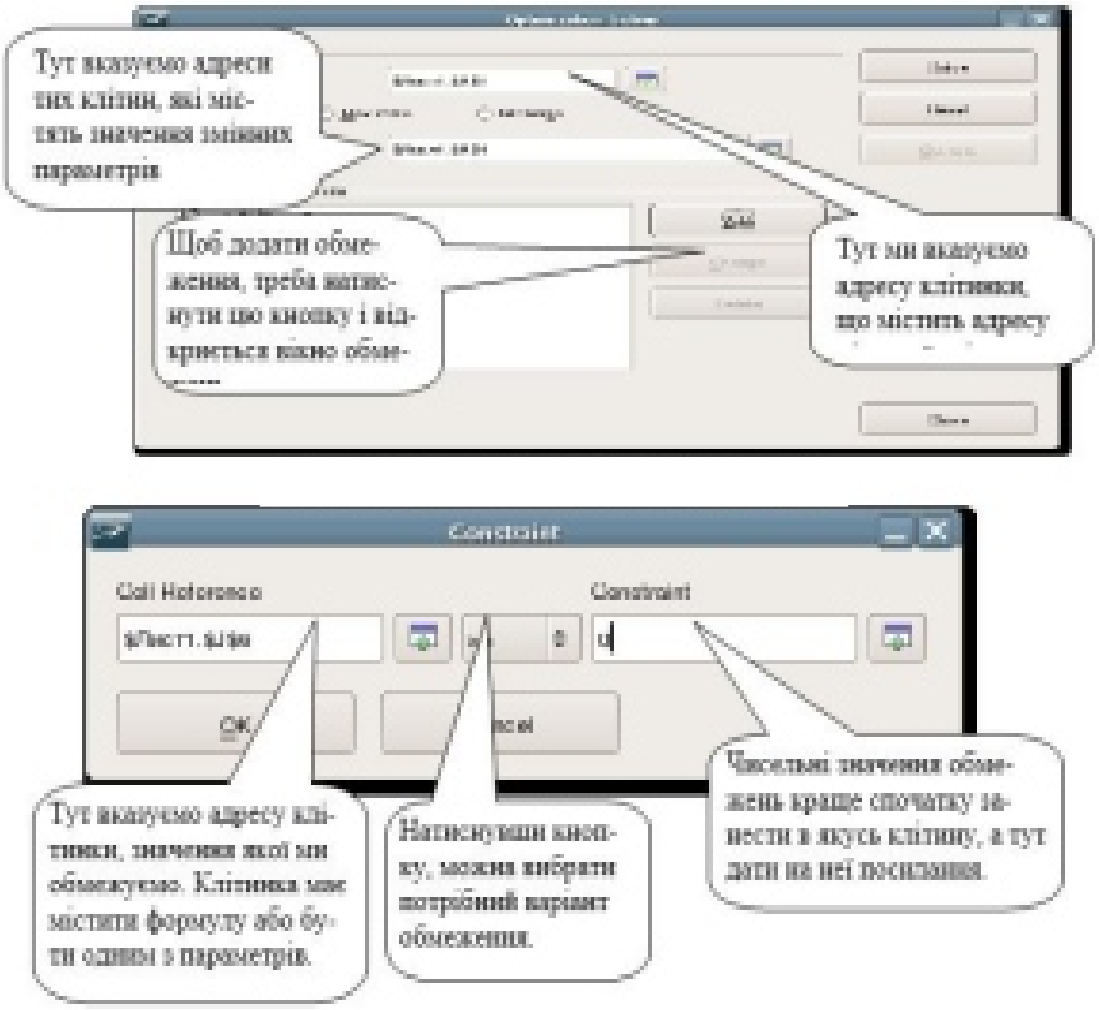
У вікні, що відкриється, відмітити три нижніх пункти і натиснути ОК.



Доступ до надбудов здійснюється через головне меню «Дані» і у правому кутку можна побачити ці надбудови.

Далі клацнути на кнопку «перейти» та відмітити вказані на рисунку позиції і натиснути «ОК» у цьому і наступних вікнах. Тепер в головному меню програми за пунктом «Дані» у правому кутку меню з'являться пункти «Data Analysis» та «Розв'язувач».

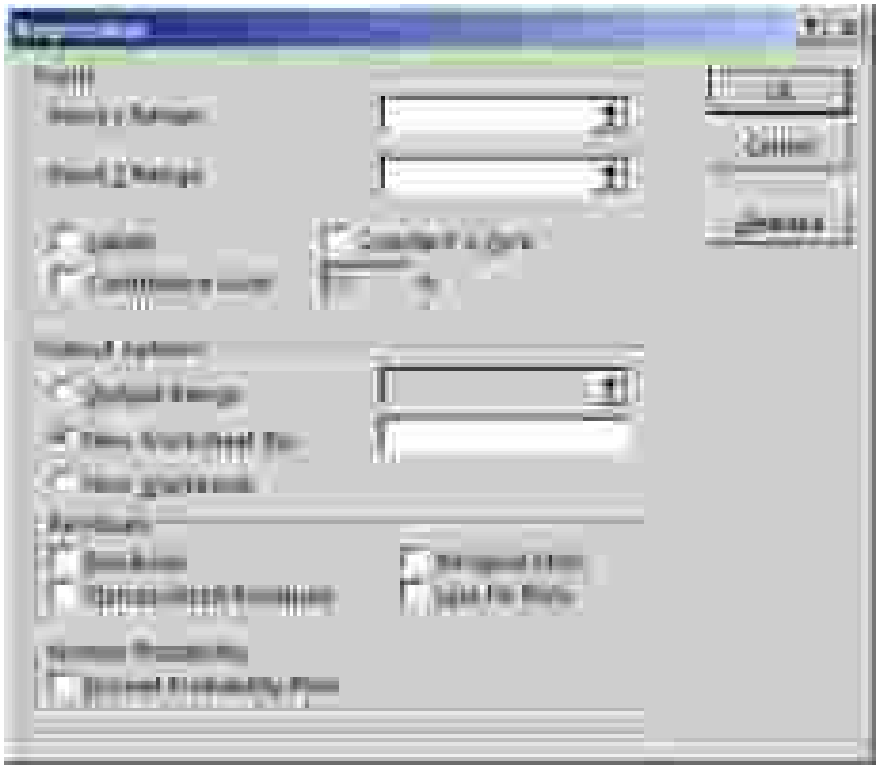
Якщо використовується програма «Розв'язувач», потрібно скористатися комірками вікна програми у порядку, вказаному нижче.



При використанні функції «Data Analysis» («Аналіз даних») ви обираєте

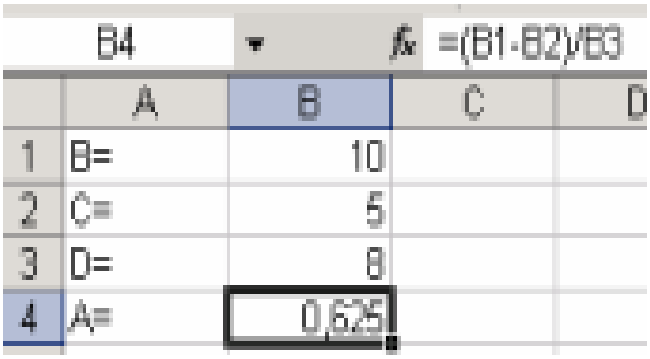
зі списку потрібний вид аналізу

Наприклад, якщо обираєте пункт «Регресія», то побачите наступне вікно



Після кожного пункту в посібнику подано приклад використання електронних таблиць при розрахунках за формулами, наведеними в цьому підрозділі. Наприклад, якщо потрібно провести розрахунки за формулою

$$A = \frac{B - C}{D},$$



для наступних числових значень параметрів B=10, C=5, D=8, то в підрозділі буде наведено малюнок, в якому видно фрагмент вікна Excel, де колонку А займають тестові визначення невідомих у формулі, колонку В – їх числові значення. Вікно f_x містить саму формулу розрахунку, де вказано адреси клітинок, які містять числові дані.

Додаткових пояснень за такими малюнками надаватися не буде, оскільки студенти повинні знати порядок складання формул в електронних таблицях Excel з курсу «Інформатика та комп'ютерна техніка».

В.3. Набуті компетенції

Після засвоєння матеріалу цього посібника, студенти мають отримати наступні компетенції

- Пояснювати моделі соціально-економічних явищ з погляду фундаментальних принципів і знань на основі розуміння основних напрямів розвитку економічної науки.
- Застосовувати відповідні економіко-математичні методи та моделі для вирішення економічних задач.
- Застосовувати набуті теоретичні знання для розв'язання практичних завдань та змістовно інтерпретувати отримані результати.
- Виконувати дослідження за встановленим замовленням.
- Демонструвати вміння абстрактно мислити, застосовувати аналіз та синтез для виявлення ключових характеристик економічних систем різного рівня, а також особливостей поведінки їх суб'єктів.

РОЗДІЛ 1
ОСНОВНІ ПОНЯТТЯ

Засвоєння матеріалів цього розділу забезпечить студентів основними поняттями, що у подальшій частині посібника будуть зустрічатися без додаткових пояснень.

1.1. Модель

Модель – речова, знакова або уявна (мислена) система, що відтворює, імітує, відображає принципи внутрішньої організації або функціонування, певні властивості, ознаки та/або/і характеристики об’єкта дослідження (оригіналу). Розрізняють фізичні, математичні та ін. моделі. Слово «модель» походить від латинського modulus, що означає міра, такт, ритм, величина. Воно пов'язане також із словом modus – копія, зразок.

Смислове навантаження терміна “модель” багатопланове:

а) зразок, взірцевий примірник чогось;

б) тип, марка конструкції;

в) те, що є матеріалом, натурою для відтворення;

г) зразок, з якого знімається форма для відливання в іншому матеріалі;

д) комп’ютерна модель,

е) розрахункова модель,

ж) теоретична модель (процесу, конструкції тощо).

Наприклад, модель – опис об'єкта (предмета, явища або процесу) на якій-небудь формалізованій мові, складений з метою вивчення його властивостей. Такий опис особливий корисний у випадках, коли дослідження самого об'єкта ускладнене або фізично неможливе.

Найчастіше в ролі моделі виступає інший матеріальний або уявний об'єкт, що замінює в процесі дослідження об'єкт-оригінал. Процес побудови моделі називається моделюванням. Таким чином, модель виступає як своєрідний інструмент для пізнання, який дослідник ставить між собою і об'єктом, і за допомогою якого вивчає об'єкт, що його цікавить.

Макетна модель – це реально існуюча модель, що відтворює модельовану систему у деякому масштабі

Математична модель – система математичних співвідношень, які описують досліджуваний процес або явище.

При одержанні математичної моделі використовують загальні закони природознавства, спеціальні закони конкретних наук, результати пасивних та активних експериментів, імітаційне моделювання за допомогою ЕОМ. Математична модель дозволяє передбачити хід процесу, розрахувати цільову функцію (вихідні параметри процесу), керувати процесом, проектувати системи з бажаними характеристиками.

Для створення математичних моделей можна використовувати будь які математичні засоби – мову диференційних або інтегральних рівнянь, теорії множин,абстрактної алгебри, математичну логіку, теорії ймовірностей та інші. Процес створення математичної моделі називається математичним моделюванням. Це найзагальніший та найчастіше використаний в науці, зокрема, в кібернетиці, метод досліджень.

Якщо відношення задаються аналітично, то їх можна розв'язати в замкнутому вигляді (явно) відносно шуканих змінних як функції від параметрів моделі, або в частково замкнутому вигляді (неявно), коли шукані змінні залежать від одного або багатьох параметрів моделі. До моделей цього класу належать диференційні, інтегральні, різницеві рівняння, ймовірнісні моделі, моделі математичного програмування та інші.

Якщо не можна здобути точний розв'язок математичної моделі, використовуються чисельні (обчислювальні) методи або інші види моделювання.

У залежності від того, якими є параметри системи та зовнішні збурення математичні моделі можуть бути детермінованими та стохастичними. Останні мають особливо важливе значення при дослідженні і проектуванні великих систем зі складними зв'язками і властивостями, які важко врахувати. Математичний опис неперервного процесу (напр., диференційними рівняннями) являє собою неперервну математичну модель

Якщо ж математична модель описує стан системи тільки для дискретних значень незалежної змінної і нехтує характером процесів, які протікають у проміжках між ними, то така модель є дискретною (тут важливим є вибір кроку дискретності, від якого залежить точність опису реального об'єкта його математична модель). Якщо параметри об'єкта, для якого розробляють Математичної моделі, можна вважати незалежними від часу, то така система описується стаціонарною моделлю, характерна особливість якої – постійні коефіцієнти. У протилежному випадку математична модель є нестаціонарною.

Дискретна модель – математична чи імітаційна модель, змінні якої приймають тільки дискретні значення, тобто змінюються від одного значення до іншого і не приймають проміжних значень (наприклад, модель, що прогнозує рівні запасів організації, ґрунтуючись на відвантаженнях, які змінюються, і платежах). Протилежність: неперервна модель.

Алгебраїчна система (*алгебраїчна структура*) – в математиці це не порожня множина з заданим на ній набором операцій та відношень, що задовольняють деякій системі аксіом.

Основним завданням абстрактної алгебри є вивчення властивостей аксіоматично заданих алгебраїчних систем.

Формально: об'єкт $\langle A;\ \Omega_F; \Omega_R \rangle$ де:

- A – не порожня множина,
- Ω_F – множина алгебраїчних операцій визначених на A,
- Ω_R – множина відношень визначених на A.

Множина A називається **носієм** алгебраїчної системи. Множини Ω_F, Ω_R називається **сигнатурою** алгебраїчної системи.

Якщо алгебраїчна система не містить операцій, вона називається моделлю, якщо не містить відношень, то – алгеброю.

Якщо не розглядають ніяких аксіом, яким мають задовольняти операції, то алгебраїчна система називається універсальною алгеброю заданої сигнатури Ω_F. **Концептуальна модель** має такі ознаки:

1. Формулювання змістовного і внутрішнього представлення, що поєднує концепцію користувача і розробника моделі. Вона включає в явному виді логіку, алгоритми, припущення й обмеження.

2. Абстрактна модель, яка виявляє причинно-наслідкові зв'язки, властиві досліджуваному об'єкту в межах, визначених цілями дослідження. По суті, це формальний опис об'єкта моделювання, який відображає концепцію (погляд) дослідника на проблему.

Аналітична модель – один з класів математичного моделювання.

Перевагою аналітичної моделі є те, що розв'язки можна аналізувати математичними методами. Недоліком аналітичних моделей є спрощення реальних ситуацій з метою отримання аналітичних розв'язків.

В економіці – модель, що складається з системи розв'язних рівнянь, наприклад, система розв'язних рівнянь, що представляють закони попиту та пропозиції на світовому ринку.

1.2. Математичне моделювання

Математичне моделювання – метод дослідження процесів або явищ шляхом створення їхніх математичних моделей, дослідження цих моделей.

В основу методу покладено ідентичність форми рівнянь і однозначність співвідношень між змінними в рівняннях оригіналу і моделі, тобто, їх аналогії. Математичні моделі досліджуються, як правило, із допомогою аналогових обчислювальних машин, цифрових обчислювальних машин, комп'ютерів.

На початку 60-их років XX сторіччя було розроблено один із методів математичного моделювання – квазіаналогове моделювання. Цей метод полягає

в дослідженні не досліджуваного явища, а явища або процесу іншої фізичної природи, яке описується співвідношеннями, еквівалентними відносно отримуваних результатів.

Розрізняють *геометричне. (наочне) моделювання*, здійснюване на макетах або об'ємних моделях. Вони передають в наочній формі просторові властивості об'єкту, його зовнішній вигляд, співвідношення і взаємозв’язок його частин. Такі, наприклад, модель, яка відтворює форми літака, і макет мікрорайону.

За допомогою методів *фізичного моделювання* вивчають фізико-хімічні, технологічні, економічні процеси, що відбуваються в оригіналі. Раніше ці процеси вивчалися за допомогою аналогових пристроїв, придатних для моделювання різноманітних динамічних і статичних процесів. Але поява персональних комп’ютерів із сучасним математичним забезпеченням має величезні можливості для моделювання поведінки складних систем (технічних, біологічних і економічних). Як правило, фізичне моделювання здійснюється на базі логіко-математичної моделі процесу, що вивчається, і який грає роль проміжної ланки між об'єктом і його фізичною моделлю.

Фундаментальне значення у всіх областях науки і техніки має інформаційне моделювання. В якості моделей тут використовуються схеми і графіки, креслення, а також символи, що формуються в слова, формули і рівняння. Ці матеріальні утворення можуть слугувати моделями лише у тому випадку, коли визначені допустимі для них операції перетворення і правила їх виконання. Найважливішу роль грає інформаційне моделювання, здійснюване засобами математичного і логічного апарату. Його називають логіко-математичним моделюванням. Використовувані при цьому символи (букви і цифри) і їх послідовності (формули, рівняння і нерівності) описують властивості оригіналу, що вивчаються, і є його логіко-математичною (або математичною) моделлю.

На рис. 2.1 наведено класифікацію прийомів та методів моделювання.

Аналіз економічних явищ, планування розвитку народного господарства, розробка ефективних методів управління, фізичне моделювання процесів в

економічних системах і, нарешті, автоматизація планово-економічних розрахунків засновані на їх математичному або, як його зазвичай називають, економіко-математичному моделюванні. Точність і обґрунтованість аналізу і управління залежать від об'єктивності і точності віддзеркалення в моделях реальних економічних процесів, зв'язків між параметрами економічної системи, обмежень, що накладаються на неї зовнішніми умовами, достовірністю використовуваної інформації.

На рис. 2.2 представлено класифікацію методів моделювання соціально-економічних процесів і явищ.

Пояснимо наведені вище схеми для глибшого розуміння напрямків застосування методів моделювання при моделюванні соціально-економічних явищ.

Моделі лінійного програмування (параметричного, непараметричного, детерміністичного, стохастичного, цілочисельного, нецілочисельного) – вирішення завдань оперативно-календарного планування, технічної підготовки виробництва (складання карт, розкрою листових і смугових матеріалів, знаходження оптимальної шихти, виробничої суміші і т. п.), транспортних завдань і деяких завдань техніко-економічного планування (визначення оптимальної виробничої програми і т. і.);

- нелінійного програмування – ухвалення рішень в області техніко-економічного планування, оперативно-календарного планування, з питань загальної господарської політики підприємства, фінансування і кредитування;
- динамічного програмування – вибір політики заміни устаткування, оптимальний розподіл амортизаційних відрахувань на заміну устаткування і відновлення його, визначення оптимальних умов розширеного відтворення;
- блокового програмування – вирішення завдань великого розміру шляхом розбиття їх па ряд моделей з меншим числом змінних н обмежень;

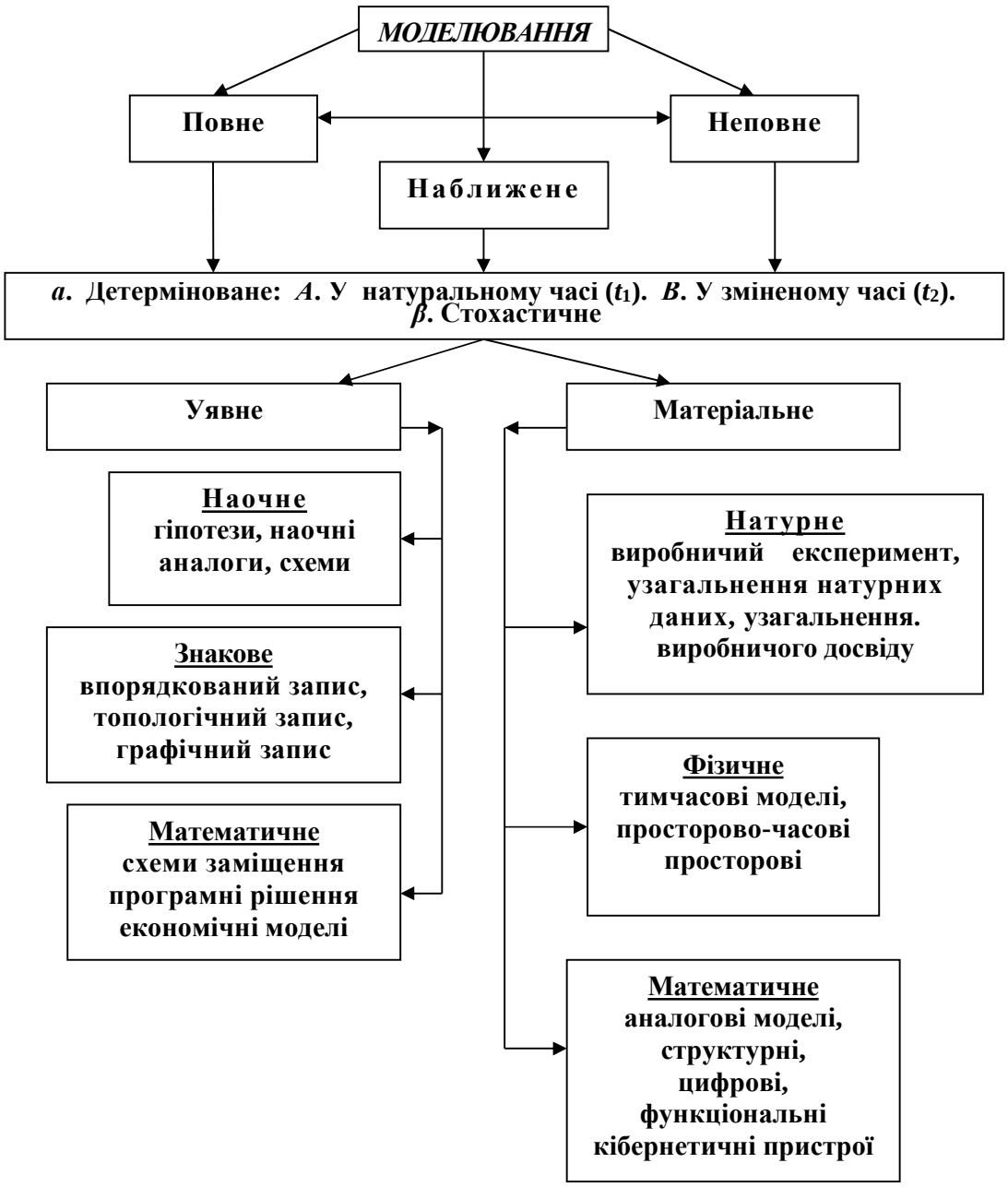


Рис. 1.1. Класифікація методів та прийомів моделювання

- балансових методів аналізу – вирішення проблем пропорційного розвитку виробництва на різних рівнях, складання техпромфінплану підприємства в матричній формі;
- мережевого планування – управління здійсненням крупних індивідуальних і рідше серійних розробок: є формою виразу ухваленого

рішення (з вказівкою цілей, завдань, термінів виконання, виконавців, резервів), інструментом оперативного керівництва виконанням рішення і контролю за ходом виконання;

- транспортних завдань на мережі – розробка оптимальних схем прикріплення постачальників до споживачів;
- теорії аналізу кореляцій і регресії і теорії дисперсійного аналізу – вивчення статистичних взаємозв'язків в економічних процесах встановлення різних нормативів – трудових, вартісних, по витраті матеріалів і інших;
- теорії масового обслуговування – встановлення оптимальних співвідношень між розмірами основного і допоміжного виробництв і окремими частинами усередині кожного з них, якщо процеси в них мають елементи нерегулярності і можуть бути представлені як масове обслуговування;

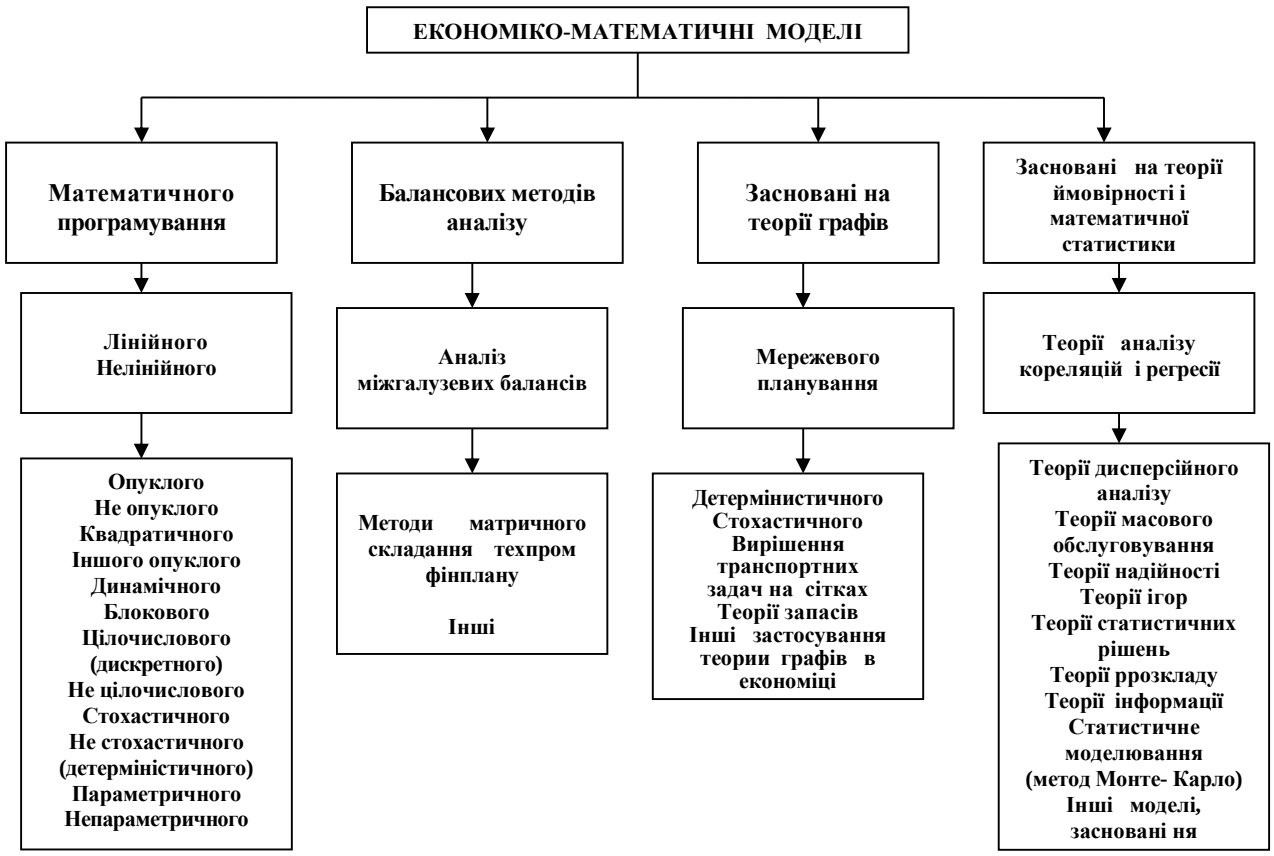


Рис. 1.2. Структура методів моделювання

- теорії надійності – вирішення проблем надійності і довговічності устаткування, підвищення якості продукції і роботи;
- теорії запасів – встановлення оптимальних розмірів оборотних фондів на підприємстві, вирішення деяких завдань оперативно-календарного планування в серійному і масовому виробництві, визначення оптимальних заділів;
- теорії ігор і теорії статистичних рішень – управління процесами взаємовідношення підприємства з ринком, страхування від стихійних лих, створення сезонних запасів сировини і матеріалів;
- теорії інформації – вдосконалення інформаційних потоків в управлінні, вирішення інших завдань, не інформаційних, але схожих за типом даних процесів;
- теорії розкладів – визначення раціональної послідовності запуску деталей у виробництво, встановлення оптимальної тривалості виробничого циклу виробів;
- статистичного моделювання – вирішення завдань теорій масового обслуговування, ігор, статистичних рішень, надійності, інформації, запасів, для яких немає власного математичного апарату.
- наочні і знакові – при оптимізації управлінських вирішень всіх видів, як і взагалі у всій економічній роботі, тому виділити переважну область їх застосування важко. Такі, як натуральні, фізичні, застосовуються при оптимізації управлінських рішень ще рідко, епізодично, що також утрудняє можливість виділення якихось конкретних областей переважного їх застосування.

1.3. Об’єкт. Його параметри і фактори

Об’єктом прийнято називати деяке явище, підприємство, механізм, технологічний процес, які є предметом вивчення дослідника.

Частіше, об’єкт зображується як прямокутник до якого проведені стрілки, деякі з яких входять у прямокутник, а деякі виходять. Ці стрілки позначають ті явища, які можна спостерігати і вимірювати за межами об’єкта. Ці явища називаються факторами. Розрізняють вхідні (позначаються стрілками до прямокутника) і вихідні (від прямокутника) фактори або входи та виходи. На рис. 1.3 показано приклад зображення об’єкта та вхідних (X_i) і вихідних (Y_i) факторів. Вхідні фактори розділяються на фактори управління – такі фактори, значеннями яких може керувати дослідних за власним розсудом; збурення – фактори, значеннями яких керувати не можна, але їх можна виміряти; перешкоди – фактори, значеннями яких не тільки не можна керувати, але і величину перешкод частіше виміряти теж неможливо. Для того, щоб розділити ці вхідні фактори за змістом, інколи вживають окремі позначення для кожної з цих груп, наприклад, $\vec{U},\vec{Y},\vec{Z}$ відповідно. Тоді вихідні фактори позначаються як $\vec{X}$.

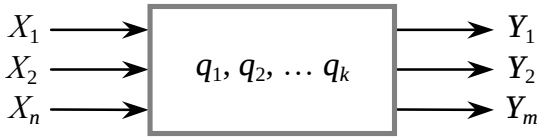


Рис. 1.3. Графічна схема об’єкта

Кожен об’єкт має власну структуру, яка, після взаємодії зі вхідними факторами, перетворює їх на вихідні фактори за певними правилами. Ці правила можуть бути виражені як вербально (словами) так і за допомогою системи рівнянь. Внутрішня структура об’єкта характеризується параметрами, які представляють собою як правила перетворення вхідних факторів так і чисельні значення. На рис.1.3 вони позначаються як q_i, але існують і інші позначення такі як a_i, h_i тощо. Довільний реальний об’єкт має незліченну кількість властивостей (характеристик), і за кожною з них його можна віднести до тієї чи іншої системи як її елемент. Якщо, скажімо, розглядати університет як окрему систему, то з погляду його ректора, проректора з фінансово-господарських

питань, головного бухгалтера, начальника служби охорони він складатиметься з різних підсистем та елементів, наділених неоднаковими функціональними властивостями.

Іншим прикладом може бути такий об’єкт як прибуток після сплати податків. Факторами управління для нього будуть – дохід та витрати, факторами збурення – ставки податків, а фактори перешкод у цьому випадку можна вважати економічну кон’юнктуру, яка наперед невизначеним чином впливає на витрати та доходи. Вихідним же фактором розглянутого об’єкту є сума чистого прибутку.

Введемо ще одне поняття – *внутрішній стан об’єкту*. З його допомогою кількісно характеризуються його істотні властивості. Так, з внутрішнім станом об’єкту – цеху, можуть бути співставлені наявні потужності, трудові ресурси, запаси предметів праці і т. д. Отже, внутрішній стан елементу відображає те, що цікавлять нас потенційні характеристики реального об'єкту – кількість речовини, енергії, інформації, його пропускну спроможність і ін. Зіставимо з ним n – вимірний вектор g (з безперервними або дискретними компонентами) так, що попарно помітним станам відповідають відповідні значення цього вектору. Його компоненти інколи $(g_1, g_2, \dots, g_k)$ називають координатами стану або точками елементу, що відображає в n - вимірному просторі станів. Вектори і множина допустимих значень їх компонентів характеризують можливі стани елементів і інтенсивності їх входів і виходів.

Кількісні характеристики властивостей реального об'єкту в загальному випадку залежать від зовнішніх дій, і їх зміни обумовлюють зміни його стану. Так, виробнича потужність цеху змінюється з інтенсивністю, визначуваною інтенсивністю капітальних вкладень, що направляються на її приріст. Інтенсивність зміни стану оперативного накопичувача ЕОМ визначається інтенсивністю надходження в нього інформації.

1.4. Система

Системою є сукупність об'єктів і процесів, званих компонентами або елементами, взаємозв'язаних і таких, що взаємодіють між собою, які утворюють єдине ціле, таке, що володіє властивостями, не властивими складовим його компонентам, узятим окремо.

Функціонування елементів системи як єдиного цілого забезпечується *зв'язками* між її елементами. У технічній системі вони формуються при її проектуванні, в біологічній вони виникають природним чином в процесі зародження і розвитку організму. У економічних системах зв'язки можуть організовуватися в плановому порядку або складатися стихійно під впливом ринкового механізму. Склад елементів і спосіб їх об'єднання визначають структуру системи.

Кажучи про систему, зазвичай мають на увазі сукупність елементів (об’єктів), яка реалізовує між ними певні відносини, що цікавлять нас, з фіксованими властивостями. Економічною системою є, наприклад, підприємство як сукупність цехів, біологічною – система кровообігу, що включає серце і кровоносні судини, технічною – електронно-обчислювальна машина, що складається з комплексу пристроїв для обробки інформації.

Ці приклади ілюструють інтуїтивні уявлення про систему як об'єкт, об'єднуючий множину матеріальних елементів, які функціонують як єдине ціле.

Будь-який об'єкт, прийнятий як відправний, може бути представлений як елемент (або підсистема) деякої системи вищого рангу і як система по відношенню до деякої сукупності підсистем нижчого рангу. Рухаючись вгору по ступенях цієї ієрархії, ми прийдемо до «універсальної» системи – Всесвіту, рух вниз приведе нас до її первинного елементу – елементарної частинки. Тому при аналізі і проектуванні конкретної системи виникає проблема визначення тієї «ділянки» ієрархії, яка входить в її компетенцію, і вибору елементу, що приймається як «первинний».

Взаємодії реальних об'єктів, що охоплюються конкретною системою, і її взаємодії із зовнішнім середовищем такі ж різноманітні, як властивості об'єкту і середовища. Система має вхідні та вихідні фактори, має і свої параметри, але ці останні вже визначаються параметрами об’єктів, які складають систему. Систему також можна описати словесно, математично і графічно.

При аналізі і проектуванні системи беруть до уваги лише ті зв'язки, які істотно впливають на її функціонування; останніми нехтують, а у разі потреби систему захищають від їх «паразитного» впливу, що розглядається як обурення (перешкоди). Вживаючи поняття вхідних факторів (входів) елементу мають на увазі, що вони відображають найбільш істотні зв'язки (матеріально-речові і інформаційні) між об'єктами. Таким чином, поняття «система» є абстракцією не тільки відносно властивостей охоплюваних нею реальних об'єктів, але і відносно зв'язків між ними.

При описі системи для кожного з її елементів треба брати відповідні йому рівняння зв'язку між його входами і виходами з тією обмовкою, що функціональні змінні перейменовуються відповідно до прийнятих в системі. Об'єднані таким чином рівняння складуть систему рівнянь, які дають математичний опис системи.

На рис. 1.4 представлена спрощена схема підприємства як об'єкту управління. Аналіз таких об'єктів заснований на комплексному його розгляді як єдиного цілого, такого, що складається з взаємозв'язаних елементів, а також на з'ясуванні впливу цих елементів на функціонування всієї системи.

Частина системи, яка сприймає дію навколишнього середовища, називається входом системи, частина, яка впливає на навколишнє середовище і інші системи, називається виходом.

На вхід системи поступає три типи дій:

– вхідні керовані змінні $\vec{U} = \{u_1, u_2, \ldots, u_n\}$ є вектором управлінь. У виробничо-економічних системах це, як правило, цілеспрямовано змінні ресурси (трудові, енергетичні, матеріальні);

– вхідні некеровані (але контрольовані) змінні $\vec{Y} = \{y_1, y_2, \ldots, y_r\}$, під якими у виробничій економіці розуміють, як правило, якість змінних цілеспрямовано ресурсів (якість сировини, кваліфікація фахівців, види енергетичних ресурсів і ін.). Їх називають збурення;

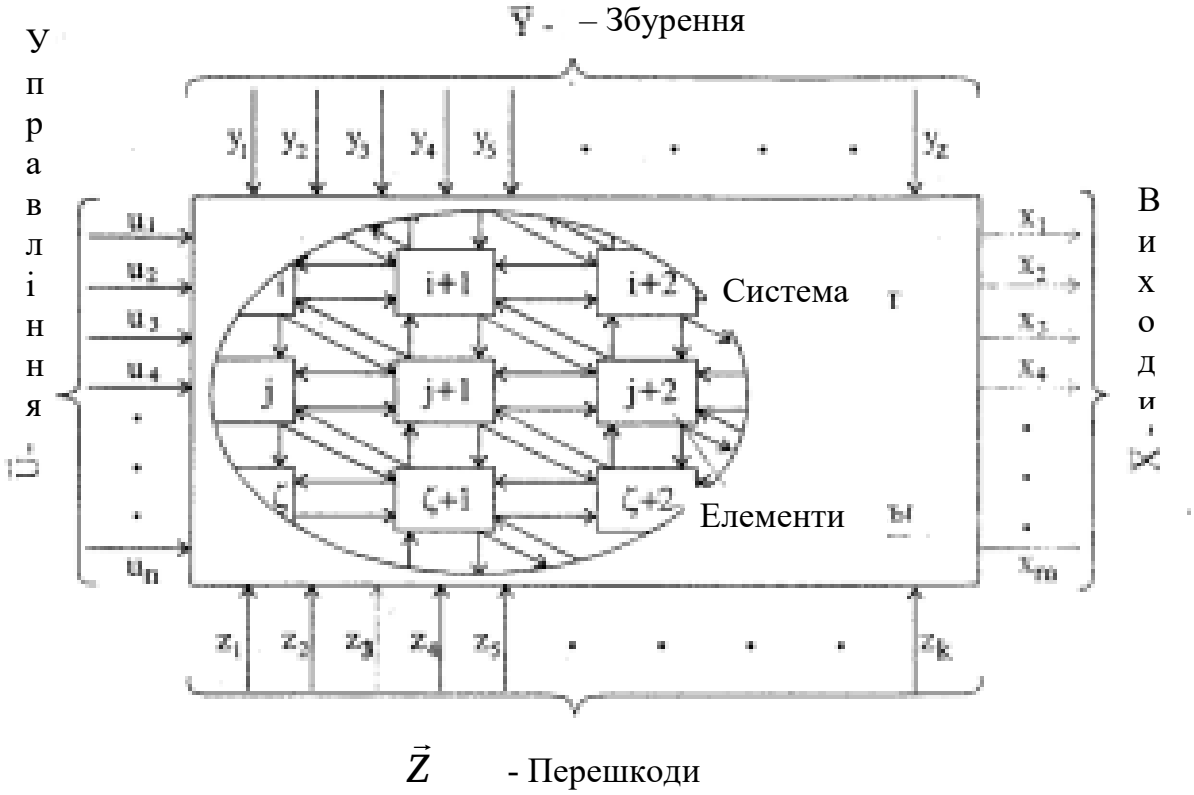


Рис. 1.4. Спрощена схема підприємства як об'єкту управління.

– неконтрольовані чинники $\vec{Z} = \{z_1, z_2, \ldots, z_k\}$ є вектором перешкод. По суті, це вектор, про який менеджерові відомо дуже мало або взагалі нічого не відомо. Тоді цей вектор взагалі не враховується. Якщо деякі статистичні характеристики компонентів вектора відомі, то їх слід враховувати при розробці моделі системи. Про наявність збурень, що діють на систему можна здогадатися тоді, коли при незмінних факторах керування та перешкод, вихідні фактори починають змінюватися.

Будь-яка економічна, виробничо-економічна система призначена для перетворення векторів вхідних дій $\vec{Y}, \vec{U}, \vec{Z}$ у вектор вихідний змінної $\vec{X} =$

$\{x_1, x_2, \ldots, x_m\}$. У цьому і полягає суть виробничих процесів, що перетворюють матеріальні, енергетичні і трудові ресурси в продукт, що оцінюється економічними показниками (наприклад, собівартістю).

На рис. 1.4 вихідні змінні $x_1, x_2, ..., x_m$ характеризують результати роботи досліджуваної системи, а вектор $\vec{X}$ носить назву вектора стану.

Промислове виробництво, представлене у вигляді структури на рис. 1.4, можна розглядати як керований об'єкт. Процес виробництва тієї або іншої продукції завжди порушуватиметься унаслідок зміни вектора обурень $\vec{Y}$ об'єкту, що діє на вході. На практиці обуреннями є все ті випадкові дії на об'єкт, які відхиляють параметри виробничого процесу від заданих рівнів. На сучасних виробництвах можна виділити певні групи збурень:

y_1 – технологічні, такі, що характеризують відхилення параметрів процесів і засобів праці; y_2 – психофізичні і медичні, пов'язані із захворюванням працівників і коливаннями їх індивідуальної продуктивності праці залежно від зовнішніх умов; y_3 – соціальні, пов'язані з порушенням трудової дисципліни; y_4 – кліматичні; y_5 – організаційні і інформаційні, пов'язані з недосконалістю організації виробництва, планування, обробки і відображення інформації і ін. Таким чином, можна записати, що $\vec{Y} = \{y_1, y_2, y_3, y_4, y_5\}$.

Необхідно відзначити, що перераховані вище збурення діють на об'єкт із зовнішнього середовища, тому можуть бути охарактеризовані як первинні збурення (чинники, що збурюють). Треба також сказати, що дія первинних збурень породжує вторинні збурення, точка додатку яких вже не співпадає з первинними обуреннями із-за наявності внутрішніх зв'язків між елементами, що входять в об'єкт.

Організована (керована) виробнича система (об’єкт) повинна функціонувати таким чином: при зміні компонент $y_1, \ldots, y_r$ вектору збурень $\vec{Y}$ компоненти $u_1, \ldots, u_n$ вектору управлінь $\vec{U}$ змінюються так, щоб компоненти $x_1, x_2, \ldots, x_m$ вектору стану $\vec{X}$ відповідали плановим значенням. Такий об'єкт називається керованим.

Важливою властивістю складних систем є **емерджентність** – тобто наявність таких специфічних властивостей системи, які не випливають з властивостей, притаманних її елементам, а виникають у процесі їхньої взаємодії як наслідок відповідних кооперативних ефектів. Саме емерджентні властивості економічних систем є найменш доступними для спостереження та вимірювання, що вельми утруднює дослідження таких систем та управління ними. Загальні закономірності появи нових властивостей, породжуваних об’єднанням економічних об’єктів, явищ та процесів, можна виявити та кількісно описати, лише проаналізувавши значний обсяг інформації. Емерджентність легко зрозуміти на прикладах статистичної рівноваги: деякі ознаки можуть бути характерними для всієї статистичної сукупності явищ, не будучи характерними окремим об'єктам і для виведення його властивостей на основі розглянутих властивостей складових його частин.

1.5. Класифікація систем

Частинами, складовими різного роду складні утворення, тобто системи, можуть бути не тільки якісь фізичні предмети, але і деякі уявні, ідеальні об'єкти. Тому всі системи можна підрозділити на **фізичні** (або емпіричні) і **абстрактні** (або концептуальні). Фізичні складаються з реально існуючих (природних або штучних) об'єктів: машин, виробів, устаткування, працівників і так далі Абстрактні – з об'єктів, що існують лише в думці людини: поняття, ідеї, гіпотези, плани і тому подібне. Наприклад, абстрактними системами будуть сукупність принципів, складових підстав якого-небудь учення, сукупність вживаних в дослідженнях методів, таких, що створюють певну методологію, і так далі.

Є також і системи змішаного типу. Так, уявляючи собі особу якої-небудь людини як систему, ми розглядаємо сукупність його фізичних даних і індивідуально виражених психологічних властивостей (характер, темперамент, здібності н інші), що є абстрактними поняттями.

Як друга ознака класифікації систем можна прийняти їх походження. Системи, які виникають в природних процесах, називають природними: Галактика, сонячна система, клімат, гірські системи і тому подібне. Якщо людина змінила систему шляхом перетворення складових її об'єктів, властивостей і взаємозв'язків, тоді вони називаються системами, зробленими людиною (штучними): транспорт, телефон, телеграф, конвеєр, ракетна система. **Природні** і **штучні** системи можуть бути фізичними або абстрактними.

Системи можна розділити на види також за ознакою структури їх побудови і значущості тієї ролі, яку грають у них окремі складові частини порівняно з роллю інших частин тієї ж системи. У деяких системах одному зі складових їх об'єктів або одній з частин може належати домінуюча роль. Значення цього підрозділу системи набагато перевершує значення інших частин. Такий компонент системи виступатиме як центральний, такий, що визначає функціонування системи. Інші частини системи підкоряються дії центрального підрозділу. Такі системи називаються **централізованими.** Серед природних фізичних систем прикладом централізованої системи служить атом, в якому електрони обертаються навколо того, що займає центральне положення в системі ядра, а також сонячна система. Із зроблених людиною систем централізованою назвемо систему виробництва, в якому всі цехи працюють заради забезпечення безперебійної роботи одного з цехів, наприклад, складального конвеєра на машинобудівному заводі. У інших системах всі складові їх об'єкти приблизно однаково значущі. Структурно вони розташовуються не навколо деякого центрального об'єкту, як супутники, а взаємозв'язані послідовно або паралельно і мають приблизно однакове значення для функціонування системи. Такі системи називаються децентралізованими. В масштабі країни система транспорту, в якій раціонально поєднуються всі його види (залізничий, автомобільний, повітряний, річковий, морський), може розглядатися як **децентралізована** система. Децентралізованою (у технічному аспекті) є також система електростанцій, що працюють на єдине енергетичне кільце.

Системи можна класифікувати також по ступеню визначеності їх дії (функціонування). Якщо складові системи, об'єкти або частини, і зв'язки між ними функціонують таким чином, який точно передбачений, систему називають **детермінованою** (визначеною). Будь-який подальший стан такої системи можна точно передбачити, знаючи попередній стан і програму переходу системи в наступний стан. Системи, в яких складові їх об'єкти (частини) і зв'язки між ними функціонують так, що не можна точно затверджувати про послідовність їх станів, детально передбачати їх поведінку, називають **стохастичними**, тобто випадковими, імовірнісними або такими, що нерегулярно функціонують. У таких системах майбутні стани, тобто поведінка їх, описується (передбачається) за допомогою методів теорії вірогідності і математичної статистики. Саме такі системи частіше зустрічаються у виробничій практиці, в економіці. Їх важче досліджувати і описувати. Ними, в першу чергу, і займається економічна кібернетика.

Системи існують в певному навколишньому їх середовищі, обумовлюються нею і можуть і більшою або в меншій мірі обмінюватися з нею речовиною, енергією, інформацією. У зв'язку з цим вони діляться на **відкриті *і* закриті**. Ті, які з певною регулярністю і певним чином обмінюються з навколишнім середовищем речовиною, енергією або інформацією, називаються відкритими (незамкнутими) по якомусь з цих ознак (відкриті для енергії, відкриті для інформації, для речовини). Якщо такий обмін з навколишнім середовищем не спостерігається або ж він незначний і їм нехтують, системи будуть закритими (замкнутими).

Системи можуть по-різному реагувати на дію зовнішнього середовища: одні залишаються пасивними по відношенню до неї, інші пристосовуються до її дій, до зовнішніх змін. Залежно від цього їх ділять ще на дві групи: пасивні до змін зовнішнього середовища називають **неадаптивними**, а реагуючі на ці зміни, здатні пристосовуватися до них, називають **адаптивними**.

Системи, що складаються з людей, називають **соціальними** системами. Природно, в такі системи входять і різні об'єкти, з яких складаються фізичні

неживі системи – машини, будівлі, споруди. Проте найбільш важливою, є така, в якій визначна роль належить організаційній структурі і поведінці людей. Як приклад приведемо такі системи: виробничий колектив, громадські організації, технічні, спортивні і інші суспільства.

Системами, що складаються з одних тільки машин і механізмів і що діють автоматично, без участі людей, є **машинні** системи. Виділення їх майже завжди до певної міри умовно, оскільки вони не здатні виробляти свій власний початковий стан і підтримувати своє функціонування. При більш загальному розгляді ці системи виступають зазвичай як частини більші, що включають і людей.

Для сучасного виробництва і для інших областей нашого життя характерна наявність сукупності машин, що виконують певні складні функції за участю людини, причому основне навантаження лягає не на його м'язи, а на психічні процеси сприйняття, запам'ятовування, мислення. Такі системи виділяють в самостійну групу **людино-машинних** систем. Сюди ж відносяться і людино-машинні системи, що управляють, в яких в процесі взаємодії людини і обчислювальних машин виробляються рішення – сигнали, що управляють.

Системи, що існують тривалий час в порівнянні з обмеженим періодом діяльності людей в них, називають **постійними**. Наприклад, підприємство можна розглядати як постійну систему при розробці планів його розвитку на перспективу. Системи, які створюються на заданий період часу, а потім ліквідовуються, називаються **тимчасовими**. Група, створена для розробки деякого науково-дослідного проекту, а потім розформована, є тимчасовою системою.

Якщо властивості і функції системи протягом тривалого часу істотно не змінюються або змінюються у формі циклів, що повторюються, її називають **стабільною**. Система з властивостями, що змінюються, і функціями і без циклів, що строго повторюються, в цих змінах називається **нестабільною**.

Залежно від кількості об'єктів, включених в системи, їх підрозділяють на **великі** і **малі.**

Прості системи мають просту структуру і виконують якісь нескладні функції. **Складні** системи складаються з великої кількості взаємозв'язаних і взаємодіючих між собою частин (мають складну структуру) і виконують якусь достатньо складну функцію. Приналежність тієї або іншої конкретної системи до класу простих або складних є відносною. Іноді мета і завдання досліджень дозволяють абстрагуватися і розглядати складну систему як просту. Причому результати від цього не страждають, а проводити дослідження легко.

Отже, окрім описаних вже систем на види їх можна класифікувати на статичні і динамічні. Ті системи, в яких показники, що характеризують стан елементів, постійні в часі (параметри), а зв'язки між елементами жорсткі, називаються **статичними**. Ті ж системи, в яких показники, що характеризують стан елементів, змінні в часі, а зв'язки між елементами – динамічні, називаються **динамічними**.

У загальному випадку в будь-якій складній системі всі кількісні показники (і ті, що характеризують стан елементів, і ті, що визначають стан їх входів і виходів) змінюються в часі, тобто набір їх є $A = (a_1, a_2, a_3, \dots, a_n, \dots, a_N)$ якась функція від часу t:

$$A = f(t) = (a_1(t), a_2(t), a_3(t), \dots, a_n(t), \dots, a_N(t)) \qquad (1.1)$$

Якщо показниками набору виступають величини, що характеризують стан елементів системи, і відомий конкретний вираз функції $f(t)$, то можна вивчити зміни, що відбуваються в системі з плином часу. Таким чином, функція (1.1) як би відображає рух системи в часі, через що її називають траєкторією або лінією поведінки системи. Значення цієї функції при $t = 0$ визначає стан системи на початку даного періоду, а при якомусь значенні стану $t = t_i$ її стан через t одиниць часу.

Деякі з показників a_i на відрізку часу, впродовж якого досліджується система, можуть не змінюватися або змінюватися настільки мало, що при прийнятій точності розрахунків цими змінами нехтують. Такі показники

називаються **параметрами**. Показники, які змінюються в часі істотно і ці зміни враховуються в дослідженнях, називають **змінними**. Показники, що характеризують зв'язки між елементами системи, завдяки яким здійснюється їх узгоджене функціонування, тобто показники стану входів і виходів елементів, також можуть змінюватися з часом або ж залишатися незмінними. Зв'язки, що не змінюються, називаються **жорсткими**, **постійними** або **статичними**. Для набору показників a_i, що характеризують жорсткі зв'язки, функція, що визначає лінію поведінки системи, постійна:

$$A = f(t) = (a_1(t), a_2(t), a_3(t), \dots, a_n(t), \dots, a_N(t))$$
$$A = f(t) = [a_1(t),\ a_2(t),\ a_3(t),\ \dots,\ a_n(t),\ \dots,\ a_N(t)] = \text{const}, \qquad (1.2)$$

а її перша похідна дорівнює нулю:

$$A' = f'(t) = [a_1'(t),\ a_2'(t),\ a_3'(t),\ \dots,\ a_n'(t),\ \dots,\ a_N(t)] = 0. \qquad (1.3)$$

Зв'язки, що змінюються з часом, між елементами називаються **гнучкими**, змінними або **динамічними**. Для таких зв'язків функція траєкторії руху системи в часі не є постійною, а її похідна не рівна нулю.

1.6. Соціально-економічна система

Соціально-економічна система - це цілісна сукупність взаємозв'язаних і взаємодіючих соціальних і економічних інститутів (суб'єктів) і відносин з приводу розподілу і споживання матеріальних і нематеріальних ресурсів, виробництва, розподілу, обміну і споживання товарів і послуг.

Соціально-економічну систему, як, втім, і всяку іншу, характеризують системні якості. У їх ряду можна відзначити особливі економічні відносини, які зв'язують єдністю походження всі останні, з якого потім розвиваються все

більш складні відносини. Воно є найпростішим для даних умов способом розподілу ресурсів і підтримки пропорцій.

Соціально-економічна система неминуче локалізована в економічному часі і просторі, а також по відношенню до її альтернативних варіантів. Вона має певні історичні, географічні, етнічні, духовні, політичні і економічні межі. Це у свою чергу означає, що вона може втілюватися в конкретних державно-політичних утвореннях або у формі інших, менших по масштабу, суспільно-господарських організацій. У міру посилення ефекту глобалізації як соціально-економічну систему правомірно розглядати все людство. Цим обумовлюється історичність дослідження: будь-яка система, що вивчається, з одного боку, неминуче історично обумовлена, а, з іншою, історично обумовлені всі категорії і закони цієї системи.

Не всі риси даної системи виникають одночасно, а спочатку розвиваються прості соціальні і економічні форми, а на їх основі все більш і більш складні.

Складність економічної системи полягає передусім у тому, що зміна структури, зв'язків та поводження довільного економічного суб'єкта впливає на решту економічних суб'єктів і спричиняє зміну системи в цілому. Водночас будь-яка зміна в системі на макрорівні позначається на структурі, зв'язках та поводженні економічних суб'єктів.

Ще однією ознакою складності економічної системи є наявність великої кількості як прямих, так і зворотних зв'язків (матеріальних, інформаційних) між її елементами та підсистемами.

Основні властивості, притаманні соціально-економічним системам, які необхідно враховувати під час їх дослідження:

1.Цілісність, яка означає, що зміна будь-якого компоненту системи впливає на її інші компоненти і приводить до зміни системи в цілому. Таке явище можна, наприклад, прослідкувати у разі діалектичної взаємодії продуктивних сил і виробничих відносин, коли при зміні засобів виробництва міняються відповідно виробничі відносини і система в цілому. Тобто, ми в

даному випадку маємо справу з взаємозалежністю компонентів економічної системи.

2. Ієрархічність. Це означає, що кожна система може бути розглянута як елемент вищого порядку. Наприклад, економіка України, як перехідна, може бути розглянута як один з елементів системи-світу.

3. Інтегративність (емерджентність), яка припускає, що система в цілому володіє властивостями, відсутніми у її елементів (наприклад, розподіл праці, який можливий тільки за наявності деякої кількості виробників). Вірно і зворотне, тобто, елементи можуть володіти властивостями, які не властиві системі в цілому.

4. Динамічність економічних процесів, що полягає у зміні параметрів та структури економічних систем під впливом зовнішніх і внутрішніх факторів.

5. Стохастичний характер економічних явищ, з огляду на який для їх опису застосовуються статистичні методи дослідження, а це означає, що поводження економічних систем не піддається точному детальному опису та прогнозуванню.

6. Закономірності економічних процесів можна виявити тільки на підставі достатньої кількості спостережень.

7. Здатність до самоорганізації, тобто здатність без регулюючого впливу покращувати свої показники.

8. Її підсистеми не мають чітких меж: один і той самий елемент (економічний суб’єкт) може одночасно брати участь у різних процесах функціонування економіки, може бути елементом багатьох її підсистем.

9. Економічні процеси не можна ізолювати від зовнішнього середовища та спостерігати їх у «чистому» вигляді.

При аналізі і розробці кібернетичних соціально-економічних систем потрібно одразу визначити певну класифікацію підсистем, аналіз структури і зв’язків яких дозволить точніше їх описати і створити.

Для прикладу табл. 1.1 представлена ієрархія і деталізація управління виробництвом, розбита на підсистеми.

Таблиця 1.1

Зразковий перелік об'єктів управління і типових завдань економіко-організаційного управління підприємством

Індекс підсистеми	Підсистема	Супутні процеси об'єктів управління і функціональних завдань
А	Основні фонди	1. Придбання, надходження, зведення. 2. Стан нових, відновлених, готівки. 3. Наявність. 4. Розподіл і внутрішні переміщення. 5. Використання і завантаження. 6. Вибуття. 7. Знос, амортизація і плата за фонди. 8. Поточний ремонт. 9. Капітальний ремонт
В	Оборотні фонди	1. Надходження, придбання, освіта. 2. Якість тих, що поступили, готівки відновлених. 3. Наявність. 4. Розподіл і внутрішні переміщення. 5. Використання. 6. Вибуття. 7. Відновлення
С	Трудові ресурси	1. Надходження. 2. Атестація тих, що поступили, навчилися і є в наявності. 3. Розподіл, внутрішні переміщення, використання трудових ресурсів. 4. Використання робочого часу. 5. Використання фонду зарплати. 6. Вибуття. 7.Навчання і підвищення кваліфікації
І	Інтелек-туальні ресурси	Виконання сторонніми організаціями для виробничого підприємства: 1) наукові досліджень: 2) спеціального математичного забезпечення; 3) конструкторських розробок; 4) технологічних розробок; 5) нормативних даних
ЕІ	Відвантаження і реалізація продукції	1. Маркетинг. 2. Продукція, що підлягає відвантаженню. 3. Відвантаження і постачання готовій продукції. 4. Реалізація продукції. 5. Відвантаження і реалізація неліквідів (надлишків основних і оборотних фондів)

Індекс підсистеми	Підсистема	Супутні процеси об'єктів управління і функціональних завдань
E2	Фінансова діяльність	Наявність і використання: 1) держбюджетні асигнування; 2) платежі до держбюджету; 3) кредити банку; 4) розрахунки за кредити банка; 5) надходження грошей на рахунок підприємства в банці; 6) витрати з рахунку підприємства в банці; 7) розрахунок за відвантажену продукцію; 8) розрахунок за надані послуги; 9) розрахунок
Dl	Виробнича діяльність по випуску	1. Запуск-випуск товарної продукції (основною і допоміжною). 2. Якість продукції, сортність.
D2	Виробнича діяльність обслуговуючого типу	1. Виконання сервісних послуг своїм підрозділам. 2. Поточна конструкторська підготовка виробництва. 3. Поточна технологічна підготовка виробництва. 4. Розрахунково-обчислювальні роботи. 5. Оперативне коректування норм і нормативів. 6. Функціонування інформаційно-довідкової служби.
D3	Виробнича діяльність по виконанню інших операцій	1. Будівництво. 2. Послуги стороннім організаціям. 3. Конструкторські розробки 4. Технологічні розробки для випуску нової продукції. 5. Розробка нових форм. 6. Науково-дослідні роботи.
P	Розвиток виробництва	1. Впровадження нової техніки 2. Впровадження нових матеріалів. 3. Організаційно-технічні заходи. 4. Зростання продуктивності купа і НОТ. 5. Формування і розподіл: 1) повернення і собівартість: 2) балансовий прибуток і рентабельність; 3) інтегральні показники діяльності підприємства; 4) фонди економічного

У табл. 1.2. наведено категорії об’єктів систем життєзабезпечення, розподілені за ярусами об’єктів управління.

Таблиця 1.2

Компоненти тіла об'єктів систем життєзабезпечення A, B, C

Перший ярус тіла об'єкту управління	Компоненти другого ярусу тіла об'єкту управління								
A - основні фонди	Будівлі	Споруди	Машини і устаткування	Передаточні засоби	Транспортні засоби	Інструменти	Виробничий інвентар	Господарський інвентар	Інші основні фонди
B - оборотні фонди	Сировина, основні матеріали, напівфабрикати	Допоміжні матеріали	Паливо, електроенергія	Тара і упаковка	Запасні частини	Незавершене виробництво	Малоцінні предмети і ті, що швидко зношуються		
C - трудові ресурси	Кадри	Зарплата основна	Зарплата додаткова	Календарний час	-	-	-	-	-

1.7. Гетеросткедастичність

Гетероскедастичність (англ. heteroskedasticity) – це статистичне явище, коли дисперсія помилок (рештків) моделі змінюється залежно від значень однієї або кількох змінних моделі. Це означає, що розкид помилок залежить від рівня або величини прогнозованих змінних.

У регресійних моделях гетероскедастичність може виникати внаслідок різних причин, таких як недосконаліть моделі, систематичні зміни у залежності від значень змінних, наявність аутлаєрів або неврахування інших впливів, що впливають на дисперсію.

Наявність гетероскедастичності може негативно впливати на оцінки параметрів моделі. Зокрема, оцінки стандартних помилок можуть бути

неточними, що призводить до некоректних статистичних висновків та невірної інтерпретації статистичних тестів.

Для виявлення гетероскедастичності застосовують різні методи, такі як графічний аналіз залишків, тести на гетероскедастичність (наприклад, тест Бройша-Паґана, тест Гольдфельда-Квандта, тест Вайта, тест Глейзера, метод Ейткената інші) або застосування спеціальних методів оцінювання, які враховують гетероскедастичність, наприклад, метод найменших квадратів з вагами.

Якщо гетероскедастичність виявлена, можуть бути застосовані корекційні методи, які дозволяють отримати більш надійні оцінки параметрів моделі. Одним із поширених методів є робочі оцінки (англ. robust standard errors), які дозволяють коригувати оцінки стандартних помилок з урахуванням гетероскедастичності.

Тест Гольдфельда-Квандта (Goldfeld-Quandt test) - це статистичний тест для перевірки гетероскедастичності (нерівномірності дисперсії) у регресійних моделях. Цей тест часто використовується в економетриці, коли розглядаються моделі, що включають одну залежну змінну та одну або більше незалежних змінних.

Ідея тесту Гольдфельда-Квандта полягає в порівнянні дисперсії залишків (решткових) між двома підгрупами спостережень. Припустимо, що наша вибірка розподіляється на дві підгрупи відносно певного критерію (наприклад, величини незалежних змінних). Потім проводиться регресійний аналіз для обох підгруп, і залишки (решткі) розраховуються для кожного з двох підгруп. Тест перевіряє, чи дисперсії залишків однорідні між цими двома підгрупами.

Формулюється нульова гіпотеза (H_0) про гомоскедастичність, тобто рівномірність дисперсії залишків у всій вибірці. Альтернативна гіпотеза (H_1) стверджує, що дисперсія залишків є нерівномірною (гетероскедастичність).

Статистика тесту Гольдфельда-Квандта обчислюється шляхом ранжування підгруп за зростанням змінної, за якою проводиться розбиття, і обчислення дисперсій залишків у двох підгрупах. Потім порівнюється

дисперсія залишків у двох кінцях розташування підгруп за допомогою F-статистики.

Якщо F-статистика перевищує критичне значення на заданому рівні значущості, нульова гіпотеза (H_0) про гомоскедастичність відхиляється, і зроблюється висновок про наявність гетероскедастичності. У протилежному випадку, якщо F-статистика не перевищує критичного значення, немає підстав відкидати нульову гіпотезу, і можна припустити, що дисперсія залишків є рівномірною.

1.8. Індивідуальне завдання № 1
Засвоєння основних понять економетрики

Мета роботи: Набути навичок з розуміння основних понять економетрики.

Порядок виконання:

1. Студенти ознайомлюються з основною термінологією економетрики.
2. Студенти самостійно обирають галузь народного господарства.
3. Студенти створюють есе щодо застосування вивчених термінів в обраній галузі. Тобто описати, яка це система, її властивості, підсистеми компоненти (див. табл. 1.1-1.2).
4. Есе має складатися з титульного листа, змісту, вступу, основної частини, висновків, списку використаних джерел і мати не менше 8 сторінок з ілюстраціями.

Контрольні запитання

1. Охарактеризуйте предмет та методи економетрики.
2. Дайте визначення основних методів економетрики.
3. Чим економетрика відрізняється від статистики?
4. Які головні елементи економетрики?
5. Що таке об’єкт?
6. Подайте поняття факторів та параметрів об’єкта.

7. Поясність умовність поділу на системи і об’єкти.
8. Визначте відміну факторів управління, збурення, перешкод та виходів, як факторів системи.
9. Наведіть класифікацію та властивості систем.
10. Як ви розумієте термін «соціально-економічна система»?
11. Чим характеризуються соціально-економічні системи?
12. У чому полягають емерджентні властивості економіки?
13. Дайте опис соціально-економічних систем різних типів (підприємство, фінансово-кредитні установи, галузі економіки, органи державного управління, економіки в цілому) як кібернетичних систем.
14. Що являє собою гетероскедастичність?

У цьому розділі студенти ознайомилися з основними поняттями економетрики; зрозуміли відмінність і єдність таких понять, як об’єкт та система; вникнули у поняття соціально-економічна система.

Розділ 2.
ІДЕНТИФІКАЦІЯ ДАНИХ

Вивчивши матеріали цього розділу студенти опанують прийоми визначення основних статистичних характеристик та перевірки достатності вибірки.

Будь-який економетричний розрахунок починається з формування таблиці спостережень, в якій вхідні фактори розташовуються у лівих колонках, а вихідні – у правих. Запис чергового спостереження здійснюється в один рядок таблиці водночас усіх факторів. Якщо попередня гіпотеза про зв'язок вхідних і вихідних факторів включає в себе і час, цей фактор записується у першій колонці. Таким чином, загальний вид таблиці спостережень наступний (табл. 2.1). Така таблиця називається вибіркою числових значень факторів соціально-економічної системи.

Таблиця 2.1

Приклад формату таблиці спостережень за факторами соціально-економічної системи

Час або дата спостереження	Вхідні фактори				Вихідні фактори			
	X_1	X_2	…	X_n	Y_1	Y_2	…	Y_m

З таблиці видно, що кількість вхідних факторів – n, а вихідних – m. Саме для таблиць такого виду і буде вестися подальше викладення матеріалу, причому, всі фактори – вхідні і вихідні – будуть позначатися як X_i.

2.1. Статистичний аналіз соціально-економічних систем

Цей аналіз проводиться окремо для кожного фактору полягає у визначенні декількох параметрів, що їх характеризують. Це середнє, дисперсія, стандарт та варіація.

Середнє – це один з найбільш розповсюджених прийомів узагальнень. Правильне розуміння сутності середньої визначає її особливу значимість в умовах ринкової економіки, коли середнє через одиночне і випадкове дозволяє виявити загальне і необхідне, виявити тенденцію закономірностей економічного розвитку.

$$M_X = \frac{1}{N}\sum_{i=1}^{N} X_i, \tag{2.1}$$

де X_i –окреме значення фактору; N – число одиниць сукупності (кількість вимірів цього фактору).

Але середня величина – це абстрактна, узагальнююча характеристика ознаки досліджуваної сукупності, вона не показує будівлі сукупності, що дуже істотно для її пізнання. Середня величина не дає представлення про те, як окремі значення досліджуваної ознаки групуються навколо середньої, чи зосереджені вони поблизу чи значно відхиляються від неї. У деяких випадках окремі значення ознаки близько примикають до середньої арифметичної і мало від неї відрізняються. У таких випадках середня добре описує всю сукупність. В інших, навпаки, окремі значення сукупності далеко знаходяться від середньої, і середня погано описує всю сукупність.

Коливання окремих значень характеризують показники варіації, через яку виявляється більшість статистичних закономірностей. Під ***варіацією*** в статистиці розуміють такі кількісні зміни величини досліджуваної ознаки в межах однорідної сукупності, що обумовлені перехресним впливом дії різних факторів.

Аналіз систематичної варіації дозволяє оцінити ступінь залежності змін у досліджуваній ознаці від визначаючих її факторів. Наприклад, вивчаючи силу і характер варіації у сукупності, можна оцінити, наскільки однорідною є дана сукупність у кількісному, а іноді і якісному відношенні, а отже, наскільки характерною є обчислена середня величина. Ступінь близькості даних окремих одиниць до середнього виміряється низкою абсолютних, середніх і відносних показників. Серед них:

Дисперсія – показник, що характеризує розсіювання значень ознаки щодо його середньої величини

$$D_X = \frac{1}{N-1}\sum_{i=1}^{N} X_i^2 - M_X^2, \qquad (2.2)$$

де X_i – окреме значення ознаки; M_X – середня арифметична ознаки; N – число значень ознаки.

Середнє квадратичне відхилення (або математичний стандарт чи просто стандарт) – це узагальнююча характеристика абсолютних розмірів варіації ознаки в сукупності. Середнє квадратичне відхилення є мірилом надійності середнього значення. Чим менше середнє квадратичне відхилення, тим краще середня арифметична відбиває собою всю вибірку. Середнє квадратичне відхилення – це квадратний корінь з дисперсії.

$$\sigma_X = \sqrt{D_X}\,, \qquad (2.3)$$

де D_X – дисперсія ознаки.

Незважаючи на логічну подібність, дисперсія є більш чутливішою до варіації, а, отже, й частіше застосовуваним показником.

Оскільки числові характеристики випадкової величини ми знаходимо за вибіркою кінцевого розміру, то ми не можемо визначити їх точно, а знаходимо

тільки якусь оцінку, виникає питання, а на скільки ж воно відрізняється від справжнього значення середнього чи дисперсії?

Нехай нас цікавить величина інтервалу ε, на який відхилиться від справжньої оцінки числової характеристики, розраховане за результатами експериментальної вибірки. При цьому ми повинні наперед визначити ймовірність β, значення якої викликало б у нас довіру до цього інтервалу (тобто високу ймовірність – 0.8, 0.9, 0.95...). Цей інтервал так і називається – "довірчим".

Отже нам треба зробити дію, зворотну визначенню ймовірності того, що справжнє значення числової характеристики випадкової величини ($Чх[X]$) буде відрізнятися від його оцінки $O[X]$ не більше ніж на величину ε

$$P(|Чх[X] - O[X]| < \varepsilon) = \beta. \qquad (2.4)$$

Коли буде знайдено ε, то справжнє значення числової характеристики буде знаходитися в межах $O[X] - \varepsilon < Чх[X] < O[X] + \varepsilon$.

Розмір довірчого інтервалу для кожної числової характеристики можна знайти із застосуванням функції Лапласа

– для середнього $\varepsilon_m = \ddot{\sigma}_x \Phi^{-1}(\beta);$ (2.5)

– для дисперсії $\varepsilon_D = \ddot{D}_x \Phi^{-1}(\beta) \sqrt{\frac{0{,}8N+1{,}2}{N(N-1)}},$ (2.6)

де, $\ddot{\sigma}_m = \sqrt{\frac{D_X}{N}}$; $\Phi^{-1}(\beta)$– зворотне значення функції Лапласа, тобто таке значення аргументу (квантиля – t), при якому функція Лапласа дорівнює β.

Функція Лапласа має вигляд $\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$, (2.7)

але цей інтеграл у явному вигляді взяти неможливо, тому використовують таблиці його значень або функцію Excel [illegible].

Для прискорення розрахунків масивів зі значною кількістю факторів, можна скористатися підпрограмою «Описова статистика» (Description statistic), яка знаходиться по меню Excel за шляхом Дані-Data analysis - Description statistic.

Інтерфейс додатку показано на рис. 2.1.

Рис. 2.1. Вікно Descriptive Statistics

Далі представлено переклад позначень Description statistic в різних версіях MS Excel.

Англійська	Українська
Mean	Середнє
Standard Error	Стандартна помилка
Median	Медіана
Mode	Мода
Standard Deviation	Стандартне відхилення
Sample Variance	Дисперсія вибірки
Kurtosis	Ексцес
Skewness	Асиметрія
Range	Розмах вибірки
Minimum	Мінімум
Maximum	Максимум
Sum	Сума
Count	Розмір вибірки

В деяких випадках перед початком статистичного аналізу потрібно виконати нормування чисельних значень нашої вибірки. Воно провадиться за формулою

$$X_i^н = \frac{X_i - M_X}{\sigma_X}. \qquad (2.8)$$

Таке нормування з імовірністю 98% переведе всі значення X у діапазон [-4;+4] з середнім, що дорівнює 0, та стандартом, що дорівнює 1.

Окреслена процедура стандартизації (нормування) даних є необхідною при використанні багатьох багатовимірних статистичних методів, а саме – зниження розмірності простору ознак (факторний, компонентний аналіз), класифікації об'єктів (кластерний аналіз) і ін. Стандартизацію варто використовувати в тому випадку, коли змінні виміряні в шкалах, суттєво розрізняються в величинах (мікрони одиниць – мільярди одиниць).

Якщо в процесі розрахунків за нормованими даними виникає потреба виконати денормування, то потрібна формула

$$X_i = \sigma_X X_i^н + M_X. \qquad (2.9)$$

Мірою відносного відхилення значень випадкової величини відносно оцінки його середнього служить варіація та коефіцієнт варіації

$$var(X) = \frac{\ddot{D}(x)}{\ddot{M}(X)}; \qquad K\,var(X) = \frac{\ddot{\sigma}(x)}{\ddot{M}(X)}. \qquad (2.10)$$

№ п/п	X_1	X_2	X_3	X_4
1	87	0,39	560	2770
2	25	0,82	430	2590
3	67	0,29	270	2870
4	62	0,52	860	1920
5	53	0,54	790	2770

Приклад. Економічний процес було досліджено за 4-ма параметрами. Було отримано 5 точок значень цих параметрів. Провести нормування цих параметрів.

Рішення цієї задачі будемо виконувати у Excel. Розрахуємо середні для кожного параметра із застосуванням функції AVERAGE(), у дужках через двокрапку вкажемо діапазон адрес клітинок, які містять зміни значення першого фактору для всіх 5-ти точок. Далі знаходимо стандарт, використовуючи функцію STDEVA(), де так само подано діапазон клітинок для 1-го фактору. І нарешті, за допомогою формули STANDARDIZE(X; M_X; σ_X) виконуємо нормування. Тут перше число – адреса клітинки, яка має бути нормована, 2-е – адреса клітинки, де є середнє, 3-є – клітинка, де є стандарт.

№ п/п	X_1	X_2	X_3	X_4
1	1,25	-0,61	-0,09	0,48
2	-1,49	1,54	-0,62	0,02
3	0,36	-1,11	-1,27	0,74
4	0,14	0,04	1,13	-1,73
5	-0,26	0,14	0,85	0,48

Вікно функції STANDARDIZE представлене на рис. 2.2.

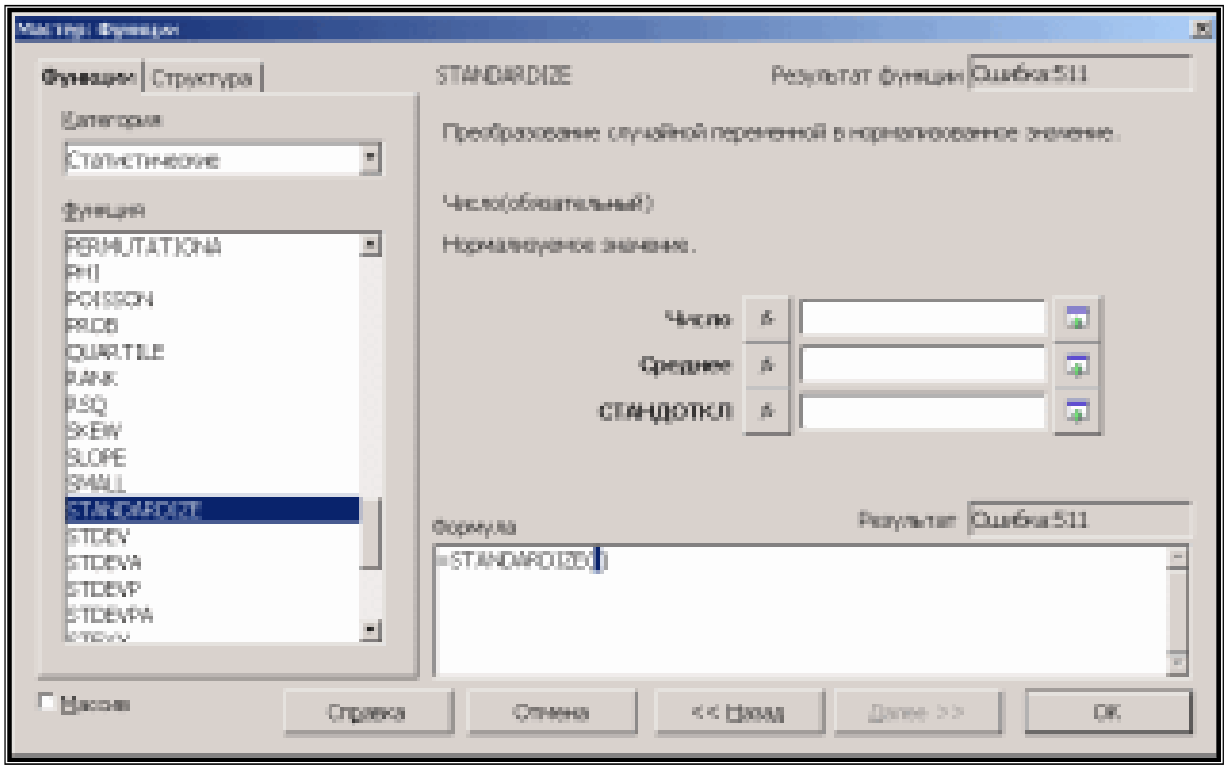


Рис. 2.2. Вікно функції STANDARDIZE з Excel

Знати закон розподілу кожного фактора соціально-економічної системи потрібно, щоб скористатися всіма вже раніше зробленими висновками щодо можливих характеристик цієї випадкової величини.

Для визначення того, якому закону підлягає випадкова величина, необхідно вибрати відомий закон (чи рівномірний, чи експоненціальний, чи нормальний, чи ще який) і висунути так звану «**нуль-гіпотезу**» про те, що математичне сподівання та середнє квадратичне відхилення (чи дисперсія) для цього закону дорівнюють оцінкам цих величин, отриманих з результатів розрахунку за вибіркою випадкової величини (2.1)-(2.10) з певною довірчою ймовірністю p.

Далі, розбиваємо область існування випадкової величини на діапазони з урахуванням $d_{op} = \frac{x_{max}-x_{min}}{1+3.332lnN}$ і знаходимо відносні частоти k_i. Для кожного діапазону знаходимо ймовірності попадання випадкової величини в конкретний діапазон за формулою

$$P(x_i<x<x_{i+1}) = F(x_{i+1}) - F(x_i), \qquad (2.11)$$

де x_i, x_{i+1} – значення випадкової величини на верхній і нижній межах i-го діапазону, $F(x)$ – прийнята нами як нуль-гіпотеза, функція розподілу, у якій параметрами математичного сподівання та дисперсії використовуються їхні оцінки, розраховані з експериментальної вибірки.

Нуль-гіпотеза приймається, якщо критерій узгодження Пірсона (або «хі-квадрат»)

$$\chi_P^2 = n\sum_{i=1}^{d}\frac{(p_i - k_i)^2}{p_i}, \qquad (2.12)$$

буде менший або дорівнювати табличному значенню цього критерію при достатньо великому значенні довірчої ймовірності. Тут n – розмір вибірки, k_i – частоти на відповідних діапазонах; p_i – імовірності попадання випадкової величини в той же діапазон, по якому розраховані і відносні частоти, розраховані

за формулою (2.11): d – загальна кількість діапазонів, на які розбита область існування випадкової величини.

Табличні значення критерію Пірсона $\chi^2(r,\ p)$ можна отримати, скориставшись функцією

ХИ2ОБР(довірча ймовірність;число степенів свободи)

або **CHIINV(довірча ймовірність;число степенів свободи)**

Інколи цю задачу вирішують через визначення рівня довірчої ймовірності. Тобто, за розрахованим значенням χ^2 та за числом степенів свободи знаходять, якій імовірності вони відповідають. А потім приймають рішення, чи можна довіряти отриманим результатам з такою ймовірністю. Для таких розрахунків існує функція

ХИ2РАСП(розраховане значення χ2;число степенів свободи)

або **CHIDIST(розраховане значення χ2;число степенів свободи)**

На рис. 2.2 наведено приклад розрахунку за критерієм Пірсона. Вибірка значень факторів, що досліджується, не наведена.

	A	B	C	D	E	F
7	мін=	9	макс=	43		
8	k=	4	del=	8,5		
9	Діапазони		K	k	P	(k-P)^2 /P
10	мін	мах				
11	9	17,5	3	0,3	0,1866	0,0689
12	17,5	26	1	0,1	0,1391	0,011
13	26	34,5	1	0,1	0,1036	0,0001
14	34,5	43	5	0,5	0,0772	2,3144
15					Xi-кв.=	23,944
16	r=	2				
17	P=	0,95				
18	Xi таб=	0,10259				

Рис. 2.3. Приклад розрахунку

2.2. Кореляційний аналіз факторів соціально-економічних систем

Завданням дисперсійного аналізу є вивчення впливу одного або декількох чинників на дану ознаку. Часто дослідник має у своєму розпорядженні не одну вибірку даних, а декілька. Наприклад, можна визначати зміну у часі валюти балансу підприємств роздрібної торгівлі, та металургійних комбінатів. При побудові математичної моделі потрібно вирішити для себе, чи можна ці вибірки поєднати в одну чи ні? Вирішенням цієї задачі займається декілька додаткових статистичних характеристик.

Основне завдання кореляційного аналізу полягає у виявленні взаємозв'язку між випадковими змінними шляхом оцінки коефіцієнтів кореляції і детермінації, а також перевірки значущості отриманих значень. У економетриці кореляційний аналіз застосовується для відбору факторів, що викликають найбільший вплив на досліджуваний показник і оцінки якості побудованих економетричних моделей. Мірою взаємозв'язку між двома змінними x і y є вибіркова коваріація, що обчислюється за правилом:

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})] \quad (2.13)$$

У практичних розрахунках зазвичай використовується вибірковий парний коефіцієнт парної кореляції, який визначається за наявного набору фактичних даних

Коефіцієнт кореляції – параметр, який характеризує ступінь лінійного взаємозв'язку між двома вибірками, розраховується за формулою

$$r_{xy} = \frac{\sum(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\cdot\sum(y_i-\bar{y})^2}} = \frac{cov(x,y)}{\sigma_x\sigma_y} \quad (2.14)$$

Де σ_x, σ_y – середнє квадратичне відхилення вибірки x та y відповідно.

Коефіцієнт кореляції змінюється від –1 (строго обернена лінійна залежність) до 1 (строго пряма пропорційна залежність). При значенні 0 лінійної залежності між двома вибірками немає. На рис. 2.4 показані можливі кореляційні зв’язки.

а) строга позитивна кореляція; б) сильна позитивна кореляція; в) нульова кореляція; г) помірна негативна кореляція; ґ) строга негативна кореляція; д) нелінійна кореляція

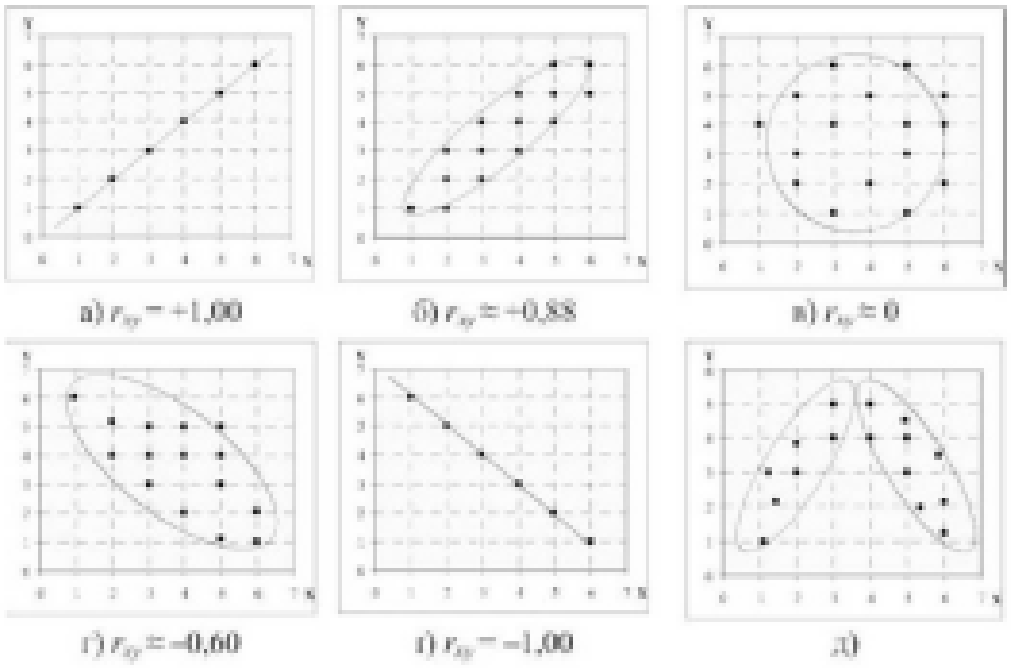


Рис. 2.4. Види кореляційних зв’язків

Незважаючи на те, що задовільного рівня коефіцієнта кореляції по усім факторам може бути не отримано, ми не можемо казати про повну відсутність зв’язку між досліджуваними змінними. Тому виникає необхідність визначення значимості отриманих коефіцієнтів шляхом отримання границь надійного інтервалу, в який потрапляє окремо досліджуваний коефіцієнт.

Для значимих параметрів зв’язку ці границі визначають надійні інтервали з наперед заданою надійністю γ. Для цього використовують z-перетворення Фішера. Перетворення виконується за формулою

$$z = \frac{1}{2} ln \frac{1+r}{1-r}, \tag{2.15}$$

де z – z-перетворення Фішера, r – коефіцієнт кореляції.

Надійний інтервал для z визначається як

$$z - t_\gamma \sqrt{\frac{1}{n-1-3}} \leq z \leq z + t_\gamma \sqrt{\frac{1}{n-1-3}}, \tag{2.16}$$

де n – кількість спостережень, t_γ – квантиль стандартного нормального розподілу, який можна знайти у відповідних статистичних таблицях..

В  є функції, що автоматично розраховують кореляцію

CORREL(адреси клітинок першого масиву даних;

адреси клітинок другого масиву даних)

Якщо потрібно дослідити взаємну кореляцію більше двох факторів, зручно використовувати програму Кореляція пакету аналізу даних.

Для оцінки значущості вибіркового коефіцієнта парної кореляції застосовується t-критерій Стьюдента. При цьому фактичне значення цього критерію визначається за формулою

$$t_{\text{факт}} = \sqrt{\frac{r^2}{1-r^2}(n-2)} \tag{2.17}$$

Отримане значення порівнюється з табличним критичним значенням, яке залежить від рівня значущості α і числа ступенів свободи $\nu = n - 2$.

Критичне значення може бути знайдено за відповідними таблицями, а при використанні табличного процесора Excel – за допомогою функції

СТЬЮДРАСПОБР (α; ν)

або [illegible]

При отриманні $t_{\text{набл}} > t_{\text{кр}}$ значення коефіцієнта кореляції r визнається значимим, тобто між змінними є лінійна кореляційна залежність.

Критерій Стьюдента використовується, щоб визначити, наскільки вірогідно, що дві вибірки узяті з генеральних сукупностей, мають одне і те ж середнє.

Якщо ймовірність невелика (менше 0,55), можна вважати, що вибірки мають істотно відмінні середні, а отже, їх не можна поєднувати в одну для побудови математичної моделі.

Для реалізації цього закону розподілу існує функція T.TEST

T.TEST(Дані 1; Дані 2; Режим; Тип)

Тут: **Дані 1** перший масив даних; **Дані 2** другий масив даних; **Режим** = 1, то функція використовує односторонній розподіл, якщо **Режим** = 2, то двосторонній розподіл; **Тип** є типом t - тесту (для перевірки за критерієм Стьюдента). Тип 1 означає двосторонній. Тип 2 означає дві вибірки, рівну вірогідність. Тип 3 означає дві вибірки, нерівну вірогідність.

Приклад

Визначити відмінність дисперсій двох вибірок, представлених нижче

X_1	0,45	0,59	0,78	0,04	0,44	0,32	0,92
X_2	0,34	0,29	0,88	0,87	0,68	0,28	0,88

Скористаємося функцією F.TEST Результат розрахунку показано на рис. 3.3. Імовірність 0,98 означає, що ці вибірки статистично взяті з однієї генеральної сукупності. Отже, їх можна поєднати в одну для побудови економіко-математичної моделі.

Рис. 2.5. Приклад роботи функції F.TEST 

F-тест (Фішера) визначає односторонню вірогідність того, що дисперсії аргументів массив1 і массив2 розрізняються неістотно. Ця функція використовується для того, щоб визначити, чи мають дві вибірки різні дисперсії. Наприклад, якщо дані результати тестування для приватних і суспільних шкіл, то можна визначити, чи мають ці школи різні рівні різнорідності учнів за наслідками тестування.

Якщо ймовірність буде невеликою (<0,55), це означає, що за дисперсією вибірки відрізняються істотно, а отже, при імітаційному моделюванні не можна використовувати дані з вибірок водночас.

Для реалізації цього тесту існує функція FTECT 

FTECT(масив1;масив2)

Пояснимо критерій Фішера.

Нехай є N нормально розподілених генеральних сукупностей з рівними дисперсіями та, можливо, з різними математичними сподіваннями.

Із кожної сукупності робимо вибірку об'єму $\{n_i\}, i = 1,2 \dots, N$

Тоді $\sum_{i=1}^{n} n_i = n$- об'єм усієї вибірки.

Позначимо j варіант випадкової величини X з i-тої сукупності x_i , Тоді

середня арифметична вибірки із *i*-тої сукупності буде $x_i = \frac{1}{n_i}\sum_{j\ 1}^{n_i} x_i$, а середня усієї вибірки буде $\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i n_i$

При рівні значущості α треба перевірити основну гіпотезу про рівність математичних сподівань сукупностей, що розглядаються

При рівності дисперсій статистична характеристика буде мати розподіл Фішера з N - 1 та n - N степенів свободи. Тому в якості статистичної характеристики для перевірки цієї гіпотези візьмемо функцію

$$F = \frac{\frac{1}{N}\sum_{i\ 1}^{N}(x_i-\bar{x})^2 n_i}{\frac{1}{n\ N}\sum_{i\ 1}^{N}\sum_{j=1}^{n_i}(x_i\ -\bar{x})^2}, \qquad (2.18)$$

Критичну область у цьому випадку знаходять з урахуванням умови $P(F > f_a) = \alpha$, де – f_a критичне значення розподілу Фішера.

Приклад. Є дані про вартість (в тис. гривень) проданих трьох видів виробів певним магазином в окремі дні тижня

вівторок	середа	четвер	п'ятниця	субота
10.2	10.8	10.7	13.0	12.0
11.5	9.8	11.5	13.2	11.5
12.0	12.1	12.0	11.5	11.8

Припускаючи нормальний закон розподілу одержаної суми кожного дня та рівність дисперсій, перевірити гіпотезу $H_0: a_1 = a_2 = a_3 = a_4 = a_5$ при рівні значущості α = 0,05.

Розв'язання. Умови прикладу дозволяють застосувати до розв'язання задачі критерій дисперсійного аналізу.

У цьому випадку маємо: N = 5;n_i = 3, i = 1,2,..., 5; n = 15. Знаходимо

$$x_1 = 11.2, x_2 = 10.8, x_3 = 11.4, x_4 = 12.6, x_5 = 11.8, \bar{x} = 11.6$$

Зробимо обчислення сум

$\sum_{i=1}^{5}(x_i - \bar{x})^2 n_i$ =4.86

$\sum_{i=1}^{5}\sum_{j\ 1}^{3}\left(x_i\ - \bar{x}\right)^2$ =11.96

Тепер за формулою (2.18) знайдемо значення статистичної характеристики

$$F_{cn} = \frac{\frac{1}{5}4.8}{\frac{1}{15-5}11.96}=0,81.$$

Із таблиці критичних значень розподілу Фішера зі степенями вільності N - l = 5-l = 4 та n - N = 15-5 = 10 і рівнем значущості α = 0,05 знаходимо $F_{\text{кр}} = f_{0.05} = 3.48$

Одержали, що $F_{cn} = 0,81 < F_{\text{кр}} = 3.48$ тому гіпотеза H_0 може бути прийнята.

Загальні принципи порівняння вибірок наступний:

1. Розраховуємо кореляцію, t-тест та F-тест для кожної пари вибіркових значень.
2. Вибірки можна об'єднувати в одну якщо кореляцію позитивна, а тести дають значну ймовірність (>0.65).

2.3. Визначення достатності обсягу вибірки

Для визначення вибіркових значень асиметрії та ексцесу застосовують точні розрахункові формули, які аналогічні тим, що є в MS Excel.

Ексцес розраховують як

$$\hat{E} = \left(\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^4\right) - \frac{3(n-1)^2}{(n-2)(n-3)} \qquad (2.19)$$

де σ – середньоквадратичне відхилення, n – кількість спостережень, x_i – і-те значення спостереження, $\bar{x}$ – середнє значення.

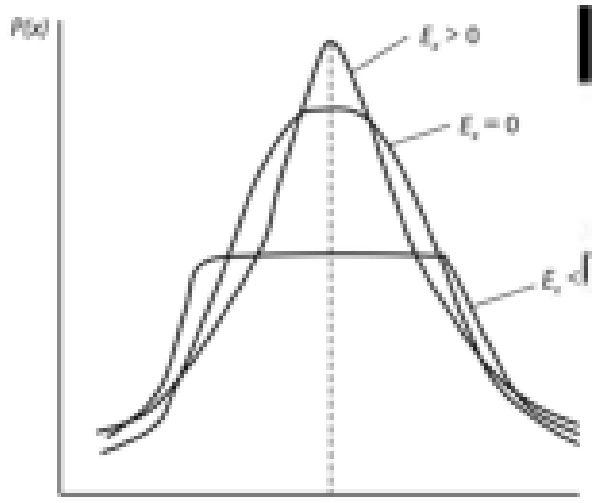


Рис.2.6. Значення ексцесу при різних видах вибірки

Асиметрія визначає зміщення ознаки в сукупності відносно її середньої величини. Додатна асиметрія – це зрушення розподілу у бік позитивних відхилень, від'ємна, у бік негативних.

$$\hat{A} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^3 \qquad (2.20)$$

де σ – середньоквадратичне відхилення, n – кількість спостережень, x_i – і-те значення спостереження, $\bar{x}$ – середнє значення.

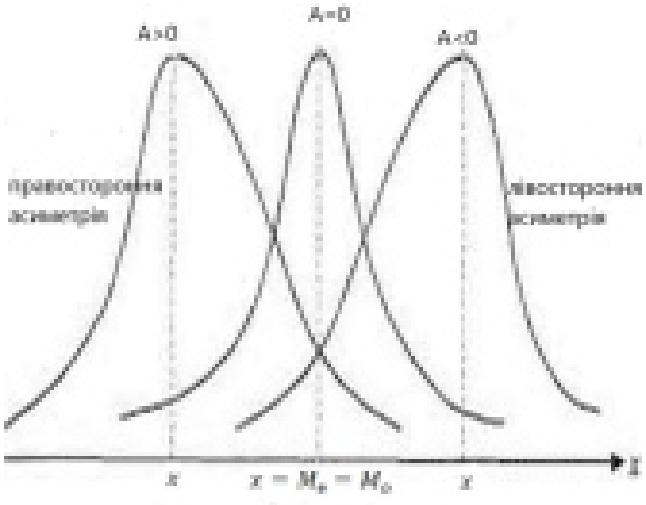


Рис.2.7. Значення асиметрії при різних видах вибірки

Дисперсії ексцесу та асиметрії

$$D_{\hat{E}}^2 = \frac{24n(n-1)^2}{(n+5)(n+3)(n-2)(n-3)}; \quad (2.21)$$

$$D_{\hat{A}}^2 = \frac{6\ (n-1)}{(n+3)(n+1)(n-2)}. \quad (2.22)$$

Критерій згоди для цих показників.

$$\left|\hat{E}\right| \leq 2D_{\hat{E}}\ ; \quad (2.23)$$

$$\left|\hat{A}\right| \leq 2D_{\hat{A}}\ . \quad (2.24)$$

Якщо його дотримано, значить вибірка є достатньою для подальшого використання. Якщо ні, вибірку потрібно збільшити.

Середня помилка вибірки розраховується як

$$S = \sqrt{\frac{\sigma^2}{n}}. \quad (2.25)$$

2.4. Індивідуальне завдання № 2

Засвоєння розрахунків основних статистичних показників

Мета роботи: Набути навичок з уміння розрахувати основні статистичні показники

Порядок виконання:

1. За кожною колонкою основної таблиці розрахувати середнє, розмах, моду, медіану, дисперсію, середньоквадратичне відхилення.

2. Перевірити розрахунки шляхом застосування підпрограми Description statistic (описова статистика).

Дані необхідно згенерувати наступним чином (табл. 2.2).

Таблиця 2.2. Дані для розрахунку

x, тис. грн.	y, тис. грн
100+20*N	70+5*N
130+10*N	90+5*N
150+10*N	105+5*N
160+10*N	95+5*N
175+10*N	130+5*N
180+10*N	120+5*N
185+10*N	140+5*N
200+10*N	150+5*N
215+10*N	150+5*N
220+10*N	145+5*N
240+10*N	165+5*N
250+10*N	172+5*N
260+10*N	155+5*N
270+10*N	143+5*N
280+10*N	167+5*N

де N – номер студента за списком групи.

Методичні рекомендації та приклад

Середні можна розрахувати зі застосуванням функції AVERAGE(), у дужках через двокрапку вказується діапазон клітинок значень.

Далі знаходимо стандарт, використовуючи функцію STDEVA(), де так само подано діапазон клітинок.

Дисперсія розраховується за допомогою функції VARA().

Мода розраховується функцієюMODE.SNGL().

Медіана розраховується функцією MEDIAN()

Розмах вибірки розраховується MAX()-MIN().

У дужках всіх функцій потрібно задавати діапазон значень, для яких шукаються дані характеристики.

Рохгляномо рохрахунок на прикладі. В таблиці 2.3 представлено кількість підприємств за їх розмірами за регіонами у 2019 році (Дані наведено без урахування результатів діяльності банків, бюджетних установ, тимчасово

окупованої території Автономної Республіки Крим, м. Севастополя) та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Таблиця 2.3. Кількість підприємств за їх розмірами за регіонами у 2019 році (Дані наведено без урахування результатів діяльності банків, бюджетних установ, тимчасово окупованої території Автономної Республіки Крим, м. Севастополя) та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Область	Усього, одиниць
Вінницька	10295
Волинська	6292
Дніпропетровська	31191
Донецька	10299
Житомирська	7306
Закарпатська	6388
Запорізька	15652
Івано-Франківська	8595
Київська	21077
Кіровоградська	8755
Луганська	3732
Львівська	20480
Миколаївська	12278
Одеська	25871
Полтавська	11439
Рівненська	5956
Сумська	6222
Тернопільська	5092
Харківська	25051
Херсонська	8511
Хмельницька	7864
Черкаська	9709
Чернівецька	4235
Чернігівська	6443

На рисунку 2.8 фрагмент розрахунку з використанням функцій, а в таблиці 2.4 представлено результат застосування Description statistic.

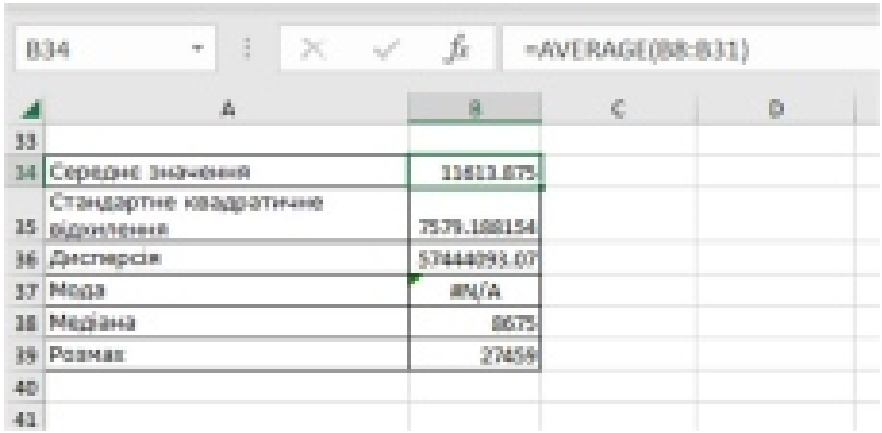


Рис. 2.8. Фрагмент розрахунку

Таблиця 2.4 - Description statistic

Column1	
Mean	11613.88
Standard Error	1547.095
Median	8675
Mode	#N/A
Standard Deviation	7579.188
Sample Variance	57444093
Kurtosis	0.823041
Skewness	1.337206
Range	27459
Minimum	3732
Maximum	31191
Sum	278733
Count	24

Звіт по практичній роботі повинний містити:

1) Вихідні дані.
2) Розрахунки двома методами.
3) Висновки.

2.5. Індивідуальне завдання № 3

Засвоєння розрахунків ексцесу, асиметрії та дисперсії

Мета роботи: Набути навичок з уміння розрахувати ексцесу, асиметрії та дисперсії

Порядок виконання:

1. Розрахувати для кожної колонки основної таблиці ексцес, асиметрію та дисперсію.
2. Знайти середню помилку вибірки.
3. Розрахувати критерій згоди для асиметрії та ексцесу.
4. Перевірити їх за критерієм згоди Стьюдента для довірчої ймовірності 0,95.
5. В разі не відповідності критеріїв згоди, збільшити вибірку 5 точок і повторити розрахунки.

Методичні рекомендації та приклад

Розрахуємо на прикладі даних з другого завдання (див. табл.2.3).

Для розрахунку ексцесу та асиметрії зручно розрахувати додаткові стовпчики $\left(\frac{x_i-\bar{x}}{\sigma}\right)^4$ і $\left(\frac{x_i-\bar{x}}{\sigma}\right)^3$, де $\bar{x}$ – середнє, σ – середньоквадратичне відхилення.

Використовуючи формули 2.19 та 2.20 отримаємо $\hat{E} = 0{,}823$ та $\hat{A} = 1{,}337$. Етапи розрахунку представлено в таблиці 2.5.

Якщо використовувати Description statistic, то отримаємо ті ж самі значення.

Тепер перевіримо критерій згоди цих показників за формулами 2.21.-2.22.

$$D_{\hat{E}}^2 = \frac{24 * 24 * (24-1)^2}{(24+5)(24+3)(24-2)(24-3)} = 0.8423$$

$$D_{\hat{A}}^2 = \frac{6 * 24 * (24-1)}{(24+3)(24+1)(24-2)} = 0.223$$

Далі застосуємо формули 2.23-2.24.

*2*$D_{\hat{E}}$ =2*(0,8423)^0.5=1.83*

Таким чином 0.8423<1.83 і критерій згоди для ексцесу виконується.

$2*D_{\hat{A}} = 2*0{,}223\text{^}0.5 = 0.9445$

Таким чином 1,337>0.9445 і критерій згоди для асиметрії не виконується.

Таблиця 2.5 - Розрахунок

	Усього, одиниць	$\left(\frac{x_i-\bar{x}}{\sigma}\right)^4$	$\left(\frac{x_i-\bar{x}}{\sigma}\right)^3$
Вінницька	10295	0.001	-0.005
Волинська	6292	0.243	-0.346
Дніпропетровська	31191	44.515	17.234
Донецька	10299	0.001	-0.005
Житомирська	7306	0.104	-0.184
Закарпатська	6388	0.226	-0.328
Запорізька	15652	0.081	0.151
Івано-Франківська	8595	0.025	-0.063
Київська	21077	2.430	1.946
Кіровоградська	8755	0.020	-0.054
Луганська	3732	1.170	-1.125
Львівська	20480	1.873	1.601
Миколаївська	12278	0.000	0.001
Одеська	25871	12.521	6.656
Полтавська	11439	0.000	0.000
Рівненська	5956	0.311	-0.416
Сумська	6222	0.256	-0.360
Тернопільська	5092	0.548	-0.637
Харківська	25051	9.879	5.573
Херсонська	8511	0.028	-0.069
Хмельницька	7864	0.060	-0.121
Черкаська	9709	0.004	-0.016
Чернівецька	4235	0.898	-0.923
Чернігівська	6443	0.217	-0.318
	сума	75.411	28.193

Середня помилка вибірки розраховується як $S = \sqrt{\frac{\sigma^2}{n}} = 1547.095$

2.6. Індивідуальне завдання № 4

Засвоєння методики кореляційного аналізу

Мета роботи: Набути навичок з уміння розрахувати коефіцієнти парної кореляції

Порядок виконання:

1. Розрахувати коваріацію та коефіцієнт кореляції для всіх пар параметрів з основної таблиці.
2. Перевірити значимість коефіцієнтів за критерієм Стьюдента.
3. В разі недотримання критерію, збільшити вибірку на 5 точок.
4. Визначити статистичну достовірність коефіцієнтів кореляції.

Таблиця 2.6. Вихідні дані

x, тис. грн.	y, тис. грн	z, тис. грн
100+20*N	70+5*N	75+15*N
130+10*N	90+5*N	80+15*N
150+10*N	105+5*N	115+15*N
160+10*N	95+5*N	100+15*N
175+10*N	130+5*N	120+15*N
180+10*N	120+5*N	110+15*N
185+10*N	140+5*N	100+15*N
200+10*N	150+5*N	140+15*N
215+10*N	150+5*N	160+15*N
220+10*N	145+5*N	150+15*N
240+10*N	165+5*N	185+15*N
250+10*N	172+5*N	152+15*N
260+10*N	155+5*N	110+15*N
270+10*N	143+5*N	120+15*N
280+10*N	167+5*N	140+15*N

Методичні рекомендації та приклад

Спочатку згенеровано дані в таблиці 2.7

Таблиця 2.7. Дані

Розмір вибірки	x, тис. грн.	y, тис. грн	z, тис. грн
1	400	170	375
2	330	190	380
3	350	205	415
4	360	195	400
5	375	230	420
6	380	220	410
7	390	250	400
8	400	255	440
9	415	260	460
10	420	245	450
11	440	265	475
12	450	272	452
13	460	255	410
14	470	243	420
15	480	267	435

За допомогою MS Excel Data Analysis розраховано коваріаційну та кореляційну матриці, таблиці 2.8 та 2.9 відповідно.

Таблиці 2.8. Коваріаційна матриця

	x, тис. грн.	y, тис. грн	z, тис. грн
x, тис. грн.	1,919.33		
y, тис. грн	966.27	940.43	
z, тис. грн	634.27	689.83	780.43

Таблиці 2.9. Кореляційна матриця

	x, тис. грн.	y, тис. грн	z, тис. грн
x, тис. грн.	1		
y, тис. грн	0.71921598	1	
z, тис. грн	0.51823965	0.80521427	1

Далі розрахуємо фактичне значення критерію Стьюдента $t_{\text{факт}} = \sqrt{\frac{r^2}{1-r^2}(n-2)}$, що подано в таблиці 2.10.

Таблиця 2.10. Фактичне значення критерію Стьюдента

	X Y	X Z	Y Z
t факт	3.73232616	2.18482481	4.89598621

Критичне значення критерія Стьюдента на рівні значущості 0.05 та 13 ступенів свободи дорівнює 1.7709.

Оскільки всі фактичні значення критеріїв більше, ніж критичне значення, то парні коефіцієнти кореляції значущі на рівні значущості 0.05.

Визначемо надійні інтервали з наперед заданою надійністю γ. Для цього використовують z-перетворення Фішера. Перетворення виконується за формулою

$$z = \frac{1}{2} ln \frac{1+r}{1-r},$$

де z – z-перетворення Фішера, r – коефіцієнт кореляції.

Надійний інтервал для z визначається як

$$z - t_\gamma \sqrt{\frac{1}{n-1-3}} \leq z \leq z + t_\gamma \sqrt{\frac{1}{n-1-3}},$$

де n – кількість спостережень, t_γ – квантиль стандартного нормального розподілу, який можна знайти у відповідних статистичних таблицях.

Візьмемо t_γ=1.96 – верхній двосторонній 5%-й квантиль нормального розподілу. Результати подано в таблиці 2.11.

Таблиця 2.11. Розрахунок надійних інтервалів

Зв’язок між	X Y	X Z	Y Z
z	0.91	0.57	1.11
$t_\gamma\sqrt{\frac{1}{n-1-3}}$	0.59	0.59	0.59
$z - t_\gamma\sqrt{\frac{1}{n-1-3}}$	0.32	-0.02	0.52
$z + t_\gamma\sqrt{\frac{1}{n-1-3}}$	1.50	1.16	1.70

Контрольні запитання

1. Що являє собою середнє?
2. Який показник є мірилом надійності середнього значення?
3. Що показують коефіцієнти кореляції та коваріації?
4. Що показує асиметрія?
5. Про що свідчить значення ексцесу?
6. Яким чином перевірити значущість коефіцієнтів кореляції?
7. З якою метою проводиться тест Фішера?
8. Яким чином використовують критерій Пірсона?
9. Що являє собою «довірчий інтервал»?
10. Для чого проводиться стандартизація даних?

У цьому розділі студенти навчилися визначати основні статистичні характеристики вибірки та достатність вибірки.

Розділ 3.
ОДНОФАКТОРНІ МОДЕЛІ

Вивчивши матеріали цього розділу студенти опанують методику знайдення коефіцієнтів однофакторних моделей та визначення статистичної значимості цих коефіцієнтів.

3.1. Метод найменших квадратів»

Всі соціально-економічні явища взаємозв'язані. Зв'язок між ними має причинно-наслідковий характер. Ознаки, що характеризують причини і умови зв'язку, називаються ***факторними x,*** а ознаки, які характеризують наслідки зв'язку, — ***результативними y.*** Між ознаками ***x*** і ***y*** виникають різні за природою і характером зв'язки, а саме: функціональні і стохастичні. При ***функціональному зв'язку*** кожному значенню ознаки ***x*** відповідає одне певне значення ***y.*** Цей зв'язок виявляється однозначно у кожному окремому випадку. При ***стохастичному зв'язку*** кожному значенню ознаки ***x*** відповідає певна безліч значень ***y,*** створюючих так званий ***умовний розподіл.*** Як закон цей зв'язок виявляється тільки в масі випадків і характеризується зміною умовних розподілів ***y.*** Якщо замінити умовний розподіл середньою величиною $\bar{y}$, то утворюється різновид стохастичного зв'язку – ***кореляційний.*** У разі кореляційного зв'язку кожному значенню ознаки ***x*** відповідає середнє значення результативної ознаки $\bar{y}$.

Прикладом стохастичною, і зокрема кореляційною, зв'язку є розподіл проданих однокімнатних квартир за їх вартістю ***y*** і розміру загальної площі ***x*** (табл. 3.1).

Таблиця 3.1. Розподіл проданих однокімнатних квартир за їх вартістю у і розміру загальної площі х

Розмір загальній площі м2 ***x***	Кількість квартир з вартістю тис. ум. од.						Середня вартість квартири тис. ум. од. $\bar{y}_j$
	9—11	11—13	13—15	15—17	17—19	Разом *fi*	
До 25	26	12	2	--	--	40	10,8
25-30	4	9	12	5	--	30	13,2
30—35	--	4	6	10	4	24	15,2
35 і більш	--	--	--	--	6	6	18,0
В цілому	30	25	20	15	10	100	13,0

Більшість економічних явищ мають імовірнісний, або випадковий, або стохастичний характер, тобто, в реальних умовах дуже важко з 100-процентною упевненістю передбачати розвиток того або іншого економічного об'єкту. Природно, це факт приводить до деяких ускладнень, і вимагає наявність у управлінського персоналу навиків в області прогнозу невизначеного майбутнього – навиків виявлення *закономірностей, прихованих серед випадковостей минулого і сьогодення періодів, і перенесенню цих закономірностей в (на) майбутнє – навиків* ***екстраполювання.***

Як правило, дослідження випадкових взаємин зводиться до адаптації реальних імовірнісних взаємозв'язків до логіки функціональних залежностей, тобто іншими словами, потрібно визначити і аналітично виразити («проявити») форму передбачуваної залежності і потім досліджувати її.

Сучасна наука ще не знайшла кращого «проявника» кореляційних залежностей, чим **метод найменших квадратів**.

Суть методу найменших квадратів полягає *у відшуканні параметрів моделі тренда, яка краще всього описує тенденцію розвитку якого-небудь випадкового явища в часі або в просторі (тренд – це лінія, яка і характеризує тенденцію цього розвитку). Завдання методу найменших квадратів (скорочено - МНК) зводиться до знаходження не просто якоїсь моделі тренда, а до знаходження кращої або оптимальної моделі. У свою чергу, модель буде оптимальною, якщо сума квадратичних відхилень між спостережуваними*

фактичними величинами і відповідними їм розрахунковими величинами тренда буде мінімальною (найменшою):

$\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \rightarrow m\ n,$

де $(Y_i - \widehat{Y}_i)^2$ *- квадратичне відхилення між спостережуваною фактичною величиною і відповідним їй розрахунковою величиною тренда*

Y_i *- фактичне (спостережуване) значення явища, що вивчається*

$\widehat{Y}_i$ *- розрахункове значення моделі тренда*

n *- число спостережень за явищем, що вивчається .*

МНК самостійно застосовується досить рідко. Як правило, частіше за все його використовують лише як необхідний технічний прийом при кореляційно-регресійних дослідженнях.

Інструментарій МНК зводиться до наступних процедур**:**

Перша процедура. З'ясовується, чи існує взагалі яка-небудь тенденція зміни результативної ознаки при зміні вибраного чинника-аргументу, або іншими словами, чи є зв'язок між «у» і «х».

Друга процедура. Визначається, яка лінія (траєкторія) здатна краще всього описати або охарактеризувати цю тенденцію.

Третя процедура. Розраховуються параметри регресійного рівняння, що характеризує дану лінію, або іншими словами, визначається аналітична формула, що описує кращу модель тренда.

УВАГА!!! Слід пам'ятати, що інформаційною основою МНК може бути тільки достовірний статистичний ряд, причому число спостережень не повинне бути менше 4-х, інакше, що згладжують процедури МНК можуть втратити здоровий глузд.

Приклад. Допустимо, ми маємо інформацію про середню врожайність соняшнику по досліджуваному господарству (див. табл. 3.2).

Таблиця 3.2. Таблиця початкових даних

Номер спосте-реження	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Роки	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Врожайність, ц/га	14,2	15,6	17,5	14,5	15,3	17,0	16,6	17,5	15,0	17,7

Припущення (гіпотеза): «Оскільки рівень технології при виробництві соняшнику в наший країні за останніх 10 років практично не змінився, значить, видно, коливання врожайності в аналізований період дуже сильно залежали від коливання погодно-кліматичних умов. Чи дійсно це так?»

Перша процедура МНК: *Перевіряється гіпотеза про існування тенденції зміни врожайності соняшнику залежно від зміни погодно-кліматичних умов за аналізованих 10 років.*

У даному прикладі за «у» доцільно прийняти врожайність соняшнику, а за «х» - номер року, який спостерігається в аналізованому періоді. Перевірку гіпотези про існування якого-небудь взаємозв'язку між «х» і «у» можна виконати двома способами: вручну і за допомогою комп'ютерних програм типу «Statistica» і тому подібне. Звичайно, при наявність комп'ютерної техніки дана проблема вирішується сама собою. Але, щоб краще зрозуміти інструментарій МНК доцільно виконати перевірку гіпотези про існування зв'язку між «х» і «у» уручну, коли під рукою знаходяться тільки ручка і звичайний калькулятор. У таких випадках гіпотезу про існування тенденції краще всього перевірити візуальним способом по розташуванню графічного зображення аналізованого ряду динаміки - кореляційного поля:

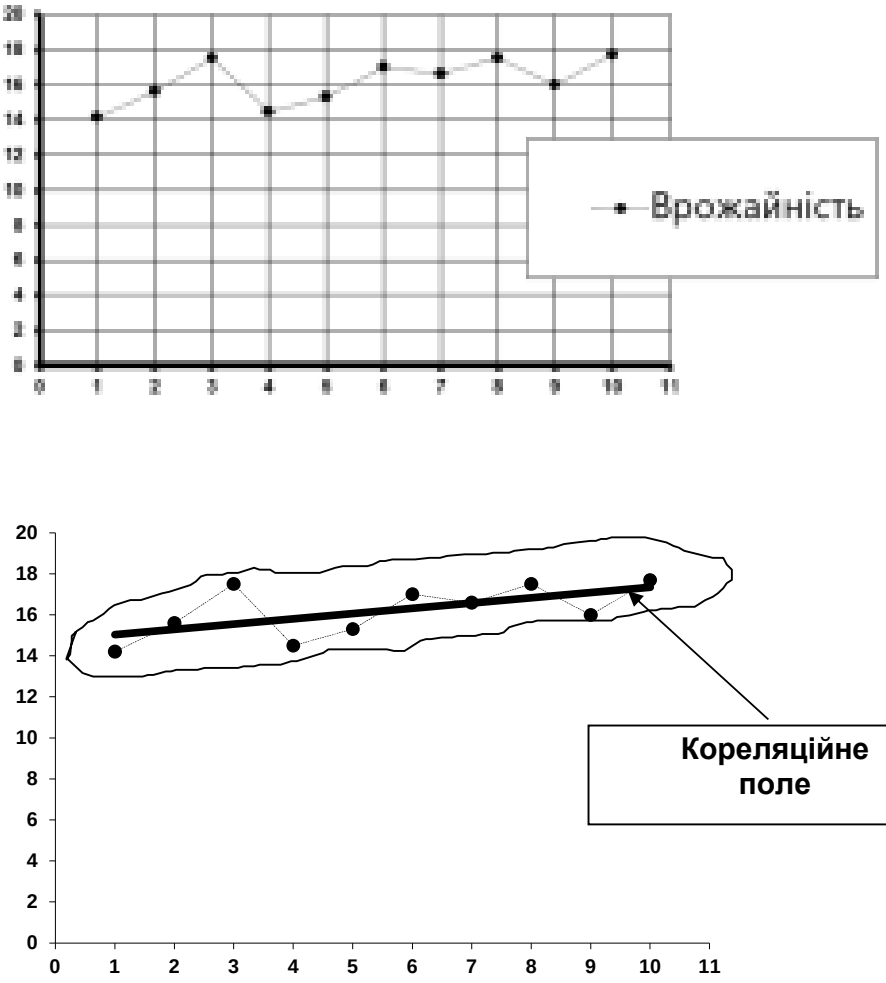


Рис.3.1. Кореляційні поля

Кореляційне поле в нашому прикладі розташоване навколо поволі зростаючій лінії. Це вже само по собі говорить про існування певної тенденції в зміні врожайності соняшнику. *Не можна говорити про наявність якої-небудь тенденції лише тоді, коли кореляційне поле схоже на круг, коло, строго вертикальне або строго горизонтальна хмара, або ж складається з хаотично розкиданих крапок.* У решті всіх випадків слід підтвердити гіпотезу про існування взаємозв'язку між «х» і «у», і продовжити дослідження.

Друга процедура МНК: *Визначається, яка лінія (траєкторія) здатна краще всього описати або охарактеризувати тенденцію зміни врожайності соняшнику за аналізований період.*

За наявності комп'ютерної техніки підбір оптимального тренда відбувається автоматично. При «ручній» обробці вибір оптимальної функції здійснюється, як правило, візуальним способом – по розташуванню кореляційного поля. Тобто, по вигляду графіка підбирається рівняння лінії, яка краще всього підходить до емпіричного тренда (до фактичної траєкторії).

Як відомо, в природі існує величезна різноманітність функціональних залежностей, тому візуальним способом проаналізувати навіть незначну їх частину – вкрай скрутно. На щастя, в реальній економічній практиці більшість взаємозв'язків достатня точно можуть бути описані або параболою, або гіперболою, або ж прямою лінією. У зв'язку з цим, при «ручному» варіанті підбору кращої функції, можна обмежитися тільки цими трьома моделями.

■ пряма $Y = a + b$ виглядає таким чином:

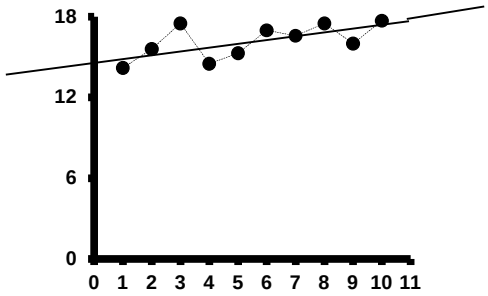


Рис.3.2. Лінійний тренд

■ гіпербола $Y = a + b/X$ має наступний вигляд:

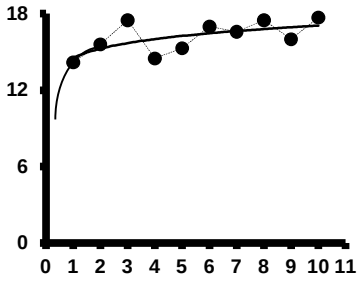


Рис.3.3. Гіпербола

■ парабола другого порядку $Y = a + b \quad + c \quad ^2$виглядає так:

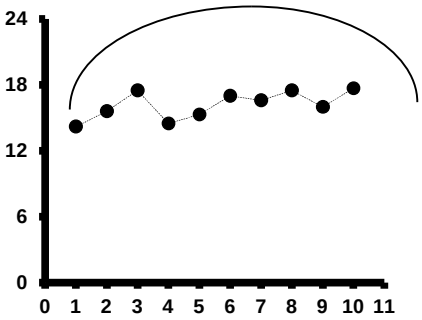


Рис.3.4. Парабола другого порядку

Неважко відмітити, що в нашому прикладі краще всього тенденцію зміни врожайності соняшнику за аналізованих 10 років характеризує пряма лінія, тому рівнянням регресії буде рівняння прямої

$$\hat{Y} = a + b \times X.$$

Третя процедура МНК. *Розраховуються параметри регресійного рівняння, що характеризує дану лінію, або іншими словами, визначається аналітична формула, що описує кращу модель тренда.*

Знаходження значень параметрів рівняння регресії, в нашому випадку параметрів a і b є серцевиною МНК. Даний процес зводиться до вирішення системи нормальних рівнянь. У свою чергу, будь-яка система нормальних рівнянь для всіх видів залежностей будується за наступною схемою:

По-перше, визначаються коефіцієнти при невідомих параметрах у виявленому рівнянні прямої лінії $Y = a + b \cdot X$. Коефіцієнт при параметрі a рівний $\boldsymbol{1}$, а при параметрі b рівний X.

По-друге, рівняння прямої по черзі, спочатку помножується на коефіцієнт при параметрі a, а потім на коефіцієнт при параметрі b. Таким чином, виходить два рівняння: перше – $Y = a + b \cdot X$ і друге – $Y \cdot X = a \cdot X + b \cdot X^2$.

По-третє, в перше рівняння з таблиці початкових даних попарно, у відповідність з номером спостереження, підставляються абсолютно всі значення X і Y. Виходить 10-ть рівнянь типу: $14{,}2 = a + b \cdot 1$; $15{,}6 = a + b \cdot 2$; $17{,}5 = a + b \cdot 3$ і так далі до десятого – $17{,}7 = a + b \cdot 10$. Після чого, всі ці рівняння складаються, і виходить одне сумарне нормальне рівняння:

$$\sum Y = a \cdot n + b \cdot \sum X$$

Потім, в друге рівняння з таблиці початкових даних точно також, у відповідність з номером спостереження, підставляються абсолютно всі значення X і Y. Виходить знову таки 10-ть рівнянь: $14{,}2 \cdot 1 = a \cdot 1 + b \cdot 1^2$; $15{,}6 \cdot 2 = a \cdot 2 + b \cdot 2^2$; $17{,}5 \cdot 3 = a \cdot 3 + b \cdot 3^2$ і так далі до десятого – $17{,}7 \cdot 10 = a \cdot 10 + b \cdot 10^2$. Після чого, всі ці рівняння складаються, і виходить ще одне нормальне рівняння: $\sum YX = a \cdot \sum X + b \cdot \sum X^2$

По-четверте, створюється система, що складається з двох нормальних рівнянь:

$$\begin{cases} \sum Y = a \cdot n + b \cdot \sum X \\ \sum YX = a \cdot \sum X + b \cdot \sum X^2, \end{cases}$$

яка досить легко вирішується методом Гауса. Нагадаємо, що в результаті рішення, в нашому прикладі, знаходяться значення параметрів $a = 14{,}78$ і $b = 0{,}2564$.

Таким чином, знайдене рівняння регресії матиме наступний вигляд:

$$\hat{Y} = 14{,}78 + 0{,}2564 \cdot X.$$

На цьому процедури МНК закінчені.

3.2. Однофакторна лінійна модель: прогноз одного чинника на підставі іншого

Коефіцієнт кореляції.

Трьома основними цілями аналізу двовимірних даних, представлених парами (*X, Y*), є: (1) опис і розуміння взаємозв'язку, (2) прогнозування і прогноз нового спостереження і (3) коректування і управління процесом.

Кореляційний аналіз дозволяє зробити висновок про силу взаємозв'язку, а *регресійний аналіз* використовується для прогнозування однієї змінної на підставі іншої (як правило, *Y* на підставі *X*).

Двовимірні дані аналізують з використанням **діаграми розсіювання** в координатах *Y* і *X*, яка дає візуальне уявлення про взаємозв'язок в даних. **Кореляція,** або точніше **лінійний коефіцієнт кореляції (r)**, є безрозмірне (що не має одиниць вимірювання) число в діапазоні від **-1** до **1**, яке характеризує силу взаємозв'язку. Рівність коефіцієнта кореляції **1** свідчить про ідеальний взаємозв'язок у вигляді прямої лінії з нахилом вгору. Рівність коефіцієнта кореляції **-1** свідчить про ідеальний взаємозв'язок у вигляді нахиленої вниз (негативно) прямої лінії. Коефіцієнт кореляції говорить про те, наскільки близько до цієї нахиленої прямої лінії розташовані точки діаграми, проте він не характеризує крутизну нахилу цієї лінії. Формула обчислення коефіцієнта кореляції для тих, хто уміє користуватися Excel **має** наступний вигляд:

$$r = \frac{\frac{1}{n\ 1}\sum_{i}^{n}{}_{1}(X_i - \bar{X})(Y_i - \bar{\ })}{S_X S_Y}. \qquad (3.1)$$

Коваріація *X* і *Y* є чисельником у формулі для коефіцієнта кореляції. Оскільки одиниці вимірювання коваріації важко інтерпретувати, зручніше працювати з коефіцієнтом кореляції.

Кореляцію не можна розглядати як причинну обумовленість. Коефіцієнт кореляції характеризує зв'язок між числами, але не пояснює її. Кореляція може бути викликана тим, що змінна *X* впливає на *Y*, або тим, що змінна *Y* впливає на *X*. Крім того, кореляція може бути викликана також тим, що на *X* і *Y* впливає якийсь прихований "третій чинник", що створює враження зв'язку між *X і Y*. *Терміном* **помилкова кореляція** позначають високу кореляцію, яка виникає завдяки дії деякого третього чинника.

2. Діаграма розсіювання.

При аналізі двовимірної діаграми розсіювання можна виявити різні взаємозв'язки. Простою, з погляду аналізу, є **лінійний взаємозв'язок,** який виражається в тому, що крапки на діаграмі розсіювання з постійним розкидом групуються випадковим чином уздовж прямої лінії. Діаграма свідчить про відсутність **взаємозв'язку, якщо** крапки розміщені випадково і при переміщенні зліва направо неможливо виявити який-небудь ухил (ні вгору, ні вниз). Двовимірна діаграма розсіяння характеризується нелінійним взаємозв'язком**, якщо** крапки на ній групуються уздовж *кривої, а не прямої* лінії. Оскільки кількість видів кривих практично безмежно, аналіз нелінійного взаємозв'язку виявляється набагато складнішим, проте взаємозв'язок можна наблизити до лінійної, застосувавши до даних відповідне перетворення. Проблема нерівної **варіації виникає** тоді, коли при переміщенні по горизонталі на діаграмі розсіювання варіація крапок по вертикалі сильно міняється. Нерівна варіація призводить до зниження надійності коефіцієнта кореляції і регресійного аналізу. Проблему нерівної варіації можна вирішити за допомогою відповідних перетворень даних або за допомогою, так званої зваженої регресії. Проблема кластеринга (розділення сукупності на групи однорідніших об'єктів) виникає у разі скупчення на діаграмі розсіювання окремих, яскраво виражених груп крапок; у **таких** випадках кожну групу слід аналізувати окремо. Деяка точка даних є **викидом** (значенням, що різко відхиляється), якщо вона не відповідає взаємозв'язку між рештою даних; значення, що різко відхиляються, можуть спотворити статистичні характеристики двовимірної сукупності даних.

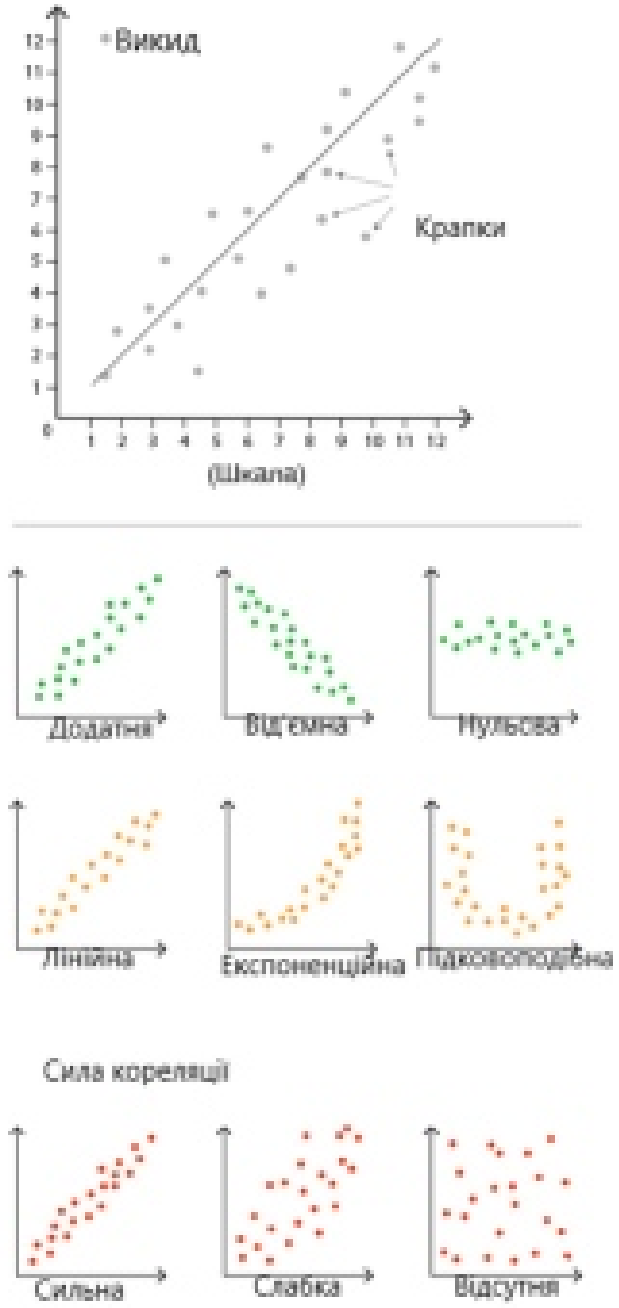


Рис. 3.5. Діаграми розсіювання

3. Регресійний аналіз.

Регресійний аналіз полягає в прогнозуванні однієї змінної на підставі іншої. **Лінійний регресійний аналіз** прогнозує значення однієї змінної на підставі іншої за допомогою прямої лінії. **Нахил** цієї лінії, виражається в одиницях вимірювання *Y* на одну одиницю *X* і характеризує крутизну підйому

або спуску (якщо ***b*** негативне) лінії. Зрушення, ***a,*** рівне значенню, яке приймає *Y* при *X*, рівному 0. Рівняння прямої лінії має наступний вигляд:

$$\boldsymbol{Y = Зрушення + (Нахил)(X) = a + bX}\ . \qquad (3.2)$$

Лінія найменших квадратів характеризується найменшою зі всіх можливих ліній сумою зведених в квадрат помилок прогнозування по вертикалі і використовується як краща лінія прогнозування, заснована на даних. **Нахил** цієї лінії, ***b***, називають також **коефіцієнтом регресії** *Y* по *X*, а зрушення ***a*** (*відрізок* відсікається на осі *Y*) називають також **постійним членом регресії**. Нижче приведені рівняння для нахилу і зрушення, відповідні лінії найменших квадратів.

Нахил рівний: $$b = r\frac{S_Y}{S_X}. \qquad (3.3)$$

Зрушення рівне: $$a = \bar{Y} - \bar{b} = \bar{Y} - r\frac{S_Y}{S_X}\bar{X}. \qquad (3.4)$$

Формула для лінії найменших квадратів має наступний вигляд:

Прогнозоване значення Y рівно:

$$\hat{Y} = a + b = (\bar{Y} - r\frac{S_Y}{S_X}\bar{X}) + r\frac{S_Y}{S_X}X. \qquad (3.5)$$

Прогнозоване значення для *Y* при заданому значенні *X* визначається шляхом підстановки цього значення *X* в рівняння для лінії найменших квадратів. Кожна з точок даних характеризується **залишком** – помилкою прогнозування, вказуючою, наскільки вище або нижче за лінію знаходиться крапка.

4. Перевірка надійності регресійної моделі.

Існують дві міри відповідності лінії найменших квадратів наявним даним. **Стандартна помилка оцінки (або прогнозу),** яку позначають S_e приблизно

указує величину помилок прогнозування (залишків) для наявних даних *в тих же одиницях, в яких зміряна і змінна* ***Y****. Відповідні* формули приведені нижче.

$$S_e = S_Y\sqrt{(1-r^2)\frac{n-1}{n-2}} \text{ (для обчислення)} \quad (3.6)$$

$$= \sqrt{\left(\frac{1}{n-2}\right)\sum_{i=1}^{n}[Y_i - (a + b\ _i)]^2} \text{ (для інтерпретації).} \quad (3.7)$$

Значення R^2, яке часто називають **коефіцієнтом детермінації,** говорить про те, який відсоток варіації ***Y*** пояснюється поведінкою ***X***.

Довірчі інтервали і перевірка гіпотез для коефіцієнта регресії пов'язані з певними припущеннями щодо аналізованої сукупності даних, які повинні гарантувати, що вона складається з незалежних спостережень, що характеризуються лінійним взаємозв'язком з рівною варіацією і приблизно нормально розподіленою випадковістю. По-перше, ці дані повинні бути довільною вибіркою з тієї, що цікавить нас генеральній сукупності. По-друге, **лінійна модель** указує, що спостережуване значення ***Y*** визначається взаємозв'язком в генеральній сукупності плюс випадкова помилка, що має нормальний розподіл. Існують параметри генеральної сукупності, відповідні нахилу і зрушенню лінії найменших квадратів, побудованої на даних вибірки:

$$\boldsymbol{Y = (a + bX) + e =}$$

= (Взаємозв'язок в генеральній сукупності) + випадковість. (3.8)

де ***e*** має нормальний розподіл з середнім значенням, рівним 0, і постійним стандартним відхиленням у**.**

Статистичні висновки (використання довірчих інтервалів і перевірки статистичних гіпотез) щодо коефіцієнтів лінії найменших квадратів ґрунтуються, як завжди, на їх стандартних помилках і значеннях з **t-таблиці** для ***n*** **- 2** мір свободи. **Стандартна помилка коефіцієнта нахилу** $\boldsymbol{S_b}$ указує приблизну величину відхилення оцінки нахилу, ***b*** (коефіцієнт регресії, обчислений на основі даних вибірки), від нахилу в генеральній сукупності, **в**,

викликаного випадковим характером вибірки. **Стандартна помилка зрушення** $\boldsymbol{S_a}$ указує приблизно, наскільки далеко оцінка зрушення $\boldsymbol{a}$ отстоит від дійсного зрушення **б** в генеральній сукупності. Відповідні формули виглядають таким чином:

стандартна помилка коефіцієнта регресії $\boldsymbol{b}$:

$$S_b = \frac{S_e}{S_X\sqrt{n-1}} \tag{3.9}$$

стандартна помилка зрушення:

$$\boldsymbol{S_a = S_e\sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_X^2(n\ \ 1)}}}. \tag{3.10}$$

Довірчий інтервал для нахилу в генеральній сукупності, b:

від $b - t_{\ b}$ до $b + t_{\ b}$.

Довірчий інтервал для зрушення в генеральній сукупності,a:

від $a - t_{\ a}$ до $a + t_{\ a}$.

Один із способів перевірки, чи є виявлений взаємозв'язок між X і Y реальним або це просто випадковий збіг, полягає в порівнянні **в** із заданим значенням $\boldsymbol{в_0 = 0}$. Про значущий зв'язок можна говорити в тому випадку, якщо **0** не потрапляє в довірчий інтервал, що базується на $\boldsymbol{b}$ і S_b, або якщо абсолютне значення $\boldsymbol{t = b/S_b}$ перевершує відповідне ***t-значення*** в ***t-таблиці.***

Ця перевірка еквівалентна перевірці значущості коефіцієнта кореляції і означає, по суті, те ж саме, що і **F-тест** для випадку, коли рівняння містить тільки одну змінну X. Зрозуміло, будь-яким з коефіцієнтів (***a*** або ***b)*** можна порівняти з будь-яким відповідним заданим значенням, скориставшись одно- або двосторонньою перевіркою (залежно від конкретних обставин) і з використанням тих же методів перевірки, що були розглянуті для середнього генеральній сукупності.

Таблиця 3.3. t – таблиця (t - критерій Стьюдента)

Довірчий інтервал							
Двосторонній	**80%**	**90%**	**95%**	**98%**	**99%**	**99,8%**	**99,9%**
Односторонній	90%	95%	97,5%	99%	99,5%	99,9%	99,95%
Рівень значущості перевірки гіпотези							
Двосторонній	***0,20***	***0,10***	***0,05***	***0,02***	***0,01***	***0,002***	***0,001***
Односторонній	*0,10*	*0,05*	*0,025*	*0,01*	*0,005*	*0,001*	*0,0005*
В цілому: ступені свободи	Критичні значення ***t***						
1	3,078	6,314	12,706	31,821	63,657	318,309	636,619
2	1,886	2,920	4,303	6,965	9,925	22,327	31,599
3	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	4,505	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
38	1,304	1,686	2,024	2,429	2,712	3,319	3,566
39	1,304	1,685	1,023	2,426	2,708	3,313	3,558
Нескінченність	1,282	1,645	1,960	2,326	2,576	3,090	3,291

3.3. Прогнозування за однофакторною моделлю

Для прогнозування *середнього* значення нового спостереження Y за умови, що $X = X_0$ (де X_0 – параметр, що цікавить дослідника, X, який ще жодного разу не зустрічався в буденній практиці), невизначеність прогнозу оцінюють за допомогою **стандартної помилки прогнозу** $S_{(прогнозоване\ Y/X0)}$, яка також має $n - 2$ мір свободи. Це дозволяє побудувати довірчі інтервали і перевірити гіпотези для нового спостереження:

$$S_{(прогнозеY/X_0)} = \sqrt{S_e^2\left(\frac{1}{n}\right) + S_b^2(X_0 - \bar{X})^2} \quad (3.11)$$

Довірчий інтервал для прогнозованого середнього значення Y при заданому значенні X_0 має наступний вигляд:

від $(a + b_0) - t_{(прогнозY/X_0)}$

до $(a + b_0) + t_{(прогнозY/X_0)}$

Точковий прогноз обчислюємо шляхом підстановки в рівняння прогнозного значення факторної змінної.

$$y_{\text{прогн}}^{\text{точк}} = a + b * x_{\text{прогн}} \quad (3.12)$$

Імовірність реалізації точкового прогнозу практично дорівнює нулю. Тому на додаток до точкового прогнозу розраховується середня помилка прогнозу або довірчий інтервал прогнозу з досить великою надійністю. Розмах прогнозного інтервалу L залежить від стандартної помилки, віддалення $x_{\text{прогн}}$ від свого середнього значення в ряді спостережень, кількості спостережень n і рівня значущості прогнозу α.

$$L = S_{ст} * t_{\alpha,n-m-1} * \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{прогн}} - \bar{x})^2}{\sum_i^n{}_1 (x_i - \bar{x})^2}} \quad (3.13)$$

Тоді фактичні значення досліджуваної ознаки з ймовірністю (1- α) потраплять в інтервал

$$y_{\text{прогн}} \in \left[y_{\text{прогн}}^{\text{точк}} - L; y_{\text{прогн}}^{\text{точк}} + L\right] \quad (3.14)$$

$$S_{\text{ст}} = S_{\text{ст}} = \sqrt{\frac{\sum_i^n{}_1 \varepsilon_i^2}{n-m-1}}, \text{де } \varepsilon_i = Y_i - Y_{\text{пр}} \quad (3.15)$$

3.4. Автокореляція: причини та наслідки

Автокореляція (autocorrelation), також відома як автокореляція часового ряду, є статистичним поняттям, яке означає кореляцію між значеннями змінної з одного і того ж ряду в різні моменти часу. Це важливий аспект в аналізі часових рядів і економетрії, оскільки може вказувати на наявність залежності між подіями в різні моменти часу.

Автокореляція виявляється у часових рядах, де спостерігається певна систематична закономірність або паттерн у величині часового ряду і її попередніх значень.

Основні причини автокореляції:

1. Структура даних.

Деякі часові ряди можуть мати в собі внутрішню структуру або паттерни, які приводять до автокореляції. Наприклад, сезонні зміни, що відбуваються з певною періодичністю (наприклад, щомісяця або щороку), можуть створювати автокореляцію.

2. Не враховані фактори.

Якщо в моделі не враховані важливі фактори, які впливають на часовий ряд, це може створити автокореляцію. Наприклад, якщо певний зовнішній фактор або подія впливає на ряд і не врахований у моделі, то може з'явитися автокореляція в залишках.

3. Перекос у вибірці.

Нерівномірний і неправильний вибір інтервалів часу при зборі даних може створити зсуви в даних, які можуть викликати автокореляцію.

4. Співвідношення змінних.

Якщо ряди взаємопов'язані та впливають один на одного, то це може створити автокореляцію. Наприклад, якщо ряд залежить від себе в минулому періоді, це може призвести до автокореляції.

5. Неспостережнувані фактори.

Іноді є незалежні фактори або події, які не враховані у моделі, але впливають на дані. Це може створити автокореляцію в залишках.

6. Помилки вимірювання.

Неточності в вимірюванні або реєстрації даних можуть викликати автокореляцію, особливо якщо ці помилки повторюються у часі.

7. Затримки в реакціях.

У деяких процесах затримки у реакціях можуть призвести до того, що вплив одного періоду буде помітний у наступному, що викличе автокореляцію.

Загалом, виявлення причин автокореляції вимагає аналізу даних, контексту дослідження та ретельного дослідження впливу різних факторів на часовий ряд.

У аналізі автокореляції використовуються такі поняття:

Корелограма (autocorrelation plot). Це графічне зображення кореляційних коефіцієнтів між значеннями ряду і його попередніми значеннями на різних відставаннях в часі. Корелограма допомагає виявити сезонність та інші структури в часовому ряді.

Коефіцієнт автокореляції (autocorrelation coefficient). Це числове значення, яке вказує на силу та напрямок автокореляції між значеннями ряду на певному відставанні. Зазвичай використовують коефіцієнт кореляції Пірсона для обчислення автокореляції.

Тести на автокореляцію. Для формальної оцінки наявності автокореляції використовуються різні статистичні тести, такі як тест Дарбі-Уотсона, тест Льюнга-Бокса та інші. Ці тести допомагають визначити, чи є статистично значуща автокореляція в часовому ряді.

Автокореляція може мати важливі наслідки для аналізу та прогнозування часових рядів. Якщо виявляється автокореляція, це може вказувати на недостатню урахованість певних факторів у моделі або на наявність певних структурних аспектів у даних, які необхідно врахувати при аналізі.

Серед найбільш спрощених випадків автокореляції залишків можна відзначити сценарій, коли e_i підкоряються авторегресійній ланцюговій динаміці

першого порядку. Це означає, що значення залишків в даному періоді залежать лише від залишків, що існували у попередньому періоді, і це можна виразити так:

$$e_i = \rho e_{i-1} + u_i, \tag{3.16}$$

Треба зазначити, що $|\rho| < 1$. Величина ρ – коефіцієнт автокореляції залишків першого порядку, який характеризує рівень взаємозв'язку кожного наступного значення з попереднім. Якщо $\rho > 0$, то автокореляція залишків є позитивною, якщо $\rho < 0$, то автокореляція залишків є негативною. В економетричних моделях негативна автокореляція спостерігається, але нечасто. Якщо $\rho = 0$, то автокореляція залишків відсутня.

На рис. 3.6. зображено різні випадки автокореляції.

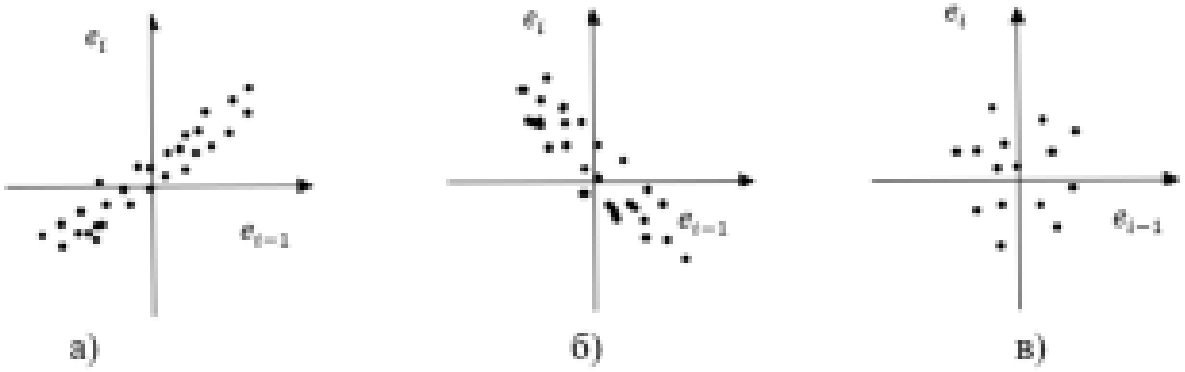


Рис. 3.6. *Види автокореляції залишків:*
а) позитивна; б) негативна; в) відсутня

Щоб дослідити автокореляцію, розглянемо економетричне рівняння багатофакторної лінійної регресії вигляду:

$$y_i = a_0 + \sum_j^m {}_1 a_j x_j + e_i, \quad i = 1 \dots n \tag{3.17}$$

де m – кількість пояснювальних змінних моделі; n – кількість вибіркових спостережень; y_i, x_{1i} , ..., x_{mi} – i -й набір вибіркових ознак. Тоді для кожного моменту часу i =1..n значення компоненти e_i визначається як

$$e_i = y_i - \widehat{y_i} \tag{3.18}$$

або

$$e_i = y_i - (a_0 + \sum_j^m {}_1 a_j x_j) \tag{3.19}$$

Якщо розглядати послідовність залишків як часовий ряд, то можна буде побудувати графік їх залежності від часу. Але метод найменших квадратів ґрунтується на тому, що залишки повинні бути випадковими, це одна з передумов.

Числовою характеристикою автокореляції слугує **коефіцієнт автокореляції** ρ . Його розраховують за безпосередніми даними рядів динаміки, коли фактичні рівні одного ряду розглядаються як значення факторної ознаки, а рівні цього ж самого ряду зсувом на один період приймаються за результативну ознаку.

Коефіцієнт автокореляції відображає ступінь кореляції між значеннями в часовому ряді та їх відстаючими значеннями на певному відставанні часу.

Коефіцієнт автокореляції може бути інтерпретований як міра схожості між значеннями в часовому ряді та їх відстаючими значеннями. Значення близьке до 1 або -1 свідчить про наявність сильної позитивної або негативної кореляції, відповідно.

Коефіцієнт автокореляції може допомогти виявити сезонність, циклічність або інші закономірності в часовому ряді. Якщо він відмінний від нуля, це може свідчити про наявність залежності між попередніми і поточними значеннями.

Високий коефіцієнт автокореляції на певних відстанях може вказувати на наявність періодичних змін в часовому ряді. Наприклад, висока автокореляція на відстані 12 може вказувати на річну сезонність.

Зв'язок з лінійною залежністю: Автокореляція також відображає ступінь лінійної залежності між значеннями ряду на різних відстанях. Великий коефіцієнт автокореляції може вказувати на високу лінійну залежність.

Типи автокореляції. Автокореляція може бути позитивною (значення зростають разом) або негативною (значення зменшуються разом). Знак автокореляції, який змінюється, може вказувати на нестабільну динаміку в часовому ряді.

Кількість періодів, за яким розраховується коефіцієнт автокореляції, називається лагом. Зі збільшенням лагу кількість пар значень, за якими обчислюється коефіцієнт автокореляції, зменшується. У загальному випадку залежність між значеннями стохастичної складової e_i у випадку автокореляції залишків можна подати формулою:

$$e_i = \rho_1 e_{i-1} + \rho_2 e_{i-2} + \cdots + \rho_s e_{i-} + u_i \qquad (3.20)$$

де ρ_1, ρ_2,..., ρ_s – коефіцієнти автокореляції 1, 2 та s-го порядків відповідно; u_i – помилка.

Послідовність коефіцієнтів автокореляції 1, 2 та s-го порядків називають автокореляційною функцією часового ряду.

Автокореляція у часових рядах та моделях має різні наслідки, які можуть впливати на аналіз, прогнозування та інтерпретацію даних. Ось деякі з наслідків автокореляції:

1. Неправильні оцінки коефіцієнтів. Автокореляція може призводити до незвичайних та ненадійних оцінок коефіцієнтів в регресійних моделях. Це може стати на шляху коректного визначення впливу різних факторів на результат.

2. Переоцінка статистичної значущості. Автокореляція може викликати переоцінку статистичної значущості коефіцієнтів. Це може призвести до надмірної впевненості у статистично значущих зв'язках, які насправді можуть бути випадковими.

3. Некоректні інтерпретації. Автокореляція може спотворювати інтерпретацію впливу факторів на результат. Помилкові залежності, виявлені через автокореляцію, можуть призвести до некоректних висновків щодо причинно-наслідкових зв'язків.

4. Недооцінка дисперсії оцінок. Автокореляція може призводити до недооцінки дисперсії оцінок параметрів моделі. Це може робити оцінки менш точними та надійними.

5. Невірні прогнози. При наявності автокореляції прогнози можуть бути неточними та неточними, оскільки вони не враховують властивості автокореляції у часовому ряді.

6. Ускладнення аналізу. Автокореляція може ускладнювати аналіз даних та робити їх інтерпретацію складнішою. Потреба в додаткових корекціях та методах може збільшити складність аналізу.

7. Важкість визначення причин. Автокореляція може бути наслідком різних факторів, таких як структура даних, формулювання моделі, систематичні помилки вимірювання та інші. Визначення точної причини може бути викликом.

Загалом, наслідки автокореляції можуть впливати на точність, надійність та правильність аналізу та інтерпретації даних. Виявлення автокореляції важливо для вжиття відповідних заходів, таких як корекція моделі або використання спеціальних методів аналізу.

Один із можливих шляхів усунення проблеми автокореляції залишків полягає в застосуванні узагальненого метода найменших квадратів до оцінки параметрів моделі (методу Ейткена).

3.5. Тест Дарбіна–Уотсона та метод Ейткена при автокореляції залишків

Для виявлення автокореляції можна, по-перше, візуально представити залишки на графіку. На рис. 3.7 представлено можливі варіанти.

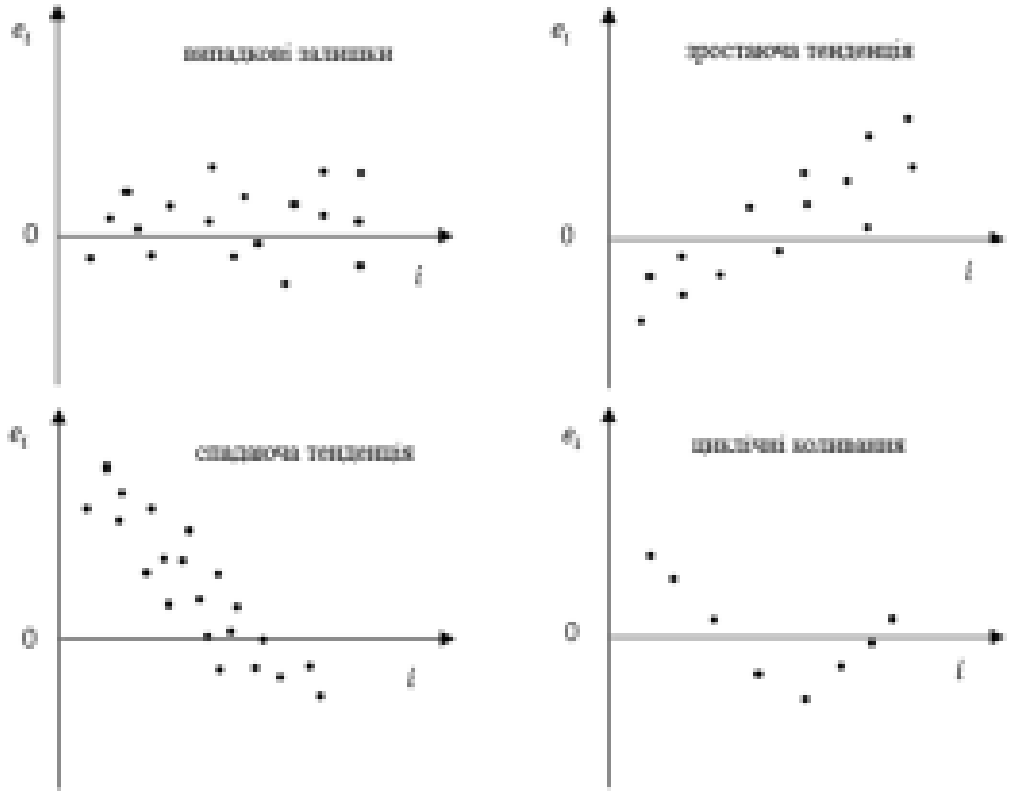


Рис. 3.7. Моделі залежності залишків від часу:

Другий метод – використання статистичного тесту Дарбіна–Уотсона.

Коефіцієнт автокореляції ρ можна оцінити методом найменших квадратів на основі відомих залишків для вибірки, звідки отримаємо формулу:

$$\rho = \frac{cov(e_{i-1}, e_i)}{\sigma_e^2} \approx \frac{\sum_{i=2}^{n} e_{i-1} e_i}{\sum_{i=1}^{n} e_i^2} \tag{3.21}$$

Але замість формули (3.21) досить часто розраховують оцінку коефіцієнта автокореляції ρ за співвідношенням:

$$\rho = \frac{\frac{1}{n-1}\sum_{i=2}^{n} e_{i-1} e_i}{\frac{1}{n}\sum_{i=1}^{n} e_i^2} = \frac{n}{n-1} \frac{\sum_{i=2}^{n} e_{i-1} e_i}{\sum_{i=1}^{n} e_i^2} \tag{3.22}$$

Отриману за формулоб (3.22) оцінку називають циклічним коефіцієнтом автокореляції.

Розглянемо найбільш відомий і поширений кількісний тест для перевірки моделі на наявність автокореляції залишків, який ґрунтується на критерії Дарбіна–Уотсона. Цей тест використовується для авторегресійних схем першого порядку наступним чином:

1. За припущенням про відсутність автокореляії залишків методом найменших квадратів створюється економетрична модель та обчислюються її залишки e_i, $i = 1..n$.

2. Розраховується критерій Дарбіна–Уотсона за формулою:

$$D = \frac{\sum_{i\ 2}^{n}(e_i - {}_{i\ 1})^2}{\sum_{i\ 1}^{n} e_i^2} \qquad (3.23)$$

Критерій Дарбіна–Уотсона та коефіцієнт автокореляції залишків першого порядку пов'язані співвідношенням:

$$DW \approx 2 \cdot (1 - \rho). \qquad (3.24)$$

Якщо в залишках існує повна позитивна автокореляція і $\rho = 1$, то $DW = 0$. Якщо в залишках є повна негативна автокореляція, то $\rho = -1$ та, відповідно, $DW = 4$. Якщо автокореляція залишків відсутня, то $\rho = 0$ і $DW = 2$. Отже, $0 \le DW \le 4$

3. За обраного рівня значущості α і заданих m факторах моделі та n вибіркових спостереженнях за статистичними таблицями DW розподілу Дарбіна–Уотсона визначаються граничні значення – нижня dL та верхня dU межі критерію Дарбіна–Уотсона, як показано в табл. 6.1 при $\alpha = 0{,}05$, 3 m =1.. , 20 n = 6.. .

Таблиця 3.4. DW-розподіл Дарбіна–Уотсона при рівні значущості α = 0,05

n	m=1		m=2		m=3	
	d_L	d_U	d_L	d_U	d_L	d_U
6	0,610	1,400	-	-	-	-
7	0,700	1,356	0,467	1,896	-	-
8	0,763	1,332	0,559	1,777	0,368	2,287
9	0,824	1,320	0,629	1,699	0,455	2,128
10	0,879	1,320	0,697	1,641	0,525	2,016
11	0,927	1,324	0,658	1,604	0,595	1,928
12	0,971	1,331	0,812	1,579	0,658	1,864
13	1,010	1,340	0,861	1,562	0,715	1,816
14	1,045	1,350	0,905	1,551	0,767	1,779
15	1,077	1,361	0,946	1,543	0,814	1,750
16	1,106	1,371	0,982	1,539	0,857	1,728
17	1,133	1,381	1,015	1,536	0,897	1,710
18	1,158	1,391	1,046	1,535	0,933	1,696
19	1,180	1,401	1,074	1,536	0,967	1,685
20	1,201	1,411	1,100	1,537	0,998	1,676

4. Формуються зони автокореляціонного зв’язку. Шкалу визначення автокореляції на основі порівняння фактично розрахованого критерію Дарбіна–Уотсона та його критичних значень зображено на рис. 3.8.

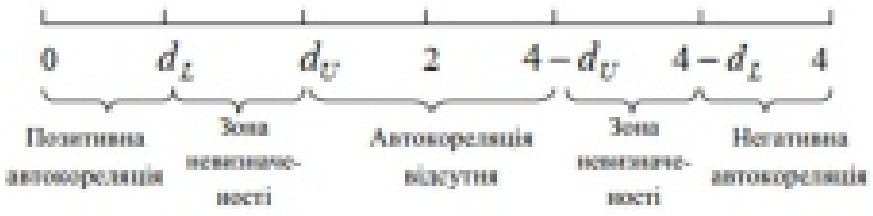


Рис. 3.8. Шкала визначення автокореляційного зв’язку

5. На основі розрахованого значення *DW* робиться висновок про наявність або відсутність автокореляції залишків:

• якщо $0 < DW < dL$, то існує позитивна автокореляції;

• якщо $4 - dL < DW < 4$, то існує негативна автокореляції;

• якщо $dL \le DW \le dU$ або $4 - dU \le DW \le 4 - dL$, то неможливо зробити висновок про наявність або відсутність автокореляції залишків;

• якщо $dU < DW < 4 - dU$, то автокореляція залишків відсутня.

Оцінку параметрів моделі з автокорельованими залишками можна виконати на основі методу Ейткена. Його доцільно застосовувати у випадках, коли залишки описуються авторегресійною моделлю першого ступеня.

Методи Кочрена–Оркатта і Дарбіна використовують для оцінки параметрів економетричної моделі тоді, коли залишки описуються авторегресійною моделлю більш високого ступеня, наприклад:

$$e_i = \rho_i e_{i-1} + \rho_2 e_{i-2} + u_i$$
$$e_i = \rho_i e_{i-1} + \rho_2 e_{i-2} + \rho_3 e_{i-3} + u_i$$

Розглянемо створення лінійної економетричної моделі методом Ейткена. У цьому випадку оператор визначення числових коефіцієнтів лінійної економетричної моделі задається матричною формулою:

$$A = (X'S^{-1}X)^{-1}X'S^{-1}Y \tag{3.25}$$

де матриця S^{-1} має вигляд:

$$S^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \dim S^{-1} = n \times n. \tag{3.26}$$

За допомогою матричних перетворень визначаються:

– матриця значень незалежних змінних моделі X:

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{mn} \end{pmatrix}, \quad \dim X = n \times (m+1); \tag{3.27}$$

– транспонована до X матриця X' розраховується за допомогою функції *TRANSPOSE* або *ТРАНСП* (*MS Excel*):

$$X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}, \quad \dim X' = (m+1) \times n; \tag{3.28}$$

– добуток матриць $X'S^{-1}$ обчислюється функцією *MMULT* або *МУМНОЖ* (*MS Excel*);

– обчислюється добуток матриць $X'S^{-1}X$;

– обернена матриця $(X'S^{-1}X)^{-1}$ визначається за допомогою функції *MINVERSE* або *МОБР* (*MS Excel*);

– розраховується добуток матриць $X'S^{-1}Y$;

– вектор значень параметрів A узагальненої моделі визначається функцією *MMULT* або *МУМНОЖ* (*MS Excel*) за формулою (3.25).

3.6. Індивідуальне завдання № 5

Засвоєння методики знайдення коефіцієнтів однофакторної моделі

Мета роботи: Набути навичок з уміння розрахувати коефіцієнти лінійної регресії методом найменших квадратів.

Порядок виконання:

1. Знайти коефіцієнти лінійних моделей Y1= f(X1), Y1= f(X2), Y1= f(X3), Y2= f(X1), Y2= f(X2), Y2= f(X3).
2. Побудувати графіки цих залежностей
3. Знайти коефіцієнт детермінації.
4. Визначити значимість коефіцієнтів моделей за критерієм Фішера.

Дані необхідно згенерувати відповідно до номеру у списку, за формулами, які наведено у табл. 3.5.

Таблиця 3.5. Дані за варіантами

X1, тис. грн	Y1, тис. грн	X2, тис. грн	X3, тис. грн	Y2, тис. грн
100+20*N	70+5*N	75+15*N	105+12*N	170+7*N
130+10*N	90+5*N	80+15*N	90+12*N	180+7*N
150+10*N	105+5*N	115+15*N	145+12*N	165+7*N
160+10*N	95+5*N	100+15*N	170+12*N	175+7*N
175+10*N	130+5*N	120+15*N	175+12*N	185+7*N
180+10*N	120+5*N	110+15*N	170+12*N	190+7*N
185+10*N	140+5*N	100+15*N	165+12*N	200+7*N
200+10*N	150+5*N	140+15*N	210+12*N	195+7*N
215+10*N	150+5*N	160+15*N	220+12*N	200+7*N
220+10*N	145+5*N	150+15*N	215+12*N	205+7*N
240+10*N	165+5*N	185+15*N	210+12*N	210+7*N
250+10*N	172+5*N	152+15*N	220+12*N	220+7*N
260+10*N	155+5*N	110+15*N	225+12*N	215+7*N
270+10*N	143+5*N	120+15*N	230+12*N	225+7*N
280+10*N	167+5*N	140+15*N	280+12*N	230+7*N

Методичні рекомендації та приклад

Критерій Фішера для регресійної моделі відбиває, наскільки добре ця модель пояснює загальну дисперсію залежною змінною. Розрахунок критерію виконується за рівнянням:

$$F = \frac{R^2}{1-R^2} \cdot \frac{f_2}{f_1}$$

де R – коефіцієнт кореляції; f_1 та f_2 – число ступенів свободи.

Перший дріб у рівнянні дорівнює відношенню поясненої дисперсії до непоясненої. Кожна з цих дисперсій поділяється на свій ступінь свободи (другий дріб у виразі). Число ступенів свободи поясненої дисперсії f_1 дорівнює кількості пояснюючих змінних (наприклад, для лінійної моделі виду Y = a +bx отримуємо f_1 = 1). Число ступенів свободи непоясненої дисперсії f_2= N-k-1, де N-кількість експериментальних точок, k-кількість пояснюючих змінних (наприклад, для моделі Y=a+bx підставляємо k=1).

Ще один приклад: для лінійної моделі виду Y = A0 + A1 * X1 + A2 * X2, побудованої за 20 експериментальними точками, отримуємо f_1 = 2 (дві змінних X1 і X2), f_2= 20-2-1 = 17.

Для перевірки значущості рівняння регресії обчислене значення критерію Фішера порівнюють з табличним, взятим для числа ступенів свободи f_1 (велика дисперсія) та f_2 (менша дисперсія) на вибраному рівні значущості (зазвичай 0.05). Якщо розрахований критерій Фішера вище, ніж табличний, то пояснена дисперсія значно більше, ніж непояснена, і модель є значимою.

Розрахунок параметрів парного рівняння регресії можна робити за допомогою електронних таблиць двома способами:

- стандартним методом найменших квадратів;
- з використанням убудованої функції LINEST () (ЛИНЕЙН(...).)

За допомогою вбудованого пакету аналізу Аналіз даних – Регресія

Розрахуємо на прикладі ціни на навфту та ціни на бензин за певний період часу (див. табл. 3.6).

Таблиця 3.6. Ціна на нафту та ціна на бензин

Ціна на нафту (X)	Ціна на бензин А-95 (Y)
1	2
92.52 ₴	49.41 ₴
94.48 ₴	49.43 ₴
91.87 ₴	49.43 ₴
91.64 ₴	49.43 ₴
89.90 ₴	49.43 ₴
91.81 ₴	49.42 ₴
92.36 ₴	49.43 ₴
93.30 ₴	49.43 ₴

Ціна на нафту (X)	Ціна на бензин А-95 (Y)
91.32 ₴	49.43 ₴
91.22 ₴	49.43 ₴
93.86 ₴	49.41 ₴
94.86 ₴	49.42 ₴
93.93 ₴	49.42 ₴
92.38 ₴	49.42 ₴
94.50 ₴	49.39 ₴
95.47 ₴	49.39 ₴
94.63 ₴	49.40 ₴
98.55 ₴	49.38 ₴
98.18 ₴	49.40 ₴
95.54 ₴	49.40 ₴
92.65 ₴	49.39 ₴
93.53 ₴	49.40 ₴
96.04 ₴	49.38 ₴
92.66 ₴	49.43 ₴
93.64 ₴	49.43 ₴
92.92 ₴	49.77 ₴
89.70 ₴	49.61 ₴
87.56 ₴	49.61 ₴
87.52 ₴	50.26 ₴

На рисунку 3.9 представлено розрахунки в Excel за формулами та за допомогою функції Linest ()

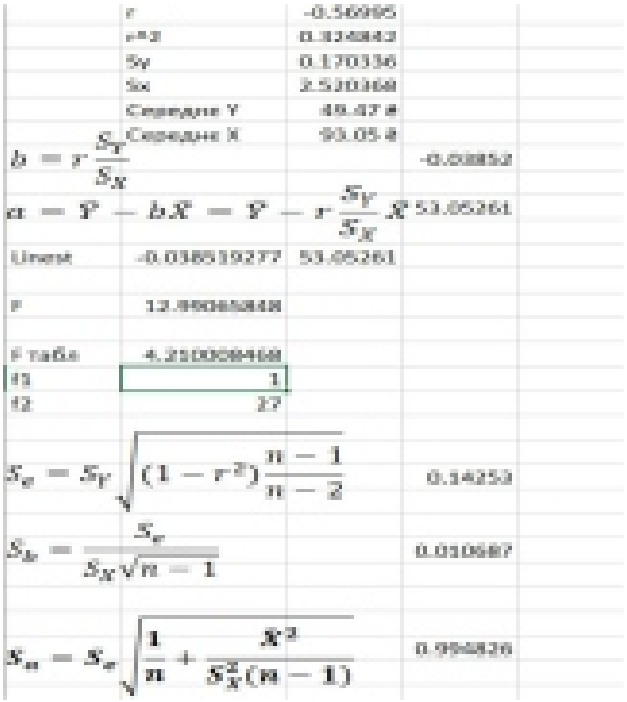


Рис.3.9. Розрахунок в Excel

Fтабл розраховуємо за допомогою функції F.INV.RT(0.05, 1,27)

Оскільки табличне значення критерію Фішера менше, ніж розраховане, то можна говорити про значущість рівняння регресії.

На рис. 3.10 представлено дані та лінію регресії.

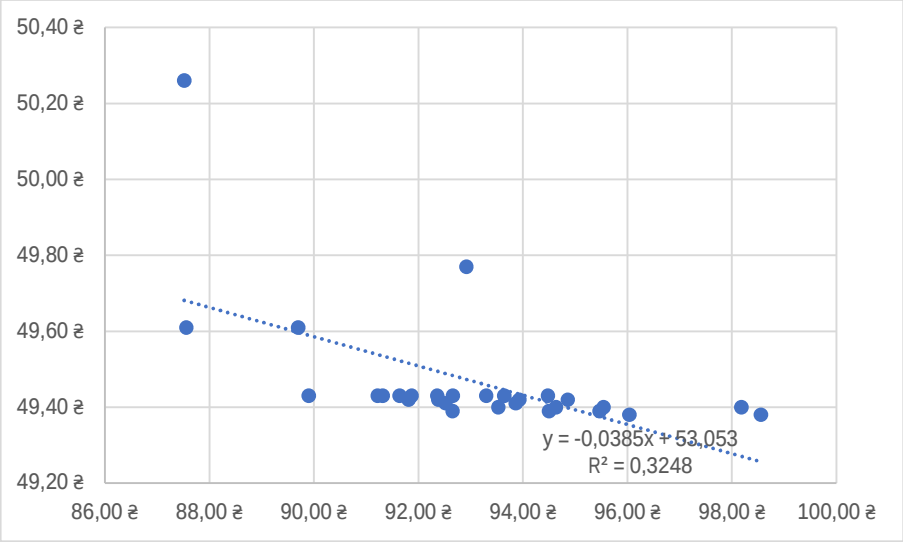


Рис. 3.10. Лінія регресії

Також можна розрахувати параметри регресії та інші коефіцієнти за допомогою Регресіїї з пакету Аналізу даних. На рисунку 3.11 представлено результат.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.569948406
R Square	0.324842325
Adjusted R Square	0.299836485
Standard Error	0.145052806
Observations	29

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.273326813	0.273327	12.99366	0.001248131
Residual	27	0.568086988	0.02104		
Total	28	0.841413793			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	53.05263413	0.994638652	53.33786	6.67E-29	51.01137383	55.09385	51.01137	55.09385444
X Variable 1	-0.038519277	0.010687106	-3.60426	0.001248	-0.06044755	-0.01659	-0.06045	-0.016591024

Рис. 3.11. Роздрук результату застосування Regression в Excel

Зауважимо, що в пакет аналізу Excel надає стандартна помилку регресії

$$S_{\text{ст}} = \sqrt{\frac{\sum_{i\ 1}^{n} \varepsilon_i^2}{n-m-1}}, \text{ де } \varepsilon_i = Y - \hat{Y}.$$

Розрахунок стандартної помилки регресії педставлено у табл. 3.7.

Таблиця 3.7. Розрахунок стандартної помилки регресії

$\hat{Y}$	$(Y-\hat{Y})^2$
49.4888	0.0062
49.4133	0.0003
49.5138	0.0070
49.5227	0.0086
49.5897	0.0255
49.5162	0.0092
49.4950	0.0042
49.4588	0.0008
49.5350	0.0110
49.5389	0.0119
49.4372	0.0007
49.3987	0.0005
49.4345	0.0002
49.4942	0.0055
49.4125	0.0005
49.3752	0.0002
49.4075	0.0001
49.2565	0.0152
49.2708	0.0167
49.3725	0.0008
49.4838	0.0088
49.4499	0.0025
49.3532	0.0007
49.4834	0.0029
49.4457	0.0002
49.4734	0.0880
49.5974	0.0002
49.6799	0.0049
49.6814	0.3348
сума	0.5681

$$S_{\text{ст}} = \sqrt{\frac{\sum_{i\ 1}^{n} \varepsilon_i^2}{n-m-1}} = \sqrt{\frac{0.5\ \ 82}{2\ \ -1-1}} = 0.14505$$

3.7. Індивідуальне завдання № 6

Засвоєння методики визначення значущості коефіцієнтів однофакторної моделі

Мета роботи: Набути навичок з уміння розрахувати значущість коефіцієнтів однофакторної моделі

Порядок виконання:

1. Оцінити значущість коефіцієнтів моделей Y1= f(X1), Y1= f(X2), Y1= f(X3), Y2= f(X1), Y2= f(X2), Y2= f(X3), отриманих у індивідуальному завданні №5, за критерієм Стьюдента.

2. Оцінити значущість коефіцієнту детермінації.

3. Побудувати довірчі інтервали для коефіцієнтів регресії.

Методичні рекомендації та приклад

t – значення (критерій Стьюдента, t-критерій або критичне значення t), знаходимо при рівні значимості 0.05 або 0.01 та числу ступенів свободи n-k-1. Для однофакторної моделі число ступенів свободи n-2. В MS Excel можна використати функцію *T.INV.2T(імовірність, число ступенів свободи).*

Помилка коефіцієнта кореляції:

$$S_r = \sqrt{\frac{1 - r^2}{n - 2}} \ ;$$

де r^2– коефіцієнт детермінації, n – кількість спостережень.

Для оцінки значущості коефіцієнтів регресії та коефіцієнту детермінації обчислюємо коефіцієнти надійності коефіцієнтів кореляції і регресії (значення статистики Стьюдента):

$$t_r = \frac{|r|}{S_r}\ ;\ \ t_a = \frac{|a|}{S_a};\ \ t_b = \frac{|b|}{S_b}\ .$$

Якщо обчислені значення більше t – значення, то коефіцієнти значущі.

Розглянемо приклад з попереднього індивідуального завдання. Було розраховано рівняння регресії, де залежною змінною є ціна на бензин, а залежною ціна на нафту.

Рівняння регресії має наступні коефіцієнти a=53.0526, b=-0.0385.

$$r^2 = 0.325$$

Розрахунок значущості представлено на рисунку 3.12.

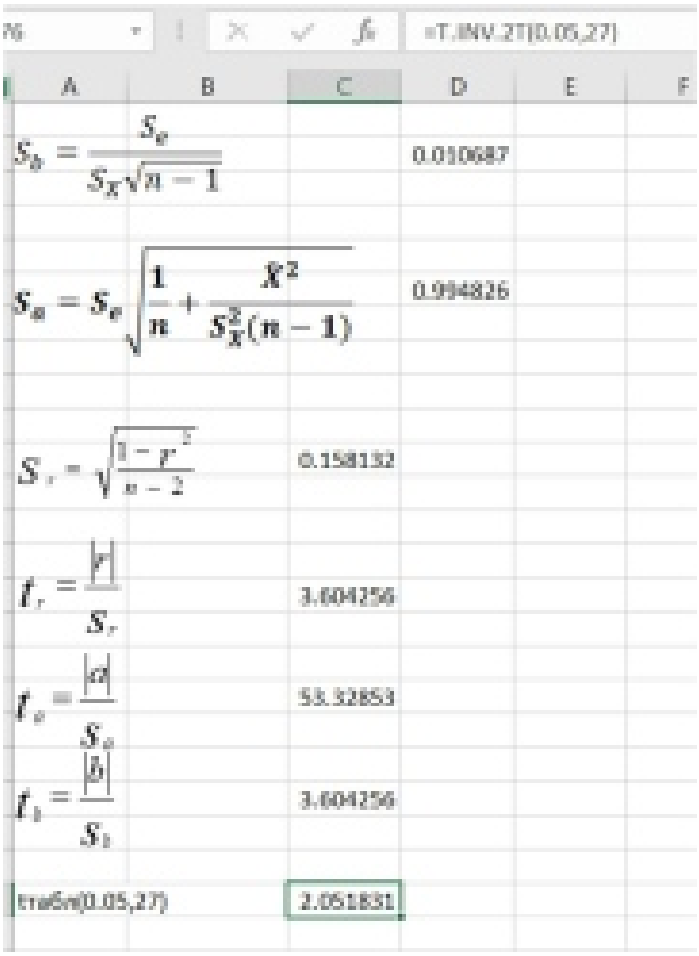


Рис.3.12. Розрахунок значимості коефіцієнтів регресії

Оскількі всі розраховані фактичні значення t більше критичного значення t на рівні значимості 0.05, то їх можна вважати значущими.

Побудуємо довірчий інтервал для нахилу в генеральній сукупності, b:

від $b - t_{b}$ = –0.0385 – 2.0518*0.010687 = –0.0604 до $b + t_{b}$. = –0.0385 – 2.0518*0.010687= – 0.01659.

Довірчий інтервал для зрушення в генеральній сукупності, a:

від $a - t_{a}$ = 53.0526 – 2.0518*0.9948 = 51.011 до $a + t_{a}$. = 53.0526+2.0518*0.9948=55.0938.

Також можна розрахувати параметри регресії та інші коефіцієнти за допомогою Регресііі з пакету Аналізу даних – Регресія (Regression). На рисунку 3.13 представлено результат.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.569949406
R Square	0.324842325
Adjusted R Square	0.299836485
Standard Error	0.145852606
Observations	29

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.275326813	0.275327	12.99066	0.001248131
Residual	27	0.56808998	0.02104		
Total	28	0.843413793			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%Lower 95.0%	Upper 95.0%
Intercept	53.05261413	0.994838852	53.32786	6.67E-30	51.01137383	55.09385	51.01137	55.09385444
X Variable 1	-0.038529277	0.010687166	-3.60426	0.001248	-0.06044755	-0.01659	-0.06045	-0.016591024

Рис. 3.13. Результат застосування Регресііі

3.8. Індивідуальне завдання № 7

Засвоєння методики прогнозування із застосуванням лінійної однофакторної моделі

Мета роботи: Набути навичок з уміння прогнозуванти із застосуванням лінійної однофакторної моделі

Порядок виконання:

1. Для моделей Y1= f(X1), Y1= f(X2), Y1= f(X3), Y2= f(X1), Y2= f(X2), Y2= f(X3), знайти точковий прогноз для точок X > 1,2Xmax та X < 1,2Xmin.

2. Для лінійних моделей Y1= f(X1), Y1= f(X2), Y1= f(X3), Y2= f(X1), Y2= f(X2), Y2= f(X3)знайти довірчий інтервал для всіх точкових прогнозів для рівнів значущості 0,15; 0,1 та 0,05.

Значення для X та Y у п'ятому індивідуальному завданні.

Методичні рекомендації та приклад

Продовжимо розглядати приклад з регресійною залежністю ціни бензина від ціни на нафту. Допоміжні розрахунки представлено в таблиці 3.8.

Нехай необхідно зробити прогноз для $X_0 = 90$.

Для розрахунку **стандартної помилки прогнозу** $S_{(прогнозоване\ Y/X0)}$,

використаємо значення S_e^2 та S_b^2, які були розраховані в попередніх роботах. Тоді

$$S_{\text{прогнозне}Y/X_0)} = \sqrt{S_e^2\left(\frac{1}{n}\right) + S_b^2(X_0 - \bar{X})^2} = \sqrt{0.14253^2\left(\frac{1}{2}\right) + 0.010687^2(90 - 93.05)^2} = 0.042$$

Довірчий інтервал для прогнозованого середнього значення Y при заданому значенні $X_0 = 90$ та рівні значущості 0.05 і ступенів свободи 29-2=27 (t=2.0518) має наступний вигляд:

від $(a + b_0) - t_{(\text{прогноз}Y/X_0)}$=(53.0526-0.0385*90) - 2.0518*0.042=49.49967

до $(a + b_0) + t_{(\text{прогноз}Y/X_0)}$=(53.0526-0.0385*90) + 2.0518*0.042=49.672

Для знаходження довірчого інтервалу для індивідуального значення розрахуємо значення L

$$S_{\text{ст}} = \sqrt{\frac{\sum_i^n{}_1 \varepsilon_i^2}{n-m-1}} = \sqrt{\frac{0.5\ \ 81}{2\ \ -1-1}} = 0.145053$$

На рівні значущості 0.05 і ступенів свободи 29-2=27.

$$L = S_{cm} * t_{\alpha,n-m-1} * \sqrt{1 + \frac{1}{n} + \frac{\left(x_{\text{прогн}} - \bar{x}\right)^2}{\sum_i^n {}_1 (x_i - \bar{x})^2}} = 0.145053 * 2.0518 *$$

$$\sqrt{1 + \frac{1}{2} + \frac{(9 - 3.05)^2}{184.2154}} = 0.3229.$$

Отже, можна побудувати інтервал для точкового прогнозу:

$$y_{\text{прогн}} \in \left[y_{\text{прогн}}^{\text{точк}} - L; y_{\text{прогн}}^{\text{точк}} + L\right].$$

$$y_{\text{прогн}} \in [49.5859 - 0.3229; 49.5859 + 0.3229].$$

$$y_{\text{прогн}} \in [49.2629; 49.9088].$$

Таблиця 3.8 – Дані та розрахунки

Ціна на нафту (X)	Ціна на бензин А-95 (Y)	$Y_{\text{прогн}}$	$(Y_i - Y_{\text{прогн}})^2$	$(x_i - \bar{x})^2$
92.52 ₴	49.41 ₴	49.4888	0.0062	0.2842
94.48 ₴	49.43 ₴	49.4133	0.0003	2.0360
91.87 ₴	49.43 ₴	49.5138	0.0070	1.3997
91.64 ₴	49.43 ₴	49.5227	0.0086	1.9969
89.90 ₴	49.43 ₴	49.5897	0.0255	9.9421
91.81 ₴	49.42 ₴	49.5162	0.0092	1.5453
92.36 ₴	49.43 ₴	49.4950	0.0042	0.4804
93.30 ₴	49.43 ₴	49.4588	0.0008	0.0610
91.32 ₴	49.43 ₴	49.5350	0.0110	3.0036
91.22 ₴	49.43 ₴	49.5389	0.0119	3.3603
93.86 ₴	49.41 ₴	49.4372	0.0007	0.6511
94.86 ₴	49.42 ₴	49.3987	0.0005	3.2649
93.93 ₴	49.42 ₴	49.4345	0.0002	0.7689
92.38 ₴	49.42 ₴	49.4942	0.0055	0.4531
94.50 ₴	49.39 ₴	49.4125	0.0005	2.0935
95.47 ₴	49.39 ₴	49.3752	0.0002	5.8414
94.63 ₴	49.40 ₴	49.4075	0.0001	2.4866
98.55 ₴	49.38 ₴	49.2565	0.0152	30.2159
98.18 ₴	49.40 ₴	49.2708	0.0167	26.2851
95.54 ₴	49.40 ₴	49.3725	0.0008	6.1847
92.65 ₴	49.39 ₴	49.4838	0.0088	0.1625
93.53 ₴	49.40 ₴	49.4499	0.0025	0.2274
96.04 ₴	49.38 ₴	49.3532	0.0007	8.9216
92.66 ₴	49.43 ₴	49.4834	0.0029	0.1545
93.64 ₴	49.43 ₴	49.4457	0.0002	0.3444
92.92 ₴	49.77 ₴	49.4734	0.0880	0.0177
89.70 ₴	49.61 ₴	49.5974	0.0002	11.2433
87.56 ₴	49.61 ₴	49.6799	0.0049	30.1742
87.52 ₴	50.26 ₴	49.6814	0.3348	30.6152
		Сума	**0.5681**	**184.2154**

Якщо потрібно розрахувати на рівні значущості 0.01, то беремо значення критерію Стьюдента при такому рівні значущості та тих же ступенях свободи, це значення 2.7707. А далі перераховуємо довірчі інтервали.

3.9. Індивідуальне завдання № 8
Засвоєння методики визначення автокореляції залишків та методу Ейткена

Мета роботи: Набути навичок з уміння розрахувати автокореляцію залишків та оцінювати її за методом Дарбіна-Уотсона та методу Ейткена

Порядок виконання:

1. Оцінити за даними X1 та Y1 автокореляцію залишків за допомогою критерію Дарбіна-Уотсона.
2. Розрахувати за методом Ейткена нові параметри регресії.
3. Оцінити знову автокореляцію.

Значення для X1 та Y1 у п’ятому індивідуальному завданні.

Методичні рекомендації та приклад

Розглянемо розрахунок автокореляції за даними в таблиці 3.9.

За допомогою пакету аналізу MS Excel, спочатку знаходимо коефіцієнти регресії (a= 1180.23, b= 0.6177) та розраховуємо $\widehat{y_i}$, помилки, критерій Дарбіна-Уотсона та коефіцієнт автокореляції.

Результати надано у таблиці 3.10

Таблиця 3.9. Вихідні дані

№	Y	X
1	1226.4	239.25
2	1014.7	387.15
3	1365.1	382.8
4	978.2	320.45
5	992.8	378.45
6	970.9	314.65
7	1868.8	339.3
8	1627.9	366.85
9	1817.7	249.4
10	1226.4	211.7
11	1350.5	374.1
12	1233.7	207.35
13	1803.1	368.3
14	1839.6	381.35

Таблиця 3.10. Розрахунок залишків

$\widehat{y}_i$	e_i	$(e_i - e_{i-1})^2$	e_i^2	$e_{i-1}e_i$
1328.0084	-101.6083999	=	10,324.27	
1419.364347	-404.6643466	91,842.91	163,753.23	41,117.30
1416.677407	-51.57740697	124,670.39	2,660.23	20,871.54
1378.164606	-399.9646059	121,373.64	159,971.69	20,629.14
1413.990467	-421.1904674	450.54	177,401.41	168,461.28
1374.58202	-403.6820198	306.55	162,959.17	170,027.02
1389.808011	478.9919891	779,113.41	229,433.33	-193,360.45
1406.825295	221.0747049	66,521.33	48,874.03	105,893.01
1334.277926	483.4220743	68,826.14	233,696.90	106,872.39
1310.991116	-84.59111577	322,638.98	7,155.66	-40,893.21
1411.303528	-60.80352775	565.85	3,697.07	5,143.44
1308.304176	-74.60417616	190.46	5,565.78	4,536.20
1407.720942	395.3790584	220,884.24	156,324.60	-29,496.93
1415.78176	423.8182396	808.79	179,621.90	167,568.86
		1,798,193.21	1,541,439.26	547,369.57

$$D = \frac{\sum_{i\ 2}^{n}(e_i - {}_{i\ 1})^2}{\sum_{i\ 1}^{n} e_i^2} = \frac{1{,}79\ \ ,1\ \ 3.21}{1{,}541{,}43\ \ .2} = 1.167.$$

Оскільки для m=1, n=14, d_L=1,045, d_U =1,35, то розрахований критерій вказує на те, що є не можна сказати точно, чи є автокореляція, тому що значення 1.167 потрапляє у зону невизначеності.

$$\rho = \frac{\frac{1}{n\ 1}\sum_{i\ 2}^{n} e_{i\ 1}e_i}{\frac{1}{n}\sum_{i\ 1}^{n} e_i^2} = \frac{n}{n-1}\frac{\sum_{i\ 2}^{n} e_{i\ 1}e_i}{\sum_{i\ 1}^{n} e_i^2} = 0.3824.$$

Розрахуємо за методом Ейткена нові параметри регресії. Результати представлено на рис. 3.14.

Рис.3.14. Розрахунок параметрів регресії за методом Ейткена

Перевіремо нові параметри на наявність автокореляції при нових параметрах (див. табл. 3.11).

Таблиця 3.11. Розрахунок нових залишків

$\widehat{y_i}$	e_i	$(e_i - e_{i-1})^2$	e_i^2
1	2	3	4
1313.37861	-86.97860953	=	7,565.28
1443.698377	-428.9983766	116,977.52	184,039.61
1439.865442	-74.76544223	125,480.97	5,589.87
1384.926717	-406.7267169	110,198.29	165,426.62
1436.032508	-443.2325079	1,332.67	196,455.06
1379.816138	-408.9161378	1,177.61	167,212.41
1401.536099	467.263901	767,691.46	218,335.55

Продовження таблиці 3.11

1	2	3	4
1425.81135	202.0886503	70,317.91	40,839.82
1322.322123	495.377877	86,018.57	245,399.24
1289.103359	-62.70335881	311,454.67	3,931.71
1432.199574	-81.69957358	360.86	6,674.82
1285.270424	-51.57042449	907.77	2,659.51
1427.088994	376.0110055	182,825.88	141,384.28
1438.587797	401.0122025	625.06	160,810.79
	сума	1,775,369.24	1,546,324.56

$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} = 1.148$$

Знову коефіцієнт Дарбіна-Уотсона попадає в інтервал від d_L=1,045 до d_U =1,35, що говорить неможливість визначення автокореляції.

Приходимо до висновку, що ми не звільнились від невизначеності в автокореляції залишків. Це означає, що вихідна гіпотеза, коли залишки описуються авторегресійною схемою першого порядку, не дотримується. Якщо залишки описуються авторегресійною схемою більш високого порядку, то доцільно виконати оцінку параметрів моделі методом Кочрена—Оркатта або Дарбіна.

Контрольні запитання

1. Що таке тренд?
2. Що являє собою кореляція?
3. В чому полягає помилкова кореляція?
4. Яким чином по діаграмі розсіювання виявити взаємозвязок X та Y?
5. На чому ґрунтуються статистичні висновки щодо коефіцієнтів лінії найменших квадратів?
6. Про що говорить коефіцієнт детермінації?

7. Як знайти довірчий інтервал для прогнозованого середнього значення Y при заданому значенні X?

8. Яким чином визначити значущість моделі регресії?

9. Яким чином можна визначити значущість параметрів моделі регресії?

10. Що являє собою автокореляція?

11. Які причини автокореляції?

12. Які наслідки спричиняє автокореляція при застосуванні регресійних моделей?

13. Які методи виявлення автокереляції?

14. В чому полягають методи усунення автокореляції?

У цьому розділі студенти навчилися будувати однофакторні регресійні моделі, визначати їх значущість, робити прогнози на їх основ та перевіряти наявність автокореляції

РОЗДІЛ 4

БАГАТОФАКТОРНІ МОДЕЛІ

Вивчивши матеріали цього розділу, студенти зможуть знаходити коефіцієнти багатофакторних моделей та визначати їх статистичну достовірність

4.1. Багатофакторна регресія: основні поняття

Навколишній нас світ багатовимірний. У переважній більшості реальних економічних завдань доводиться розглядати дані більш ніж по одному або два чинники. Проте це не є нерозв'язною проблемою: наступний крок, множинна *регресія*, є відносно нескладною процедурою, яка дозволяє вам розширити свої можливості за межі простих випадків одно- і двовимірних даних. Більш того, з відповідними базовими ідеями ви вже знайомі: поняття середнього значення, мінливості, кореляції, прогнозування, довірчих інтервалів і перевірки гіпотез.

Прогнозування єдиної змінної Y на підставі двох або декілька змінних *X* називається множинною регресією. Прогнозування єдиної змінної *Y* на підставі єдиної змінної *X* називається *простою регресією;* про просту регресію мова йшла раніше. Користуючись множинною регресією, ми переслідуємо, по суті, ті ж цілі, що і у разі простої регресії. Нижче приведений короткий огляд цих цілей, що супроводжується простими прикладами.

Перше. ***Опис і розуміння взаємозв'язку.***

а) Розглянемо взаємозв'язок між заробітною платою (Y) і поряд базових характеристик службовців, таких як стать (X_1 представлений двома значеннями, 0 і 1 позначають відповідно чоловіків і жінок), стаж роботи (X_2) і освіта (X_3).

Опис і розуміння того, як ці *X-фактори* впливають на *Y*, дозволяє, наприклад, вибудовувати систему доказів в судових процесах, що стосуються дискримінації за ознакою статі. Коефіцієнт регресії за ознакою статі є оцінкою величини різниці заробітної плати між чоловіками і жінками *з урахуванням поправки* на вік і стаж роботи. Навіть якщо вашу фірму поки що не звинувачують в дискримінації працівників за ознакою статі, все одно корисно було б виконати множинний регресійний аналіз, щоб незначні (поки що!) проблеми не переросли у великі, вирішувати які буде значно складніше.

б) Якщо ваша фірма бере участь в конкурсі на реалізацію тих або інших проектів, тоді – для тих проектів, конкурс на яких вам вдалося виграти – ви маєте в своєму розпорядженні даних, таких, що стосуються фактичних витрат (Y), оцінки прямих трудовитрат (X_1), оцінки витрат на матеріали (X_2) і витрат на управлінські функції (X_3). Допустимо, що пропозиція ціни, з якою ви виходите на конкурс, здається вам невиправдано низьким. Визначивши взаємозв'язок між фактичними витратами і оцінками, зробленими раніше, на етапі переговорів про висновок контрактів, ви зможете з'ясувати, які з оцінок ви систематично занижуєте або, навпаки, завищуєте (з погляду їх внеску у фактичні витрати).

Друге. *Прогнозування (прогноз) нового спостереження.*

а) Глибоке розуміння структури витрат у вашій фірмі може бути корисне у багатьох відношеннях. Наприклад, у вас може скластися правильніше уявлення про те, які додаткові витрати слід запланувати на сезон підвищеного попиту на продукцію вашої фірми (зокрема, можна врахувати додаткові витрати, пов'язані з виконанням наднормових робіт). Якщо ваш бізнес зазнає певні зміни, ви повинні уміти прогнозувати вплив цих змін на структуру витрат. Краще розбиратися в структурі витрат своєї фірми вам допоможе множинна регресія витрат (Y) на кожен з потенційно значущих (на ваш погляд) чинників, таких як кількість виробів (X_1), що випускаються, кількість працівників (X_2) і об'єм наднормових робіт (X_3). Результати аналізу, подібного цьому, допоможуть вам ухвалювати набагато більш продумані рішення, ніж просте рішення "посадити людей на наднормові роботи на тиждень-другий". Такий аналіз допоможе вам

виявити приховані витрати, які виявляють тенденцію до зростання із зростанням об'ємів наднормових робіт, і робити точніші прогнози фактичних витрат, засновані на інформації, що є у вас.

б) Щомісячні об'єми продажів у вашій фірмі (часовий ряд) можуть пояснюватися сезонними коливаннями попиту. Один із способів аналізу і прогнозування об'ємів продажів полягає у використанні множинної регресії, що дозволяє пояснювати об'єми продажів (Y) на підставі деякого тренда (наприклад, x_1 = 1, 2, 3 ..., вказуючого місяці від початку реєстрації об'ємів продажів) і змінної для кожного місяця (наприклад, x_2 дорівнює 1 для січня і 0 інакше, x_3 R представляє лютий, і так далі). Множинну регресію можна використовувати для прогнозування об'ємів продажів на декілька місяців вперед, а також для з'ясування довгострокових тенденцій і розуміння, в які місяці об'єми продажів, як правило, виявляються більше, ніж в інших.

Третє. *Регулювання і управління процесом.*

На вхід технологічного ланцюжка, використовуваного на целюлозно-паперовому комбінаті, поступає целюлозна маса, а на виході виходить готовий до вживання папір. Як управляти таким складним комплексом устаткування? Одного лише уважного вивчення технічної документації явно недостатньо – щоб навчитися правильно регулювати технологічний процес (з погляду мінімізації витрати електроенергії), потрібні багато років практичного досвіду. Якщо цей досвід виражається в числах, то аналіз множинної регресії дозволяє вам з'ясувати, яка саме комбінація параметрів технологічного процесу (*X-змінні)* дозволяє добитися потрібного результату (змінна *Y*).

Таким чином, прогнозування однієї змінної *Y* на підставі двох або декілька *X-змінних* називається множинною регресією. Цілями множинної регресії є: (1) опис і розуміння відповідного взаємозв'язку, (2) прогнозування (прогноз) нового спостереження, (3) регулювання і управління процесом.

Як виглядатимуть результати множинної регресії? Перш за все, ми приведемо короткий огляд вхідних даних і основних результатів. Докладніше їх пояснення буде дано пізніше.

Хай ***k*** означає кількість пояснюючих змінних (*X*-змінних); ***k*** може бути будь-яким раціональним числом. Значення змінних нерідко називаються *спостереженнями;* це можуть бути клієнти, фірми, вироби, що випускаються, і тому подібне. По "технічних" причинах у вас повинно бути, принаймні, на одне спостереження більше, ніж є *X*-змінних, тобто $n > k+1$. Практичні міркування диктують необхідність набагато більшого числа спостережень.

Вхідні дані для звичайного множинного регресійного аналізу представлені в табл. 4.1.

Таблиця 4.1. Вхідні дані для множинної регресії

	Y (залежна, або з'ясовна, змінна)	**X1 (перша незалежна, або що пояснює, змінна)**	**X2 (друга незалежна, або що пояснює, змінна)**	**X.**	**Xk (остання незалежна, або що пояснює, змінна)**
Спостереження 1	10,9	2,0	4,7	.	12,5
Спостереження 2	23,6	4,0	3,4	.	12,3
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Спостереження *n*	6,0	0,5	3,1	.	7

Зрушення, або **постійний член, *a*,** визначає прогнозоване значення *Y*, коли *всі* змінні *X* рівні 0. **Коефіцієнт регресії** для кожної *X*-*змінної* визначає вплив цієї *X*-змінної на ***Y*** за умови, що всі решта *X*-змінні не міняються: коефіцієнт регресії b_j для j-ої *X*-змінної указує, яке збільшення ***Y*** очікується, коли все *X*-змінні залишаються незмінними, за винятком змінній X_j, яка збільшується на одну одиницю. Узяті разом ці коефіцієнти регресії складають **рівняння прогнозування,** або **рівняння регресії,** вигляду:

$$\textbf{прогнозоване значення } \boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{b_1 X_1} + \boldsymbol{b_2 X_2} + \cdots + \boldsymbol{b_k X_k}, \qquad (4.1)$$

яке можна використовувати в цілях прогнозування або управління. Ці коефіцієнти $(a, b_1, b_2, \dots, b_k)$ зазвичай обчислюються методом найменших квадратів, який мінімізує суму квадратів помилок прогнозування. Як відомо, в основі процедури МНК лежить вирішення системи нормальних рівнянь. Наприклад, для трьох факторній регресії система нормальних рівнянь виглядатиме таким чином:

$$\begin{cases} \sum Y = na + b_1 \sum X_1 + b_2 \sum X_2 \\ \sum Y_1 = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 \\ \sum Y_2 = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 \end{cases} \quad (4.2)$$

Вирішення даної системи не становить великих труднощів, але за наявності персонального комп'ютера знаходження коефіцієнтів $a, b_1, b_2, \dots, b_k$ не передбачає наявності навіть елементарних навиків в області МНК, – всі процеси виконуються в автоматичному режимі.

Як і у разі простої регресії (з єдиною *X*-змінної), **стандартна помилка оцінки, S_e** указує приблизну величину помилок прогнозування. І як у разі простої регресії, R^2 є **коефіцієнтом детермінації,** який указує, який відсоток варіації *Y «*пояснюється» всіма *X*-змінними. В даному випадку мова йде не просто про квадрат коефіцієнта кореляції *Y* з *однією* X-змінної, а про квадрат коефіцієнта кореляції **r** змінної *Y* (фактичних значень) з прогнозами (які обчислюються за допомогою рівняння регресії, знайденого методом найменших квадратів). Такий показник враховує *всі X*-змінні.

Статистичний висновок починається із загальної перевірки, яку називають **F-тестом** (F-test). Мета F-теста полягає в тому, щоб з'ясувати, чи пояснюють *X*-змінні значущу частку варіації *Y*. Якщо ваша регресія *не* є значущою, говорити більше немає про що. Якщо ж регресія виявляється значущою, можна продовжити аналіз статистичних висновків, використовуючи **t-тести для окремих коефіцієнтів регресії,** які показують, наскільки значущим є вплив тієї

або інший *X*-змінної на Y *за умови, що все інші X-змінні залишаються незмінними*. Побудова довірчих інтервалів і перевірки гіпотез для окремого коефіцієнта регресії будуть, звичайно ж, ґрунтуватися на його стандартній помилці. Кожен коефіцієнт регресії має свою стандартну помилку; вони позначаються $S_{b_1}, S_{b_2}, \dots, S_{b_k}$. У табл. 2 приведені результати множинного регресійного аналізу.

Таблиця 4.2. Результати множинного регресійного аналізу

Назва	**Результат**	**Опис**
1	2	3
Зрушення або постійний член	$\boldsymbol{a}$	Прогнозоване значення для *Y*, коли всі значення *X*-змінних рівні 0
Коефіцієнти регресії	$\boldsymbol{b_1, b_2, \dots, b_k}$	Вплив кожної *X*- змінної на *Y* за умови, що всі інші *X*-змінні залишаються незмінними
Рівняння прогнозування, або рівняння регресії	**прогнозоване значення** $\boldsymbol{Y = a + b_1X_1 + b_2X_2 + \cdots + b_kX_k}$	Прогнозоване значення *Y* при заданих значеннях *X*- змінних
Помилки прогнозування, або залишки	**Y – прогнозоване значення Y**	Помилка, що виникає для кожного спостереження в результаті використання рівняння прогнозування замість фактичного значення *Y* для цього спостереження
Стандартна помилка оцінки	$\boldsymbol{S_e}$ або $\boldsymbol{S}$	Приблизна величина помилок прогнозування (типова різниця між фактичним значенням *Y* і його прогнозом виходячи з рівняння регресії)
Коефіцієнт детермінації	$\boldsymbol{R^2}$	Відсоток мінливості *Y*, який пояснюється всією групою *X*-змінних

Продовження таблиці 4.2

1	2	3
F-тест	Значущий або незначущий	Перевіряє, чи може прогноз на основі *X-змінних* як групи бути краще прогнозу на основі простої випадковості; по суті, перевіряє, чи є $\boldsymbol{R^2}$ більшим, ніж у разі відсутності взаємозв'язку між *X*-змінними і $\boldsymbol{Y}$
t-тести для окремих коефіцієнтів регресії	Значущий або незначущий, для кожної *X*- змінної	Перевіряє, чи впливає на $\boldsymbol{Y}$ конкретна *X* змінна за умови, що всі інші *X* змінні залишаються незмінними; цю перевірку виконують тільки тоді, коли F-тест значущий
Стандартні помилки коефіцієнтів регресії	$\boldsymbol{S_{b_1}, S_{b_2}, \ldots, S_{b_k}}$	Указує вибіркову оцінку стандартного відхилення кожного коефіцієнта регресії; використовується звичайним способом для знаходження довірчих інтервалів і перевірки гіпотез для окремих коефіцієнтів регресії
Число ступенів свободи для стандартних помилок коефіцієнтів регресії	$\boldsymbol{m = k - 1}$	Використовується, щоб знайти в t-таблиці відповідне значення для побудови довірчих інтервалів і перевірки гіпотез для окремих коефіцієнтів регресії

Приклад. ***Реклама в журналах***

Тарифи на розміщення рекламних оголошень в журналах визначаються кожним журналом самостійно. Чим пояснюються відмінності в тарифах? Можливо, тут якимось чином враховується цінність рекламного оголошення для рекламодавця. Журнали, що мають в своєму розпорядженні більшу читацьку аудиторію (за рівних інших умов), напевно, мають право встановлювати великі тарифи. Крім того, журнали, розраховані на спроможніші круги читачів, також мають право встановлювати вищі тарифи. Незважаючи на те, що напевно є інші, не менш важливі чинники, ми обмежимося лише вказаними двома, додавши до них ще один – переваги людей різної статі, і з'ясуємо, чи змінюють журнали свої тарифи залежно від співвідношення чоловіків і жінок в їх читацькій аудиторії.

Відповіді на деяких з цих питань можна отримати за допомогою множинного регресійного аналізу. Такий аналіз допоможе нам пояснити вплив на тарифи таких чинників, як величина читацької аудиторії, структура читацької аудиторії відносно статі і доходів читачів.

У табл. 4.3 представлена відповідна багатовимірна сукупність даних, яку нам належить проаналізувати. В якості змінної $\boldsymbol{Y}$ (з'ясовної) ми розглядатимемо вартість однієї сторінки одноразової повнокольорової реклами. Пояснюючими змінними будуть X_1, читацька аудиторія (планована в тисячах чоловік), $\boldsymbol{3_2}$, відсоток чоловіків серед планованої аудиторії, і $\boldsymbol{3_3}$, медіана доходу сім'ї. Розмір вибірки **n** = 55.

Таблиця 4.3. Тарифи на розміщення реклами і характеристики журналів

Назва журналу	**Y, тариф (одна сторінка кольорової реклами), дол.**	X_1, **планована аудиторія, тис. чоловік**	X_2, **відсоток чоловіків**	X_3, **медіана доходу сім'ї, поділ**
1	2	3	4	5
Audubon	25 315	1645	51,1	38 787
Better Homes & Gardens	198 000	34 797	22,1	41933
Business Week	103300	4760	68,1	63 667
Cosmopolitan	94100	15 452	17,3	44 237
Elle	55 540	3735	12,5	47 211
Entrepreneur	40 355	2 476	60,4	47 579
Esquire	51559	3037	71,3	44 715
Family Circle	147 500	24 539	13,0	38 759
first For Women	28 059	3 856	3,6	43 850
Forbes	59 340	4191	68,8	66 606
Fortune	60800	3 891	68,8	58 402
Glamour	85 080	10891	7,8	46331
Goff Digest	98760	6 250	78,9	61323
Good Housekeeping	166 080	25 306	12,6	38 335
Gourmet	49 640	4484	29,6	57 060

Продовження таблиці 4.3

1	2	3	4	5
Harper's Bazaar	52 805	2 621	11,5	44 992
Inc.	70 825	2166	66,9	72493
Kiplinger's Personal Finance	46580	3332	65,1	63 876
Ladies' Home Journal	127 000	17040	6,8	38442
Life	63 750	14 220	46,9	41770
Mademoiselle	55 910	4804	8,0	46694
Martha Stewart's Living	93 328	4 849	16,6	61890
McCalls	113120	16301	7,6	33 823
Money	98 250	9805	60,6	60549
Motor Trend	79 800	5 281	88,5	48 739
National Geographic	159345	32158	53,0	44 326
Natural History	20180	1775	45,0	41499
Newsweek	148 800	20 720	53,5	53 025
Parents Magazine	72 820	12064	18,2	39369
PC Computing	40 675	4606	67,0	57 916
People	125 000	33 668	34,0	46171
Popular Mechanics	78685	9036	86,9	40802
Reader's Digest	193000	51925	42,4	38 060
Redbook	95 785	13 212	8,9	41 156
Rolling Stone	78 920	8 638	59,8	43 212
Runner's World	36 850	2 078	62,9	60 222
Scientific American	37 500	2 704	70,0	62372
Seventeen	71 115	5 738	17,0	37 034
Ski	32 480	2 249	64,5	58 629
Smart Money	42 900	2 224	63,4	57170
Smithsonian	73 075	8 253	47,9	50872
Soap Opera Digest	35 070	7 227	10,3	31835
Sports Illustrated	162 000	21602	78,8	45 897
Sunset	56 000	5 276	38,7	52 524
Teen	53 250	3 057	15,4	42640
The New Yorker	62 435	3 223	48,9	49672
Time	162 000	22 798	52,4	49166
True Story	17100	3582	12,2	15734
TV Guide	146400	40917	42,8	37 396
U.S. News & World Report	98 644	9 825	57,5	52 018
Vanity Fair	67 890	4 307	27,7	52189
Vogue	63 900	8434	12,9	44 242
Woman's Day	137 000	22 747	6,7	38463

Закінчення таблиці 4.3

1	2	3	4	5
Working Woman	87 500	3312	6,3	44 674
YM	73 270	3109	14,4	43 696
Середнє значення	**83 534**	**10913**	**39,7**	**47 710**
Середньоквадратичне відхилення	**45446**	**11212**	**25,9**	**10 225**

У табл. 4.4 представлений комп'ютерний роздрук результатів аналізу множинної регресії. Наприклад, за допомогою Excel можна виконати аналіз множинної регресії. Знайдіть пункт Data Analysis (Аналіз даних) в меню Tools (Сервіс) і виберіть команду Regression (Регресія). Якщо в меню Tools (Сервіс) відсутній пункт Data Analysis (Аналіз даних), то спочатку переконайтеся, що ви вибрали елемент електронної таблиці (а не графік, наприклад). Якщо ви все ж таки не можете знайти Data Analysis (Аналіз даних), пошукайте пункт меню Add-Ins (Надбудови) і поставте відмітку біля Analysis ToolPak (Пакет аналізу). Якщо це не допоможе, то, мабуть, необхідно переустановити Excel.

Таблиця 4.4. Результат множинною регресійного аналізу тарифів на розміщення реклами в журналах (обчислення зроблені в Excel)

ВИВЕДЕННЯ ПІДСУМКІВ

Регресійна статистика	
Множествен. R	0,887
R-квадрат	0,787
Нормований R-квадрат	0,775
Стандартна помилка	21577,870
Спостереження	55

Дисперсійний аналіз

	df	***SS***	***MS***	***F***	***Значущість F***
Регресія	3	87780733202	29260044401	62,843	0,000000
Залишок	51	23745829151	465604493		
Разом	54	111525962353			

Продовження таблиці 4.4.

	Коефіцієнти	*Стандартна помилка*	*t-статистика*	*P-значення*	*Нижні 95%*	*Верхні 95%*
Y-перетин	4042,799	16884,039	0,239	0,812	-29853,298	37938,895
Змінна X 1	3,788	0,281	13,484	0,000	3,224	4,352
Змінна X 2	-123,634	137,849	-0,897	0,374	-400,377	153,108
Змінна X 3	0,903	0,370	2,442	0,018	0,161	1,645

4.2. Інтерпретація результатів багатофакторного моделювання»

Інтерпретація результатів багатофакторної регресії полягає в розумінні впливу різних незалежних змінних (пояснювальних факторів) на залежну змінну (пояснювальна змінна) в контексті обраної регресійної моделі.

Основні елементи для інтерпретації результатів багатофакторної регресії:

1. Коефіцієнти регресії (β).

Кожна незалежна змінна має свій коефіцієнт регресії, який показує, наскільки змінюється залежна змінна при зміні одиниці незалежної змінної, усі інші змінні залишаються незмінними. Позитивний коефіцієнт показує прямий зв'язок між змінними (збільшення однієї призводить до збільшення іншої), а негативний –зворотний зв'язок (збільшення однієї змінної призводить до зменшення іншої).

2. Значущість коефіцієнтів.

При інтерпретації коефіцієнтів важливо враховувати їх статистичну значущість. Значущість вказує на те, наскільки ймовірно, що зв'язок між змінною і залежною змінною не виник випадково. Зазвичай, якщо p-значення менше заданого рівня значущості (зазвичай 0.05), то коефіцієнт вважається статистично значущим.

3. Коефіцієнт детермінації ($\boldsymbol{R^2}$).

$\boldsymbol{R^2}$ вказує на частку варіації залежної змінної, яка пояснюється незалежними змінними у моделі. Вище значення $\boldsymbol{R^2}$ вказує на краще пояснення

залежної змінної змінними моделі. За допомогою коефіцієнту детермінації перевіряють значущість моделі.

4. Інтерпретація взаємодії.

У багатофакторній регресії можуть включатися взаємодію між деякими змінними. Інтерпретація взаємодії полягає в тому, як змінюється вплив однієї змінної на залежну змінну в залежності від значень іншої змінної.

5. Оцінка прогнозів.

Багатофакторна регресія дозволяє використовувати модель для прогнозування значень залежної змінної на основі значень незалежних змінних. Важливо пам'ятати про обмеження прогнозів, особливо за межами діапазону досліджених даних.

Інтерпретація результатів багатофакторної регресії має бути узгоджена з економічною теорією та контекстом дослідження, щоб зробити достовірні та зрозумілі висновки про вплив різних факторів на залежну змінну.

4.3. Коефіцієнти регресії і рівняння регресії

Зрушення, або постійний член, a, і коефіцієнти регресії, b_1, b_2, *і* b_3, обчислюються комп'ютером з використанням методу найменших квадратів.

Серед всіх можливих варіантів рівняння регресії з різними значеннями цих коефіцієнтів саме рівняння, знайдене таким методом, забезпечує мінімальну суму квадратів помилок прогнозування для вибірки, що розглядається нами, журналів. Рівняння регресії (або рівняння прогнозування) має наступний вигляд:

$$\textbf{(прогнозований тариф на розміщення реклами)} = a + b_1X_1 + b_2X_2 + b_3X_3 =$$

$$\textbf{= \$4043 + 3,79(читацька аудиторія) – 124(відсоток чоловіків)+ 0,903(медіана доходу).}$$

Зрушення, ***a* = $4 043**, інтерпретується таким чином: типовий тариф на розміщення односторінкового кольорового рекламного оголошення в журналі, у якого немає платних підписчиків, немає чоловіків серед читачів і читачі не мають доходу, складає $4043. Проте в сукупності даних, що розглядається нами, немає подібних журналів, тому зрушення, ***a,*** слід розглядати лише як допоміжну величину, необхідну для отримання оптимальних прогнозів, але не інтерпретувати це значення так буквально.

Коефіцієнти регресії інтерпретуються як вплив кожної зі змінних на розмір тарифу, якщо всі інші незалежні («змінні, що пояснюють») змінні залишаються незмінними. Часто це значення включає «поправку на» інші незалежні змінні, або «контроль» цих інших незалежних змінних. Тому коефіцієнт регресії для конкретної *X-змінної* може змінюватися (іноді значно) в результаті включення в аналіз або виключення інших *X*-змінних. Зокрема, кожен коефіцієнт регресії визначає середнє збільшення тарифу на розміщення реклами, що доводиться на одиничне збільшення відповідною йому *X*-змінної (в даному випадку термін «одиничне» означає одну одиницю вимірювання конкретної *X*-змінної).

Коефіцієнт регресії для розміру читацької аудиторії, b_1**=3,79**, указує, що – *за всіх інших рівних умов* – журнал з додатковою тисяччю читачів (оскільки у нас x_1 вимірюється в тисячах чоловік) бере (в середньому) на $3,79 більше розміщення односторінкового кольорового рекламного оголошення. Можна також вважати, що коефіцієнт регресії для розміру читацької аудиторії означає, що кожен додатковий читач збільшує для цього журналу тариф на розміщення рекламних оголошень на $0,00379, тобто збільшення складає трохи менше половини цента на одну людину. Тому, якщо у якогось іншого журналу такий же відсоток читачів-чоловіків і такий же показник медіани доходу сім'ї читачів, але читацька аудиторія на 3548 чоловік більша, то можна чекати, що тариф на розміщення рекламних оголошень в цьому журналі буде (в середньому) на 3,79×3,548 = $13,45 більше завдяки такій відмінності розміру читацької аудиторії.

Коефіцієнт регресії для відсотка чоловіків, b_2 = **-124**, указує, що (за всіх інших рівних умов) тариф на розміщення кольорових рекламних оголошень в журналі з додатковим 1% читачів-чоловіків опиниться (в середньому) на $124 менше. Це означає, що читачки представляють для журналу велику цінність, чим читачі-чоловіки. Статистичний висновок повинен підтвердити або спростувати цю гіпотезу шляхом порівняння величини впливу відсотка чоловіків (тобто -$124) з тим, на що можна було б розраховувати, якби за даних обставин все визначалося лише чистою випадковістю.

Коефіцієнт регресії для медіани доходу, b_3 = **0,903**, указує, що (за всіх інших рівних умов) в журналі з додатковим доларом медіани доходу його читачів тариф на розміщення односторінкового кольорового рекламного оголошення буде (в середньому) на $0,903 більше. Позитивний знак цього коефіцієнта абсолютно виправданий, оскільки люди з вищим рівнем доходів можуть дозволити собі витрачати більше на покупку рекламованої продукції. Якщо у якогось іншого журналу такий же відсоток читачів-чоловіків і така ж величина читацької аудиторії, але медіана доходу сімей читачів на $4 000 вища, то можна чекати, що тариф цього журналу на розміщення рекламних оголошень буде на 0,903×4000 = $3612 вище (в середньому) завдяки вищому рівню доходів його читачів.

Пам'ятаєте, що коефіцієнти регресії відображають вплив на Y одній X-змінної за умови, що все інші X-змінні залишаються незмінними. Це слід розуміти буквально.

Наприклад, коефіцієнт регресії b_3 відображає вплив медіани доходу читачів на рекламні тарифи; він обчислюється при незмінних величинах читацької аудиторії і відсотка читачів-чоловіків. У такому разі вищі рівні доходів читачів, як правило, ведуть до встановлення вищих тарифів на розміщення рекламних оголошень (оскільки b_3 є позитивним числом) – при фіксованих розмірі читацької аудиторії і відсотку читачів-чоловіків.

Яким був би цей взаємозв'язок, якби решта змінних (розмір читацької аудиторії і відсоток читачів-чоловіків) *не* фіксувалася на постійному рівні? На це

питання можна відповісти, проаналізувавши звичайний коефіцієнт кореляції (або коефіцієнт регресії, прогнозуючий *Y* на підставі тільки одній цій *X*-змінної), обчислений тільки для двох змінних: тарифу і медіани доходу. У нашому випадку вище значення медіани доходу фактично асоціюється з *нижчим* тарифом (кореляція тарифу і медіани доходу є негативною: -0,167)! Чим це пояснити? Цілком прийнятне пояснення полягає в тому, що журнали, що орієнтуються на читачів з вищим середнім рівнем доходів, не в змозі забезпечити собі масову аудиторію через те, що багатих людей серед населення країни в цілому не так вже багато. Якщо ж ця читацька аудиторія багатих людей виявиться дуже невеликою, це може взагалі спотворити ефект впливу високого рівня доходів з розрахунку на одного читача.

4.4. Приклад прогнозів

Рівняння прогнозування, або рівняння регресії, визначається в наступному вигляді:

$$\textbf{прогнозоване значення Y} = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k$$

У нашому прикладі з рекламними оголошеннями в журналах, щоб знайти прогнозовану величину тарифу на розміщення рекламних оголошень виходячи з величини читацької аудиторії, відсотка читачів-чоловіків і медіани доходу читачів для конкретного журналу, подібного тим, які складають вибірку даних, що розглядається нами, підставимо в рівняння прогнозування відповідні цьому журналу значення *X*-змінних:

(прогнозований тариф на розміщення реклами) $= a + b_1X_1 + b_2X_2 + b_3X_3$**= $4043 + 3,79(читацька аудиторія) – 124(відсоток чоловіків)+ 0,903(медіана доходу).**

Допустимо, наприклад, що ви збираєтеся заснувати новий журнал, «*Популярна статистика»,* який розрахований на читацьку аудиторію порядка 900000 чоловік, 55% яких складатимуть жінки, а медіана доходу його читачів рівна $50000. Дані в рівняння прогнозування необхідно підставити в тій же формі, що і в початковій сукупності даних (тобто тій, виходячи з якої і будувалося рівняння регресії): x_1 = 900 (читацька аудиторія в тисячах чоловік), x_2 = 45 (відсоток чоловіків) і x_3 = $50000 (медіана доходу). Прогнозоване значення для цієї ситуації визначається таким чином:

прогнозований тариф на розміщення реклами в журналі «*Популярна статистика»* =

= 4043 +3,79(читацька аудиторія) –124(відсоток чоловіків) +0,903(медіана доходу) =

$$= 4043 + 3{,}79 \times 900 - 124 \times 45 + 0{,}903 \times 50000 = \$47024.$$

Зрозуміло, розраховувати на те, що тариф на розміщення реклами в журналі складе рівно $47024, не доводиться. По-перше, навіть між журналами, даними про яких ми розташовуємо, спостерігаються випадкові коливання, тому прогнози не є ідеальними навіть для них. По-друге, прогнози можуть бути корисні лише в тій мірі, в який прогнозований журнал подібний до журналів, що належать до початкової сукупності даних. Якщо мова йде про новий журнал, то тариф на розміщення реклами в цьому журналі може визначатися не так, як для журналів із вже сталою репутацією, які ми використовували для побудови рівняння регресії.

За допомогою цього рівняння можна також прогнозувати тарифи для журналів, що належать до початкової сукупності даних. У першого журналу, «*Audubon»* x_1 = 1645 (читацька аудиторія рівна приблизно 1,6 мільйона чоловік), x_2 = 51,1 (тобто 51,1% читачів цього журналу — чоловіки) і x_3 = 38787 (медіана річного доходу читачів цього журналу складає $38787). Прогнозоване значення для цього журналу можна знайти по наступній формулі:

прогнозований тариф на розміщення реклами в журналі «*Audubon*» =

= 4043 +3,79(читацька аудиторія) –124(відсоток чоловіків)

+0,903(медіана доходу) =

= 4043 + 3,79 × 1 645 – 124 × 51,1 + 0,903 × 38787 = $38966.

Залишок, або помилка прогнозування, визначається за формулою: **Y – (прогнозоване значення Y)**.

Для журналу, що належить до початкової сукупності даних, цей показник дорівнює фактичному тарифу мінус прогнозований тариф. Для журналу «*Audubon*» фактичний тариф складає $25315, а прогнозований тариф – $38966. Таким чином, помилка прогнозування рівна 25315 – 38966 = –$13651. Негативний залишок указує на те, що фактичний тариф менше прогнозованого (у разі журналу *Audubon* приблизно на $14000). Для багатьох з нас $14000 – величезні гроші; непогано б поглянути на інші помилки прогнозування, щоб зрозуміти, якою мірою прогнозування відображає реальну ситуацію. Чому рекламні тарифи в журналі *Audubon* опинилися набагато менше їх очікуваної величини? Швидше за все тому, що для прогнозування використовувалося лише $\boldsymbol{k = 3}$ з безлічі можливих чинників, що впливають на величину рекламних тарифів (до того ж багато хто з цих чинників не дуже зрозумілий і їх досить складно зміряти).

4.5. Статистичні висновки за багатофакторною моделлю

Наскільки хороші наші прогнози? Цей розділ слід розглядати в основному як огляд, оскільки стандартне відхилення оцінки, S_e**,** і коефіцієнт детермінації, R^2, мають для множинної регресії, взагалі кажучи, ту ж інтерпретацію, що і для простій регресії. Єдина відмінність полягає в тому, що ваші прогнози тепер базуються на декількох X-змінних. Але все залишається дуже схоже, оскільки ви як і раніше прогнозуєте тільки одну змінну Y.

Типова помилка прогнозування: стандартна помилка прогнозу.

Як і у разі простої регресії, коли ми маємо справу лише з однією *X-змінною*, стандартна помилка оцінки (прогнози) указує приблизну величину помилок прогнозування.

Повертаючись до нашого прикладу з тарифами на розміщення реклами в журналах, S_e = \$21578. Це говорить про те, що фактичні тарифи на розміщення реклами в цих журналах, як правило, відхиляються від прогнозованих тарифів не більше ніж на \$21578 (мова йде про стандартному відхиленні). Іншими словами, якщо розподіл помилок є нормальним, то можна чекати, що приблизно 2/3 фактичні тарифи знаходитимуться в межах S_e від прогнозованих тарифів; приблизно 95% – в межах $2S_e$ і так далі

Ця стандартна помилка оцінки, S_e = \$21578, указує залишок варіації тарифів після того, як ви використовували *X*-змінні (величина читацької аудиторії, відсоток чоловіків і медіана доходу) в рівнянні регресії для прогнозування тарифів кожного журналу. Порівняйте цей показник із звичайним стандартним відхиленням однієї змінної для тарифів, S_Y = \$45446, обчисленим без використання інших змінних. Це стандартне відхилення, S_Y, указує залишок варіації тарифів після того, як ви використовували для прогнозування тарифів кожного журналу тільки значення Y. Відмітьте, що S_e = \$21578 менше, ніж S_Y = \$45446; помилки, як правило, виявляються менше, якщо для прогнозування тарифів використовувати рівняння регресії, а не просто $\bar{Y}$. Як бачите, *X*- змінні корисні для пояснення розмірів тарифів.

Це можна уявити собі таким чином. Якщо вам нічого невідомо про X-змінні, ви використовуватимете як оптимальну приблизну оцінку середнє значення тарифу ($\bar{Y}$ = \$83534) і помилятиметеся приблизно на S_Y = \$45446. Але якщо вам відомі такі характеристики, як величина читацької аудиторії, відсоток чоловіків і середній дохід, то для прогнозування тарифів можна скористатися рівнянням регресії; в цьому випадку ви помилитеся приблизно на S_e = \$21578. Таке скорочення помилки прогнозування (з \$45446 до \$21578) і є одним з переваг використання регресійного аналізу.

Пояснений відсоток варіації: R^2

Коефіцієнт детермінації (часто також використовують термін «квадрат множинної кореляції»), R^2**, показує**, який відсоток варіації Y пояснюється впливом всіх X- змінних.

Якщо повернутися до нашого прикладу з тарифами на розміщення реклами в журналах, то коефіцієнт детерміації, R^2 = 0,787, або 78,7%, указує на те, що незалежні змінні (X-змінні величини читацької аудиторії, відсоток чоловіків і середній дохід) пояснюють 78,7% варіацій тарифів. При цьому 21,3% залишаються непоясненими і зв'язуються з впливом інших чинників. 78,7% – досить велике значення R^2; у багатьох дослідженнях доводиться працювати із значно меншими величинами, які, проте, забезпечують достатньо якісні прогнози. Бажано, щоб значення R^2 було як можна великим (великі значення R^2 свідчать про те, що досліджуваний взаємозв'язок є достатньо сильним). У ідеальному випадку $R^2 = 100\%;$ це можливо лише у тому випадку, коли всі помилки прогнозування рівні 0 (що, як правило, свідчить про наявність помилок у іншому місці!).

Статистичний висновок у разі множинної регресії: ***F-тест***

Отримані нами до теперішнього часу результати регресії є достатньо повним описом досліджуваних ***(n*** = 55) журналів, проте статистичний висновок допоміг би нам узагальнити цей випадок на популяцію подібних ним журналів, що ідеалізувалася. Замість того щоб просто констатувати той факт, що збільшення на один відсоток числа читачів-чоловіків приводить до зменшення тарифу на розміщення реклами в середньому на $124, можна зробити статистичний висновок щодо великої генеральної сукупності журналів такого типу, з якої цілком могли б витягувати наявні дані, і спробувати з'ясувати, чи існує *насправді* який-небудь взаємозв'язок між статтю читачів журналу і тарифами на рекламу, або коефіцієнт регресії, рівний -$124, можна пояснити просто випадковістю. Чи може бути так, що виявлений нами вплив відсотка читачів-чоловіків на вартість реклами – це просто випадкове число, а не

свідоцтво наявності систематичному взаємозв'язку? Відповідь на це питання можна отримати за допомогою статистичного висновку.

Щоб не ускладнювати приклад, припустимо, що ми маємо в своєму розпорядженні випадкову вибірку з набагато більшої генеральної сукупності. Допустимо також, що ця генеральна сукупність характеризується лінійним взаємозв'язком з випадковістю, представленою моделлю **множинної лінійної регресії,** відповідно до якої спостережуване значення Y *визначається* взаємозв'язком в генеральній сукупності плюс нормально розподілена випадкова помилка. Передбачається також, що ці випадкові помилки для різних спостережень (елементарних одиниць наших даних) не залежать один від одного.

Модель множинною регресій для генеральної сукупності:

$$\boldsymbol{Y = (б + в_1 x_1 + в_2 x_2 + \ldots + в_k x_k) + е}$$

= (взаємозв'язок в генеральній сукупності) + випадковість

де **е** характеризується нормальним розподілом з середнім значенням 0 і постійним стандартним відхиленням у**,** причому ця випадковість є незалежною для кожного із спостережень (елементарних одиниць даних).

Взаємозв'язок в генеральній сукупності визначається $\boldsymbol{k + 1}$ параметрами: **б** представляє зрушення (або постійний член) для генеральної сукупності, а $в_1, в_2, \ldots в_k$ є коефіцієнтами регресії для генеральної сукупності, які показують середній вплив кожної з *X*-змінних на Y (у даній генеральній сукупності), за умови, що всі решта *X*-змінні залишаються незмінними. Якби ви мали дані про всю генеральну сукупність, то отримані вами за допомогою методу найменших квадратів коефіцієнти регресії нічим не відрізнялися б від відповідних коефіцієнтів, що описують зв'язок в генеральній сукупності. Як правило, проте, отримане методом найменших квадратів зрушення **а** є лише статистичною *оцінкою* ***б,*** а отримані методом найменших квадратів коефіцієнти регресії $\boldsymbol{b_1, b_2, \ldots, b_k}$ являють собою лише статистичні *оцінки* $в_1, в_2, \ldots, в_k$ відповідно. Існують,

звичайно ж, помилки, обумовлені процесом оцінювання, оскільки вибірка набагато менше всієї генеральної сукупності.

Чи значуща модель? Статистичний висновок починається з *F-теста*, метою якого є з'ясування, чи пояснюють *X*-змінні значущу частину варіації *Y*. *F-тест* використовується як «вхідні ворота» в статистичний висновок: якщо цей тест значущий, отже, зв'язок існує і можна приступати до її дослідження і пояснення. Якщо цей тест незначущий, то ми маємо справу з набором не зв'язаних між собою випадкових чисел – пояснювати, по суті, нічого. Пам'ятаєте, що, коли ви приймаєте нульову гіпотезу, це вважається *слабким* висновком. Ви не довели, що взаємозв'язку немає: вам просто не вистачає переконливих доводів на користь наявності такого взаємозв'язку. Взаємозв'язок цілком може існувати, але через випадковість або малий розмір вибірки ви не в змозі виявити її за допомогою тих даних, які є у вашому розпорядженні.

Нульова гіпотеза для *F-теста* стверджує, що в генеральній сукупності між *X*-змінними і *Y* прогнозуючий взаємозв'язок *відсутній*. Інакше кажучи, *Y* є чисто випадковою величиною і значення *X*-змінних не мають на *Y* ніякого впливу. Якщо подивитися на модель множинної лінійної регресії, то це твердження означає, що $\boldsymbol{Y = б + е}$, що може мати місце в тому випадку, якщо *всі* коефіцієнти регресії в генеральній сукупності рівні 0.

Альтернативна гіпотеза *F-теста* стверджує, що в генеральній сукупності між *X*-змінними і *Y* існує певний прогнозуючий взаємозв'язок. Таким чином, змінна *Y* вже не є чисто випадковою величиною і повинна залежати принаймні від однієї з *X*- змінних. Іншими словами, альтернативна гіпотеза стверджує, що *принаймні один* з коефіцієнтів регресії не рівний 0. Зверніть увагу: зовсім не обов'язково, щоб кожна з *X*-змінних впливала на *Y* – достатньо, щоб впливала хоч би одна з них.

У *F-тесті* використовуються наступні статистичні гіпотези:

$\boldsymbol{H_0}$**:** $в_1 = в_2 = \ldots = в_k = 0$**;**

$\boldsymbol{H_1}$**:** принаймні один з коефіцієнтів регресії $в_1, в_2, \ldots, в_k \neq 0$**.**

Виконати *F-тест* найпростіше, відшукавши в результатах роботи комп'ютерної програми відповідне ***p-значення*** і інтерпретуючи результуючий рівень значущості. Якщо ***p-значення*** більше, ніж 0,05, то отриманий результат не є значущим. Якщо ж це ***p-значення*** менше, ніж 0,05, то отриманий результат є значущим. Якщо $\boldsymbol{p < 0{,}01}$, тоді отриманий результат є високо значущим, і так далі.

Ще один спосіб виконання *F-теста* полягає в порівнянні значення $\mathbf{R^2}$ (відсоток варіації *Y*, який пояснюється *X*-змінними) із значеннями з таблиці критичних значень $\mathbf{R^2}$ *для* відповідного рівня тестування (наприклад, 5%). Якщо значення $\mathbf{R^2}$ *виявляється* достатньо великим, тоді регресія вважається значущою, тобто вдалося пояснити більше, ніж просто випадкову величину варіації Y. Ця таблиця індексована по n *(*кількість спостережень) і k *(*кількість X-змінних).

Традиційний спосіб виконання *F-теста* інтерпретувати дещо складніше, але він завжди дає той же результат, що і таблиця критичних значень $\mathbf{R^2}$. Класичний ***F-тест***, як правило, виконується шляхом обчислення ***F-статистики*** і порівняння її з критичним значенням з *F-таблиці* для відповідного рівня тестування. При цьому використовуються два різних числа ступенів свободи: число ступенів свободи $\boldsymbol{k_1}$ (кількість *X*-змінних, призначених для пояснення *Y* або кількість параметрів в рівнянні регресії мінус одиниця, тобто $\boldsymbol{k_1 = m - 1}$) і число мір свободи $\boldsymbol{k_2 = n - m}$ (де. $\boldsymbol{n}$ – кількість спостережень у вибірці, а $\boldsymbol{m}$ – кількість параметрів в рівнянні регресії).

В той же час ***F-статистика*** є зайвим ускладненням, оскільки значення $\mathbf{R^2}$ можна перевірити безпосередньо. Більш того, $\mathbf{R^2}$ має більш безпосередню інтерпретацію, ніж ***F-статистика***, оскільки $\mathbf{R^2}$ говорить про ту частину варіації *Y*, яка враховується (або пояснюється) *X*-змінними, тоді як ***F*** не має такої простої і безпосередньої інтерпретації в термінах початкових даних. Який би підхід – ***F*** або $\mathbf{R^2}$ – ви не використали, відповідь (про значущість або не значущість) завжди буде одним і тим же на будь-якому рівні тестування.

Чому ж за традицією використовується складніша ***F-статистика***, тоді як замість неї можна було б звернутися до тесту $\mathbf{R^2}$, що допускає зручнішу і безпосередню інтерпретацію? Можливо, все пояснюється традицією, що саме склалася, а можливо, і тим, що вже давно і з успіхом на практиці застосовуються саме *F-таблиці*. Використання осмисленого числа (такого як $\mathbf{R^2}$) дозволяє глибше зрозуміти досліджувану ситуацію і має свої переваги, особливо коли мова йде про сфері бізнесу.

Результат F-теста (рішення ухвалюється на основі *p-значення)*

Якщо ***p-значення*** більше, ніж 0,05, значить, відповідна модель не *є значущою* (ви приймаєте нульову гіпотезу про те, що *X*-змінні не допомагають прогнозувати *Y*). *Якщо* ***p-значення*** виявляється менше, ніж 0,05, значить, відповідна модель є значущою (ви відкидаєте нульову гіпотезу і приймаєте альтернативну гіпотезу про те, що *X*-змінні допомагають прогнозувати *Y*).

Результат F-теста (рішення ухвалюється на основі $\mathbf{R^2}$)

Якщо значення $\mathbf{R^2}$ менше, ніж критичне значення в таблиці $\mathbf{R^2}$, значить, відповідна модель не є *значущою*. Якщо значення $\mathbf{R^2}$ більше, ніж критичне значення в таблиці $\mathbf{R^2}$, значить, відповідна модель *є значущою*. Ця відповідь у будь-якому випадку буде такою ж, як результат, отриманий за допомогою ***p-значення.***

Результат F-теста (рішення ухвалюється на основі критерію *F*)

Якщо значення ***F*** виявляється менше, ніж критичне значення в ***F-таблиці***, значить, відповідна модель не є значущою. *Якщо значення* ***F*** виявляється більше, ніж критичне значення в ***F-таблиці***, то відповідна модель *є значущою*. Ця відповідь у будь-якому випадку буде такою ж, як результат, отриманий за допомогою ***p-значення*** або $\mathbf{R^2}$.

Пам'ятайте, що статистичний сенс терміну «*значущий»* декілька відрізняється від його буденного сенсу. Коли ви знаходите значущу модель регресії, то знаєте, що взаємозв'язок між *X*-змінними і *Y* виявляється сильніше,

ніж зазвичай можна було б чекати від чистої випадковості. Іншими словами, в цій ситуації можна говорити про наявність певного взаємозв'язку. Цей взаємозв'язок може бути сильним або корисним в тому або іншому практичному сенсі (а може, і не бути таким) – ці питання вимагають спеціального розгляду, – але він достатньо сильний, щоб не виглядати як чиста випадковість.

Якщо повернутися до нашого прикладу з тарифами на розміщення реклами в журналах, то відповідне рівняння прогнозування дійсно пояснює значущу частку відхилення в тарифах, на що указує в результатах роботи комп'ютерної програми ***p-значення*** 0,000000 праворуч від значення ***F***, рівного 62,843. У табл. 4.5 міститься частина результатів роботи комп'ютерної програми, приведених вище.

Таблиця 4.5. Результат множинною регресійного аналізу тарифів на розміщення реклами в журналах

ВИВЕДЕННЯ ПІДСУМКІВ						
Регресійна статистика						
Множинний R	0,887					
R-квадрат	0,787					
Нормований R-квадрат	0,775					
Стандартна помилка	21577,870					***p-значення***
Спостереження	55					
Дисперсійний аналіз						
	df		***SS***	***MS***	***F***	***Значущість F***
Регресія	3		87780733202	29260044401	62,843	0,000000
Залишок	51		23745829151	465604493		
Разом	54		111525962353			

Це говорить про те, що дійсно виявляється стійка залежність тарифів від цих чинників (або принаймні від одного з цих чинників), тобто тарифи не є чисто випадковими величинами. Вам як і раніше невідомо, які саме з цих *X*-змінних

реально беруть участь в прогнозуванні *Y*, але вам напевно відомо, що є принаймні одна така змінна.

Щоб з'ясувати за допомогою $\mathbf{R^2}$, чи дійсне рівняння регресії є значущим, відзначимо, що коефіцієнт детерміації $\mathbf{R^2}$ = 0,787, або 78,7%. Таблиця $\mathbf{R^2}$ для тестування на рівні 5% у разі ***n=55*** журналів і ***k=3*** змінних (табл. 4.6) дає критичне значення 0,141, або 14,1%. Для того, щоб рівняння було значущим на звичному рівні 5%, *X*-змінні повинні пояснювати лише 14,1% варіацій тарифів (Y). Оскільки вони пояснюють більше, регресію слід визнати значущою.

Таблиця 4.6. Таблиця R^2 : критичні значення для рівня 5% (значущість)

Кількість спостережень (*n*)	**Кількість *X-змінних (k)***									
	1	2	3	4	5	6	7	8	9	10
3	0,994									
4	0,902	0,997								
5	0,771	0,950	0,998							
6	0,658	0,864	0,966	0,999						
7	0,569	0,776	0,903	0,975	0,999					
8	0,499	0,698	0,832	0,924	0,980	0,999				
9	0,444	0,632	0,764	0,865	0,938	0,983	0,999			
10	0,399	0,575	0,704	0,806	0,887	0,947	0,985	0,999		
11	0,362	0,527	0,651	0,751	0,835	0,902	0,954	0,987	1,000	
12	0,332	0,486	0,604	0,702	0,785	0,856	0,914	0,959	0,989	1,000
13	0,306	0,451	0,563	0,657	0,739	0,811	0,872	0,924	0,964	0,990
14	0,283	0,420	0,527	0,618	0,697	0,768	0,831	0,885	0,931	0,967
15	0,264	0,393	0,495	0,582	0,659	0,729	0,791	0,847	0,896	0,937
16	0,247	0,369	0,466	0,550	0,624	0,692	0,754	0,810	0,860	0,904
17	0,232	0,348	0,440	0,521	0,593	0,659	0,719	0,775	0,825	0,871
18	0,219	0,329	0,417	0,494	0,564	0,628	0,687	0,742	0,792	0,839
19	0,208	0,312	0,397	0,471	0,538	0,600	0,657	0,711	0,761	0,807
20	0,197	0,297	0,378	0,449	0,514	0,574	0,630	0,682	0,731	0,777
21	0,187	0,283	0,361	0,429	0,492	0,550	0,604	0,655	0,703	0,749
22	0,179	0,270	0,345	0,411	0,471	0,527	0,580	0,630	0,677	0,722
23	0,171	0,259	0,331	0,394	0,452	0,507	0,558	0,607	0,653	0,696
24	0,164	0,248	0,317	0,379	0,435	0,488	0,538	0,585	0,630	0,673
25	0,157	0,238	0,305	0,364	0,419	0,470	0,518	0,564	0,608	0,650
26	0,151	0,229	0,294	0,351	0,404	0,454	0,501	0,545	0,588	0,629
27	0,145	0,221	0,283	0,339	0,390	0,438	0,484	0,527	0,569	0,609
28	0,140	0,213	0,273	0,327	0,377	0,424	0,468	0,510	0,551	0,590
29	0,135	0,206	0,264	0,316	0,365	0,410	0,453	0,495	0,534	0,573
30	0,130	0,199	0,256	0,306	0,353	0,397	0,439	0,480	0,518	0,556
31	0,126	0,193	0,248	0,297	0,342	0,385	0,426	0,466	0,503	0,540

Продовження таблиці 4.6

Кількість спостережень (*n*)	**Кількість *X-змінних (k)***									
	1	2	3	4	5	6	7	8	9	10
32	0,122	0,187	0,240	0,288	0,332	0,374	0,414	0,452	0,489	0,525
33	0,118	0,181	0,233	0,279	0,323	0,363	0,402	0,440	0,476	0,511
34	0,115	0,176	0,226	0,271	0,314	0,353	0,391	0,428	0,463	0,497
35	0,111	0,171	0,220	0,264	0,305	0,344	0,381	0,417	0,451	0,484
40	0,097	0,150	0,193	0,232	0,268	0,303	0,336	0,368	0,399	0,429
50	0,078	0,120	0,155	0,186	0,216	0,244	0,272	0,298	0,323	0,348
51	0,076	0,117	0,152	0,183	0,212	0,240	0,267	0,293	0,318	0,342
52	0,075	0,115	0,149	0,180	0,208	0,235	0,262	0,287	0,312	0,336
53	0,073	0,113	0,146	0,176	0,204	0,231	0,257	0,282	0,306	0,330
54	0,072	0,111	0,143	0,173	0,201	0,227	0,252	0,277	0,301	0,324
55	0,071	0,109	0,141	0,170	0,197	0,223	0,248	0,272	0,295	0,318
56	0,069	0,107	0,138	0,167	0,194	0,219	0,244	0,267	0,290	0,313
57	0,068	0,105	0,136	0,164	0,190	0,215	0,240	0,263	0,285	0,308
58	0,067	0,103	0,134	0,161	0,187	0,212	0,236	0,258	0,281	0,303
59	0,066	0,101	0,131	0,159	0,184	0,208	0,232	0,254	0,276	0,298
60	0,065	0,100	0,129	0,156	0,181	0,205	0,228	0,250	0,272	0,293
Множник 1	3,84	5,99	7,82	9,49	11,07	12,59	14,07	15,51	16,92	18,31
Множник 2	2,15	-0,27	-3,84	-7,94	-12,84	-18,24	-23,78	-30,10	-36,87	-43,87

Якщо у вас більше 60 спостережень, критичні значення можна знайти за допомогою двох множників, вказаних внизу таблиці $\mathbf{R^2}$. Для цього необхідно скористатися наступною формулою:

Критичне значення = *(Множник 1 / n) + (Множник 2 / n)*

Коли як ***p-значення*** (*Значущість F*) указується 0,000000 (див. табл. 4.5), його можна інтерпретувати як $p < 0{,}0005$, оскільки ***p-значення***, яке більше або рівне 0,0005, буде закруглено до 0,001. Використовуючи термінологію ***p-значення***, можна сказати, що регресія в даному випадку є дуже високо значущою ($p < 0{,}001$).

Щоб переконатися в цьому дуже високому рівні значущості, використовуючи безпосередньо ***F-тест,*** можна порівняти *F-статистику* 62,843 (з комп'ютерного роздруку) із значенням з F-таблиці для рівня 5% (табл. 4.7), яке знаходиться між 2,922 і 2,758 для $k_1 = m - 1 = 3$ мір свободи і $k_2 = n - m = 55 - 4 = 51$ мір свободи.

Таблиця 4.7. F-таблиця: критичні значення для рівня значущості 5%

Ступені свободи (k_2)	**Ступені свободи (k_1)**					
	1	**2**	**3**	**4**	**5**	**6**
1	161,45	199,50	215,71	224,58	230,16	233,99
2	18,513	19,000	19,164	19,247	19,296	19,330
3	10,128	9,552	9,277	9,117	9,013	8,941
4	7,709	6,944	6,591	6,388	6,256	6,163
5	6,608	5,786	5,409	5,192	5,050	4,950
6	5,987	5,143	4,757	4,534	4,387	4,284
7	5,591	4,737	4,347	4,120	3,972	3,866
8	5,318	4,459	4,066	3,838	3,687	3,581
9	5,117	4,256	3,863	3,633	3,482	3,374
10	4,965	4,103	3,708	3,478	3,326	3,217
11	4,840	3,980	3,590	3,360	3,200	3,090
12	4,747	3,885	3,490	3,259	3,106	2,996
15	4,543	3,682	3,287	3,056	2,901	2,780
18	4,410	3,550	3,160	2,930	2,770	2,660
19	4,380	3,520	3,130	2,900	2,740	2,630
20	4,351	3,493	3,098	2,866	2,711	2,599
21	4,32	3,47	3,07	2,84	2,68	2,57
22	4,30	3,44	3,05	2,82	2,66	2,55
23	4,28	3,42	3,03	2,80	2,64	2,53
24	4,26	3,40	3,01	2,78	2,62	2,51
25	4,24	3,38	2,99	2,76	2,60	2,49
26	4,22	3,37	2,98	2,74	2,59	2,47
27	4,21	3,35	2,96	2,73	2,57	2,46
28	4,20	3,34	2,95	2,71	2,56	2,44
29	4,18	3,33	2,93	2,70	2,54	2,43
30	4,171	3,316	**2,922**	2,690	2,534	2,421
60	4,001	3,150	**2,758**	2,525	2,368	2,254
120	3,920	3,072	2,680	2,447	2,290	2,175'
∞	3,841	2,996	2,605	2,372	2,214	2,099

Оскільки значення 51 в таблиці відсутній, нам відомо, що необхідне нам значення з *F-таблиці* знаходиться в діапазоні від **2,922** для 30 мір свободи знаменника і для 60 мір свободи знаменника. Оскільки дана ***F-статистика*** (62,843) набагато більше, ніж значення з *F-таблиці* (значення з діапазону від **2,758** до **2,922**), ми знову приходимо до висновку, що отриманий результат має дуже високу значущість.

Які змінні є значущими: t-тест для кожного коефіцієнта

Якщо *F-тест* є значущим, то вам відомо, що одна або декілька *X*-змінних можуть бути корисні в прогнозуванні *Y* і, отже, можна продовжувати аналіз за допомогою *t-тестів* для окремих коефіцієнтів регресії з метою з'ясовувати, які саме з *X*-змінних дійсно корисні. Ці *t-тести* визначають, чи робить значущий вплив на *Y* та або інша *X*-змінна, якщо все інші *X*-змінні залишаються при цьому незмінними. Слід пам'ятати, що, прийнявши нульову гіпотезу, ви зробили *слабкий* висновок і, по суті, тим самим не довели даремність *X*-змінної, а просто у вас не вистачило переконливих доказів наявності взаємозв'язку. Таким чином, взаємозв'язок може існувати, але унаслідок дії чинника випадковості або через невеликий розмір вибірки ви не в змозі виявити його за допомогою тих даних, які є у вашому розпорядженні.

Якщо ж *F-тест* не є значущим, то використовувати *t-тести* для окремих коефіцієнтів регресії не можна. У окремих випадках ці *t-тести* можуть бути значущими навіть тоді, коли *F-тест* не є значущим. При цьому *F-тест* вважається важливішим і необхідно робити висновок про те, що всі коефіцієнти є незначущими.

t-тест для кожного коефіцієнта заснований на оцінці коефіцієнта регресії і його стандартній помилці і використовує критичне значення з *t-таблиці* для «$n - k - 1$» ступенів свободи (де k – кількість досліджуваних чинників-аргументів). Довірчий інтервал для якого-небудь конкретного коефіцієнта регресії в генеральній сукупності (наприклад, j-го – $в_j$) визначається звичайним способом:

$$\text{від } b_j - tS_{b_j} \text{ до } b_j + tS_{b_j}$$

де t береться з *t-таблиці* для «$n - k - 1$» ступенів свободи.

t-тест є значущим, якщо задане значення «*0*» (вказуюче на відсутність впливу) *не* потрапляє в цей довірчий інтервал. Тут немає нічого нового: це звичайна процедура для двостороннього тестування.

Як альтернативний варіант можна порівняти ***t-статистику*** b_j/S_{b_j} із значенням з t-таблиці і зробити висновок про значущість, якщо абсолютне значення цієї t-статистики виявляється більше. Якщо подивитися на останні

значення в кожному із стовпців t-таблиці, можна побачити достатньо простий, приблизний спосіб визначення значущості коефіцієнтів: значущими будуть ті коефіцієнти регресії, для яких t-статистика по абсолютному значенню рівна або більше 2, оскільки для достатньо великих n *i* рівня значущості 5% значення з t-таблиці приблизно рівного 2. Як завжди, обидва методи, і на використанні t-статистики, і на використанні довірчого інтервалу, повинні у будь-якому випадку забезпечувати однаковий результат (значущість або не значущість) для кожного тесту.

Що ж саме в даному випадку тестується? В результаті ***t-mecma*** для $в_j$; ми повинні ухвалити рішення, чи надає X_j значущий вплив на Y в досліджуваній генеральній сукупності, коли всі інші *X*-змінні залишаються незмінними. В цьому випадку мова не йде про кореляцію між X_j і Y, яка ігнорує всі решта *X*-змінні. Швидше, це перевірка впливу X_j на Y після внесення поправки на решту всіх чинників. Наприклад, в дослідженнях рівня заробітної плати, мета яких полягає у виявленні можливих фактів дискримінації за ознакою статі, зазвичай роблять поправку на рівень освіти і стаж роботи. Не дивлячись на те що чоловіки в компанії можуть (в середньому) отримувати вищу заробітну плату, ніж жінки, дуже важливо зрозуміти, чи не пояснюються ці відмінності якими-небудь іншими чинниками, крім статі. В результаті включення всіх цих чинників в множинну регресію (регресія Y = заробітна плата на X_1 = стать, X_2 = освіта і X_3 = стаж роботи) коефіцієнт регресії для статі відображатиме вплив статі на рівень заробітної плати з урахуванням поправок на рівень освіти і стаж роботи.

Нижче приведені формули для гіпотез, що стосуються перевірки значущості *j-го* коефіцієнта регресії.

Гіпотези для *t-mecma j* -го коефіцієнта регресії

$$H_0: в_j = 0;$$

$$H_1: в_j \neq 0;$$

Якщо повернутися до нашого прикладу з тарифами на розміщення рекламних оголошень в журналах («Приклад. *Реклама в журналах»*), то

відповідний *t-тест* матиме ***n – k – 1*** = 55 – 3 – 1 = 51 ступенів свободи. Двостороннє критичне значення з *t-таблиці* рівне 1,960 (або, точніше, 2,008). У табл. 4.8 приведена відповідна інформація з комп'ютерного роздруку.

Таблиця 4.8. Результат множинною регресійного аналізу тарифів на розміщення реклами в журналах

ВИВЕДЕННЯ ПІДСУМКІВ

Регресійна статистика	
Множинний R	0,887
R-квадрат	0,787
Нормований R-квадрат	0,775
Стандартна помилка	21577,870
Спостереження	55

	Коефіцієнти	***Стандартна помилка***	***t-статистика***	***P-значення***	***Нижні 95%***	***Верхні 95%***
Y-перетин	**4042,799**	16884,039	0,239	0,812	-29853,298	37938,895
Змінна X 1	3,788	0,281	**13,484**	0,000	3,224	4,352
Змінна X 2	**-123,634**	137,849	-0,897	0,374	-400,377	153,108
Змінна X 3	0,903	0,370	2,442	0,018	0,161	1,645

Дві з трьох *X*-змінних є значущими, оскільки для них ***p-значення*** виявляються менше 0,05. Ще один (еквівалентний) спосіб перевірки значущості полягає в тому, щоб з'ясувати, які ***t-статистики (***у комп'ютерному роздруку відповідний стовпець позначений просто ***t***) виявляються більшими, ніж 2,008. І ще один (теж еквівалентний) спосіб перевірки значущості полягає в тому, щоб з'ясувати, які з 95% довірчих інтервалів для коефіцієнтів регресії не включають «***0»***. Як ми і припускали раніше, величина читацької аудиторії робить величезний вплив на рекламні тарифи в журналах. Таке високе значення ***t*** (**13,48**) означає, що вплив величини читацької аудиторії на рекламні тарифи є дуже високо значущим (за умови, що відсоток читачів-чоловіків і середній дохід залишаються

постійними). Вплив середнього доходу на рекламні тарифи в журналах також є значущим (за умови, що відсоток читачів-чоловіків і величина читацької аудиторії залишаються постійними).

Очевидно, що відсоток читачів-чоловіків не робить на тарифи значного впливу (за умови, що величина читацької аудиторії і середній дохід залишаються постійними), оскільки відповідний ***t-тест*** не є значущим. Не виключено, що цей відсоток робить на тарифи певний вплив тільки через дохід (середній дохід у чоловіків може бути вище, ніж у жінок). Таким чином, після внесення поправки на середній дохід можна чекати, що змінна, відповідна відсотку чоловіків, вже не нестиме додаткової інформації для прогнозування тарифів. Не дивлячись на те що оцінюваний вплив відсотка читачів-чоловіків складає **–$123,6**, його відхилення від ***0*** носить лише випадковий характер. Строго кажучи, цей коефіцієнт, –$123,6, не підлягає інтерпретації; оскільки він не є значущим, ви "не маєте права" пояснювати його. Іншими словами, його значення (–$123,6) – лише видимість, і, по суті, нічим не відрізняється від $0,00; більш того, насправді ви не можете навіть сказати, позитивне це число або негативне!

Константа, ***a*** **= $4 043**, не є значущою. Вона не відрізняється істотно від нуля. Не можна сказати нічого визначеного і про знак відповідного параметра генеральної сукупності, ***a,*** оскільки його цілком можна вважати рівним нулю.

Які змінні роблять більший вплив?

Яка з X-змінних робить найбільший вплив на Y? Хороше питання! ***На жаль, вичерпної відповіді на це питання немає, з огляду на те, що наявність взаємозв'язків між X-змінними може зробити принципово неможливим з'ясування того, яка саме з X-змінних насправді "відповідає" за поведінку змінної Y.*** Відповідь на поставлене питання залежить від конкретної ситуації (зокрема, чи можна змінювати *X*-змінні окремо). Відповідь визначається також наявністю взаємозв'язку (або кореляції) між X-змінними. Нижче ми розглянемо корисні (хоча і неповні) відповіді на це непросте питання.

Порівняння приватних коефіцієнтів еластичності.

Яка з *X*-змінних робить найбільший вплив на *Y*? Оскільки всі коефіцієнти регресії $\boldsymbol{b_1, b_2, \ldots, b_k}$ можуть бути виражені в різних одиницях вимірювання, безпосереднє їх порівняння вельми скрутно: невеликий коефіцієнт може насправді виявитися важливішим, ніж великий. Коротше кажучи, тут ми маємо справу з класичною проблемою "спроби порівняння яблук і апельсинів".

Коефіцієнт регресії $\boldsymbol{b_i}$ указує вплив зміни $\boldsymbol{X_i}$ на змінну *Y*, коли всі інші *X*-змінні залишаються незмінними. Коефіцієнт регресії $\boldsymbol{b_i}$ вимірюється в одиницях вимірювання *Y* на одну одиницю вимірювання $\boldsymbol{X_i}$. Якщо, наприклад, *Y* є об'ємом продажів в доларовому виразі, а $\boldsymbol{X_1}$ – кількість торгового персоналу, то $\boldsymbol{b_1}$ виражається в кількості доларів (об'єм продажів) на одну людину. Допустимо, що наступний коефіцієнт регресії, $\boldsymbol{b_2}$, виражається в кількості доларів (об'єм продажів) на сумарний кілометраж робочих поїздок торгових представників компанії. Безпосереднє порівняння $\boldsymbol{b_1}$ і $\boldsymbol{b_2}$ не дозволить нам відповісти на питання, який з цих двох чинників (рівень торгового персоналу або витрати на відрядження компанії) робить більший вплив на об'єм продажів, тому що різні одиниці вимірювання (долари на людину і долари на кілометр) безпосередньо порівнювати не можна.

!!! Коефіцієнт еластичності (E), правильніше приватний коефіцієнт еластичності для кожного чинника-аргументу (пояснюючою змінною) – $\boldsymbol{E_i}$, який обчислюється для лінійних регресійних моделей як

$$\boldsymbol{E_i = b_i \times \overline{X_i} / \overline{Y}} \qquad (4.3)$$

показує на скільки відсотків зміниться *Y* при зміні X_i на один відсоток. На наш погляд, саме цей показник, уникаючи різноіменності коефіцієнтів регресії, дозволяє найточніше визначити ступінь впливу різних чинників-аргументів на результативну ознаку, тобто на *Y*.

У нашому прикладі з тарифами на розміщення реклами

$$E_1 = b_1 \times \overline{X_1} / \overline{Y} = 3{,}788 \times 10913 / 83534 = 0{,}495;$$

$$E_2 = b_2 \times \overline{X_2} / \overline{Y} = (-123{,}634) \times 39{,}7 / 83534 = -0{,}059;$$

$$E_3 = b_3 \times \overline{X_3} / \overline{Y} = 0{,}903 \times 47710 / 83534 = 0{,}516.$$

Оскільки найбільше абсолютне значення приватного коефіцієнта еластичності спостерігається у третього чинника (X_3), що характеризує медіану доходу потенційних читачів, то можна з певною вірогідністю стверджувати, що саме він робить найбільший вплив на ціну однієї рекламної сторінки в досліджуваній групі журналів. Не слідує, проте, нехтувати і впливом першого чинника (X_1) – розміру читацької аудиторії. Абсолютне значення коефіцієнта «аудиторної еластичності» (0,495) не набагато поступається коефіцієнту «прибуткової еластичності» (0,516).

4.6. Складнощі і проблеми, пов'язані з множинною регресією

На жаль, на практиці множинна регресія не завжди дозволяє отримати результати, про які пишуть в підручниках. У цьому пункті приведений перелік потенційних проблем і деякі міркування з приводу того, як з ними справитися (у тих випадках, коли це можливо).

Існують три основні різновиди проблем. Нижче приведений короткий огляд кожного з цих різновидів, а потім слідує докладніший їх опис.

1. Проблема **мультиколінеарності** виникає в тих випадках, коли деякі з ваших пояснюючих змінних (X) виявляються дуже схожими. Не дивлячись на те, що ці змінні можуть добре пояснювати і прогнозувати Y (на що указують високе значення R^2 і значущий ***F-тест)***, окремі коефіцієнти регресії погано піддаються оцінці. Це пов'язано з тим, що ми не маємо в своєму розпорядженні достатньої інформації, щоб вирішити, яка (або які) із змінних забезпечує це пояснення. Одне з можливих рішень полягає в тому, щоб видалити з рівняння деякі зі змінних з метою позбавитися від сумнівів. Інше рішення полягає в тому,

щоб перевизначити якісь змінні (можливо, шляхом ділення), щоб відрізняти одну змінну від іншої.

2. Проблема **вибору змінних** виникає в тих випадках, коли доводиться мати справу з просторовим переліком потенційно корисних пояснюючих (незалежних) *X*-змінних і необхідно вирішити, які з цих змінних слід включати в рівняння регресії. З одного боку, якщо у вас дуже багато *X*-змінних, зайві з них знижуватимуть якість результатів (можливо, унаслідок все тієї ж мультиколінеарності). Частина інформації, що міститься в даних, даремно витрачається на оцінювання непотрібних параметрів. З іншого боку, якщо відкинути потрібну *X*-змінну, знизиться якість прогнозів, оскільки ви проігноруєте корисну інформацію. Одне з можливих рішень полягає в тому, щоб гарненько подумати, *чому* важлива та або інша *X*-змінна, щоб бути упевненим в тому, що кожна змінна, яка включається в розгляд, дійсно виконує важливу функцію. Інший підхід полягає в тому, щоб скористатися автоматичною процедурою, яка прагне відібрати найбільш важливі змінні.

3. Проблема **неправильного вибору** моделі пов'язана з безліччю різних потенційних невідповідностей між вашим конкретним завданням і моделлю множинної лінійної регресії, яка є фундаментом і каркасом множинного лінійного регресійного аналізу. Може вийти так, що ваше конкретне завдання не відповідає умовам і допущенням моделі лінійної множинної регресії. Аналізуючи дані, ви можете виявити деякі потенційні проблеми, пов'язані з нелінійністю, нерівною мінливістю і наявністю значень, що різко відхиляються. Проте навіть наявність подібних проблем ще ні про що не говорить. Не дивлячись на те, що гістограми деяких змінних можуть бути сильно скошеними (несиметричними), а деякі діаграми розсіювання можуть бути нелінійними, модель множинної лінійної регресії і в таких випадках цілком може бути застосовна. Існує так звана *діагностична діаграма*, яка допомагає зрозуміти, чи дійсно виявлена проблема є настільки серйозною, що її необхідно якось вирішувати. Один з можливих варіантів рішень полягає в створенні нових *X*-змінних, які формуються на основі існуючих змінних, і/або перетворенні деяких

або всіх цих змінних. Ще одна серйозна проблема виникає у разі, коли доводиться мати справу з *часовим рядом*, стосовно якого допущення моделі лінійної множинної регресії про незалежність окремих спостережень не дотримується. Проблема часових рядів не має простого рішення, проте множинну регресію можна виконати, використовуючи замість початкових даних *процентні зміни* між різними часовими періодами.

Мультіколінеарність: чи не дуже схожі між собою пояснюючі змінні?

Коли якісь з пояснюючих *X*-змінних дуже схожі між собою, у вас може виникнути проблема *мультиколінеарності*, оскільки множинна регресія не в змозі відрізнити вплив однієї змінної від впливу іншої змінної. Наслідки мультиколінеарності можуть бути *статистичними* або *обчислювальними*.

1. Статистичні наслідки мультиколінеарності пов'язані з труднощами проведення статистичних тестів для окремих коефіцієнтів регресії унаслідок збільшення стандартних помилок. Результатом може бути неможливість оголосити ту або іншу *X*-змінну значущою навіть в тому випадку, якщо ця змінна (сама по собі) має сильний взаємозв'язок с Y.

2. Обчислювальні наслідки мультиколінеарності пов'язані з труднощами в організації обчислень на комп'ютері, викликаними "нестійкістю обчислень". У крайніх випадках комп'ютер може намагатися виконати ділення на нуль і, таким чином, невдало завершити аналіз даних. Гірше за те, комп'ютер може завершити аналіз і видати безглузді і невірні результати. Ділення на нуль неможливе з математичної точки зору: наприклад, результат виконання 5/0 є невизначеним. Проте із-за невеликих помилок округлення в процесі обчислень комп'ютер може розділити не 5 на 0, а 5,0000000000968 на 0,0000000000327. В цьому випадку, замість того щоб зупинитися і повідомити про помилку, комп'ютер використовує в подальших обчисленнях безглуздий і величезний результат такого ділення: 152 905 198 779,72.

Мультіколінеарність може породжувати проблеми, а може і не породжувати їх, – все залежить від конкретних цілей виконуваного вами аналізу і ступеня мультиколінеарності. Невелика або середня мультиколінеарність

зазвичай не є проблемою. Дуже сильна мультиколінеарність (наприклад, включення однієї і тієї ж змінної двічі) завжди буде проблемою і може приводити до серйозних помилок (обчислювальні наслідки). На щастя, якщо вашою метою є в основному прогноз або прогнозування *Y*, сильна мультиколінеарність може не являти серйозної перешкоди, оскільки якісна програма множинної регресії може і в цьому випадку робити оптимальні прогнози (по методу найменших квадратів), засновані на всіх *X*-змінних. Проте якщо ви хочете використовувати індивідуальні коефіцієнти регресії для з'ясування того, як кожна з *X*-змінних впливає на Y, то статистичні наслідки мультиколінеарності, мабуть, викличуть певні проблеми, зважаючи на те, що ці впливи неможливо відокремити один від одного. У табл. 4.8 підсумовується вплив мультиколінеарності на результати регресійного аналізу.

Таблиця 4.8. Вплив мультиколінеарності на регресію

Ступінь мультиколінеарності	**Вплив на регресійний аналіз**
Незначна	Взагалі не є проблемою
Середня	Як правило, не є проблемою
Сильна	Статистичні наслідки: часто є проблемою, якщо потрібно оцінити вплив окремих *X*-змінних (тобто коефіцієнти регресії); може не бути проблемою, якщо мета полягає в прогнозі або прогнозуванні *У*
Надзвичайно сильна	Чисельні наслідки: завжди є проблемою; комп'ютерні обчислення можуть навіть виявитися неправильними через нестійкісті обчислень

Як з'ясувати, чи дійсно існує проблема мультиколінеарності? Один з простих способів відповісти на це питання полягає в аналізі звичайних двовимірних кореляцій для кожної пари змінних. *Кореляційна матриця* є таблицею, яка містить коефіцієнти кореляції для кожної пари змінних з вашої багатовимірної сукупності даних. Чим вище коефіцієнт кореляції між двома *X*-змінними, тим більше мультиколінеарність. Це пояснюється тим, що висока кореляція (близька до 1 або -1) указує на сильний зв'язок і свідчить про те, що ці

дві *X*-змінні вимірюють дуже схожі характеристики, привносячи тим самим в аналіз інформацію, яка перетинається з тою, що є.

Основний статистичний результат мультиколінеарності полягає в *зростанні стандартних помилок деяких або всіх коефіцієнтів регресії* (S_{b_i}). Це цілком природно: якщо дві *X*-змінні містять інформацію, яка перетинається, важко визначити вплив кожній з них окремо. Високе значення стандартної помилки приводить до того, що комп'ютер повідомляє вам приблизно наступне: "Я обчислив для вас коефіцієнт регресії, але результат неточний, оскільки важко сказати, ця або якась інша змінна є визначальною". В результаті довірчі інтервали для відповідних коефіцієнтів регресії значно розширюються, а *t-тести* навряд чи будуть значущими.

У разі сильної мультиколінеарності може виявитися, що регресія дуже високо значуща (виходячи з результатів F-теста), проте жоден з *t-тестів* для окремих *X*-змінних значущим не є. Комп'ютер повідомляє вас про те, що *X*-змінні, що розглядаються як єдина група, вельми сильно впливають на *Y*, але практично неможливо визначити важливість якоїсь конкретної змінної. Слід пам'ятати, що *t-тест* для конкретної *X*-змінної вимірює її вплив на Y за умови, що значення інших змінних залишаються незмінними. Таким чином, *t-тест* для змінною X_i виявляє тільки *додаткову* інформацію, привнесену змінною X_i *і* крім тієї інформації, яку несуть інші *X*-змінні. Якщо якась інша змінна дуже близька до X_i, тоді змінна X_i не привносить в регресію значущо нову інформацію.

Одне з рішень полягає в тому, щоб проігнорувати ті *X*-змінні, які дублюють інформацію, вже присутню в інших *X*-змінних. Якщо, наприклад, ваші *X*-змінні містять три різні вимірювання розміру, спробуйте або позбавитися від двох з них, або об'єднати всі три змінні в єдину міру розміру (наприклад, скориставшись їх середнім значенням).

Інше рішення полягає в тому, щоб перевизначити деякі змінні з тим, щоб кожна з *X*-змінних виконувала чітку, властиву тільки їй одній роль у визначенні Y. Розповсюджений спосіб застосування цієї ідеї до групи близьких один до одного *X*-змінних полягає в тому, щоб узяти для представлення цієї групи одну

X-змінну (можна або вибрати одну з цих *X*-змінних, або сформувати з них індекс) і представити решту змінних у вигляді відносних показників (наприклад, величина на одиницю іншого показника), побудованих з цієї представляючою *X*-змінною. Наприклад, можна представляти залежність розміру об'єму продажів (Y) за допомогою чисельності населення (X_1) і загального доходу (X_2) для кожного регіону. Проте ці змінні є мультиколінеарними (тобто чисельність населення і загальний дохід – високо корельовані величини). Цю проблему можна вирішити, пояснюючи об'єм продажів (Y) за допомогою чисельності населення (X_1) і розміру доходу на одну людину (нова змінна X_2). В результаті чисельність населення виконуватиме роль представляючої змінної, відображаючи загальну величину території, а дохід, замість того щоб повторювати вже відому нам інформацію (про величину відповідної території), перевизначається і несе нову інформацію (про добробут людей).

Вибір змінної: можливо, ми користуємося "не тими" змінними?

Результати статистичного аналізу значною мірою залежать від наявної інформації, тобто від використаних для аналізу даних. Зокрема, особливу увагу слід звернути на вибір незалежних ("що пояснюють") *X*-змінних для множинного регресійного аналізу. Включення як можна більшого числа *X*-змінних "просто так, про всяк випадок" або тому, що "створюється враження, ніби кожна з них якось впливає на Y'' – далеко не краще рішення. Роблячи таким чином, ви прирікаєте себе на можливі труднощі при визначенні значущості для регресії (*F-тест),* або – унаслідок мультиколінеарності, викликаної наявністю надмірних змінних, – у вас можуть виникнути труднощі при рішенні питання про значущість для деяких окремих коефіцієнтів регресії.

Що відбувається, коли ви включаєте одну зайву, недоречну *X-змінну?* Значення $\boldsymbol{R^2}$ в цьому випадку опиниться дещо більшим, оскільки деяку велику частку Y можна пояснити за рахунок випадковості цієї нової змінної. Проте *F-тест* значущості регресії враховує це збільшення, тому таке збільшення $\boldsymbol{R^2}$ не можна вважати перевагою.

Насправді включення додаткової *X*-змінної може принести невелику або навіть помірну *шкоду*. Оцінка того або іншого недоречного параметра (в даному випадку недоречного коефіцієнта регресії) залишає менше інформації для стандартної помилки оцінки, $\boldsymbol{S_e}$. По технічних причинах наслідком цього є менш могутній *F-тест*, який може не виявити значущість навіть у тому випадку, коли *X*- змінні в генеральній сукупності насправді пояснюють Y.

А що відбудеться у разі, коли ви проігноруєте необхідну *X*-змінну? В результаті з сукупності даних випаде важлива і корисна інформація і ваше прогнозування $\boldsymbol{Y}$ буде менш точним, чим у разі використання цієї *X*-змінної. Стандартна помилка оцінки, $\boldsymbol{S_e}$, в цьому випадку, як правило, виявляється більше (що указує на великі помилки прогнозування), а R^2, як правило, виявляється меншим (що указує на пояснення меншої частки варіації Y). Природно, якщо ви проігноруєте критично важливу ***X***-змінну, то, можливо, *F-тест* для цієї регресії просто буде незначущий.

Ваше завдання в даному випадку – включити рівно стільки *X*-змінних, скільки потрібно (тобто не дуже багато і не дуже мало), причому включити саме ті *X*-змінні, які необхідні. Якщо у вас є сумніви, можна включити деякі з *X*-змінних, щодо яких ви не упевнені. У такому разі корисний суб'єктивний метод (заснований на пріоритетному переліку ***X***-змінних). Існує також безліч різних автоматичних методів.

Класифікація переліку *X*-змінних за пріоритетами

Хороший спосіб визначити круг важливих *X*-змінних полягає в тому, щоб уважно проаналізувати вирішувану задачу, наявні дані і цілі, яких ви хочете добитися. Потім необхідно скласти список *X*-змінних, класифікованих за пріоритетами. Зробити це можна таким чином.

1.Виберіть змінну Y, яку вам необхідне пояснити, зрозуміти або прогнозувати.

2.Виберіть *X*-змінну, яка, як вам здається, є найбільш важливою у визначенні або поясненні Y. Якщо це викликає у вас складнощі, оскільки всі *X*-змінні здаються вам однаково важливими, ухваліть *вольове* рішення.

3.Виберіть найважливішу серед *X*-змінних, що залишилися, поставивши собі питання: "Зважаючи на першу змінну, яка з X-змінних, що залишилися, несе більше нової *інформації*, що пояснює поведінку змінної Y?".

4.Продовжуйте вибирати за цим принципом найважливіші з *X*-змінних, що залишилися, до тих пір, поки не класифікуєте за пріоритетами весь перелік *X*-змінних. На кожній стадії ставте собі питання: "Зважаючи на вже відібрані X-змінні, яка з X-змінних, що залишилися, несе більше нової *інформації*, що пояснює поведінку змінної Y?"

Потім обчисліть регресію, використовуючи лише ті *X*-змінні зі складеного вами списку, які здаються вам найважливішими. Обчисліть ще декілька регресій, включаючи в свій аналіз деякі з *X*-змінних (або всі ці змінні), *що* залишилися, і з'ясуйте, чи дійсно вони впливають на прогнозування змінної *Y*. Нарешті, виберіть той результат регресії, який здається вам найбільш корисним.

Не дивлячись на те, що описана процедура виглядає достатньо суб'єктивною (оскільки залежить в основному від вашої суб'єктивної думки), їй властиві дві важливі переваги. По-перше, коли необхідно зробити вибір між двома *X*-змінними, які практично однаково пояснюють поведінку змінної Y, остаточний вибір залишається за вами (автоматизована процедура може в цьому випадку зробити менш змістовний вибір). По-друге, ретельно класифікувавши по пріоритетах свої незалежні *X*-змінні, ви можете глибше розібратися в досліджуваній ситуації. Таке прояснення вирішуваного завдання може виявитися не менш корисним, чим результати множинної регресії!

Проблема неправильного вибору моделі.

Перш за все, слід пам'ятати, що маса серйозних проблем виникає у разі, коли доводиться мати справу з часовими, *а не з одночасно зрізаними*, *наборами даних*, стосовно яких допущення стандартної моделі лінійної множинної регресії

про незалежність окремих спостережень не дотримується. Проблема часових рядів не має простого рішення, проте множинну регресію можна виконати, використовуючи замість початкових даних *процентні зміни* між різними часовими періодами.

4.7. Визначення мультиколінеарності. Алгоритм Феррара-Глобера

Алгоритм Феррара–Глобера (Farrar-Glauber test) дає можливість детально дослідити наявність мультиколінеарності. Він базується на регресійному аналізі та оцінює коефіцієнти регресії для кожної залежної змінної залежно від інших змінних. Згодом, аналізується кореляція між оцінками коефіцієнтів регресії та визначається наявність мультиколінеарності. Алгоритм містить три етапи згідно з видами статистичних критеріїв, за якими перевіряється мультиколінеарність:

- всього масиву пояснювальних змінних (χ^2 – «хі»-квадрат),
- кожної пояснювальної змінної з рештою змінних (F-критерій),
- кожної пари пояснювальних змінних (t-критерій).

Про мультиколінеарність дасть змогу свідчити зіставлення отриманих фактичних значень критеріїв з критичними значеннями цих критеріїв.

Перший етап полягає у перевірці на загальну мультиколінеарність. Порядок дій наступний:

1. Обчислити кореляційну матрицю парних коефіцієнтів кореляції. Для цього варто використати надбудову Анализ данных→Корреляция (Data Analysis →Correlation) електронних таблиць MS Excel. Іншим способом визначення парних коефіцієнтів кореляції між пояснювальними змінними є застосування функції КОРРЕЛ або CORREL (MS Excel) та формування з них кореляційної матриці.

2. Розрахувати детермінант кореляції det r_{xx}, використовуючи функції МОПРЕД або MDETERM (MS Excel)

3. Визначити критерій χ^2:

$$\chi^2 = -[n-1-(2m+5)/6]\ln(\det r_{xx}). \quad (4.4)$$

Отримане значення порівнюється з критичним значенням $\chi^2(\alpha;\gamma)$ при $\gamma = m\cdot(m-1)/2$ ступенях свободи та рівні значущості α. Якщо $\chi^2 > \chi^2(\alpha;\gamma)$, то в масиві пояснювальних змінних існує мультиколінеарність. У протилежному випадку ($\chi^2 \le \chi^2(\alpha;\gamma)$) із довірчою ймовірністю $p = 1-\alpha$ вона відсутня, тому необхідності у подальших розрахунках за алгоритмом Феррара–Глобера немає. Для визначення критичного значення $\chi^2(\alpha;\gamma)$ зручно застосувати функцію *ХИ2ОБР* або *CHISQ.INV.RT* (*MS Excel*)

Другий етап дослідження за алгоритмом Феррара–Глобера полягає у перевірці мультиколінеарності кожної пояснювальної змінної з рештою. Його можна виконати двома способами.

За першим способом необхідно скористатися інформацією, яку надає обернена матриця $C = r_{xx}^{-1}$. Матриця C розраховується з використанням функції *МОБР* чи *MINVERSE* (*MS Excel*). Далі розраховуються F-критерії за співвідношенням

$$F_j = (c_{jj} - 1)\frac{n-m}{m-1} \quad (4.5)$$

де c_{jj} – діагональні елементи матриці C, $j = 1..m$.

Фактичні значення розрахованих за формулою 4.5 критеріїв порівнюють з критичним значенням F-критерію $F(\alpha;\nu_1;\nu_2)$ при $\nu_1 = m-1$ та $\nu_2 = n-m$ ступенях свободи та рівні значущості α . Якщо $F_j > F(\alpha;\nu_1;\nu_2)$, то відповідна j-та незалежна змінна мультиколінеарна з іншими. Критичне значення Fкритерію знаходять за допомогою функції *FРАСПОБР або F.INV.RT* (*MS Excel*) .

Якщо коефіцієнт детермінації для кожної змінної $R_j^2 = 1 - \frac{1}{c_{jj}}$ має значення, яке близьке до 1, то можна стверджувати про наявність мультиколінеарності.

За другим способом для кожної змінної побудуємо регресійну залежність від інших факторів. Спочатку, наприклад, будуємо лінійну регресію виду (наприклад для 3-х факторів)

$^{y}{}_1 = c_0 + c_1x_1 + c_2x_2$

Для кожної моделі знаходимо залишки ε_i і за ними – величину R^2 для кожної моделі. Розраховуємо F-статистику Фішера за формулою

$$F = \frac{\frac{R^2}{k}}{\frac{1-\ ^2}{n\ \ k-1}} \tag{4.6}$$

і порівнюємо їх із $F_{табл.}$

Третій етап дослідження за алгоритмом Феррара–Глобера полягає у перевірці на мультиколінеарність кожної пари пояснювальних змінних та потребує:

• по-перше, знаходжуються частинні коефіцієнти кореляції

$$r_k = \frac{-c_{kj}}{\sqrt{c_{kk} * c_{jj}}} \tag{4.7}$$

де c_k – елемент матриці C, що міститься в k-му рядку та j-му стовпці; c_{kk} і c_{jj} – діагональні елементи матриці C; $j \neq k$, j =1..m, k =1..m;

• по-друге, обчислюються t-критерії

$$t_k = \frac{r_{kj}\sqrt{v}}{\sqrt{1-r_{kj}^2}} \tag{4.8}$$

де $\nu = n - m$, $j \neq k$, j =1..m, k =1..m.

Фактичні значення критеріїв порівнюються з критичним значенням $t(\alpha;\nu)$ при ν ступенях свободи та рівні значущості α. Якщо $|t_k| > t(\alpha;\nu)$, то між незалежними змінними x_k і x_j існує мультиколінеарність. Критичне значення

двостороннього t-критерію можна знайти із застосуванням функції *СТЬЮДРАСПОБР* або *T.INV.2T* (*MS Excel*).

Також можна розрахувати часткові коефіцієнти кореляції для випадку з трьох факторів за наступною формулою:

$$r_{xy.z} = \frac{r_{xy} - r_x \; r_y}{\sqrt{(1-r_x^2)(1-r_y^2)}} \tag{4.9}$$

Тоді критерій Стьюдента для трьох факторів буде:

$$t_{xy} = \frac{r_{xy.z}\sqrt{n-m}}{\sqrt{1-r_{x\;.z}^2}} \tag{4.10}$$

Якщо табличне більше розрахованого – мультиколінеарності немає.

Відкинувши одну зі змінних мультиколінеарної пари, можна легше за все позбутися мультиколінеарності в економетричній моделі. Якщо $F_j > F(\alpha;\nu_1;\nu_2)$, тобто коли x_j залежить від усіх інших пояснювальних факторів, то необхідно вирішувати питання про її вилучення з переліку змінних. Якщо $t_k > t(\alpha;\nu)$, то змінні x_k і x_j тісно пов'язані між собою. Виходячи з цього, тобто аналізуючи рівень критеріїв Фішера та Стьюдента, можна зробити обґрунтований висновок про те, яку зі змінних необхідно вилучити з моделі або замінити іншою. Але на практиці вилучення певного чинника може суперечити логіці економічних зв'язків, і заміна масиву незалежних змінних завжди має узгоджуватись з економічною доцільністю, що випливає з мети дослідження.

Розглянемо алгоритм Феррара-Глобера на прикладі з рекламою в журналах. На основі данних таблиці 4.3 будемо робити обчислення.

1. Обчислення кореляційної матриці парних коефіцієнтів кореляції.

За допомогою Excel знайдено кореляційну матрицю (див. табл. 4.9).

Таблиця 4.9. Кореляційна матриця

	Column 1	*Column 2*	*Column 3*
Column 1	1	-0.13428	-0.35316
Column 2	-0.13428	1	0.563807
Column 3	-0.35316	0.563807	1

2. Розрахувати детермінант кореляції det r_{xx} із використанням функції МОПРЕД або MDETERM (MS Excel)

В нашому випадку det r_{xx}= 0.592842.

3. Визначити критерій χ^2:

$\chi^2 = -[n-1-(2m+5)/6] \ln(\det r_{xx}) = 27.27417$

При n=55, m=3

$\chi^2(\alpha;\gamma)$ при $\gamma = m\cdot(m-1)/2=3(3-1)/2=3$ ступенях свободи та рівні значущості α.

$\chi^2(0.05;3)=7.81$

Оскільки розрахований χ^2 більше табличного, то мультиколінеарність існує.

На рис. 4.1 представлено розрахунки в MS Excel.

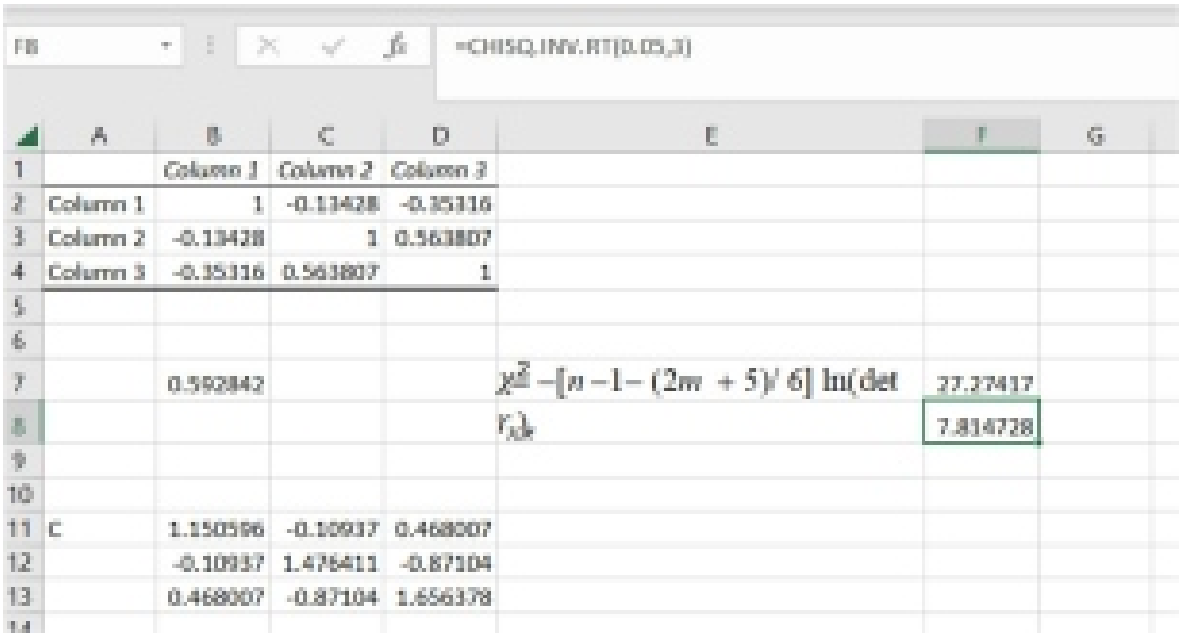


Рис. 4.1. Розрахунок в MS Excel критерію хі-квадрат

Переходимо до другого етапу дослідження за алгоритмом Феррара–Глобера – це перевірка мультиколінеарності кожної пояснювальної змінної з рештою. Для цього беремо інформацію, яку надає обернена матриця $C = r_{xx}^{-1}$. Матриця *C* розраховується з використанням функції *МОБР* чи *MINVERSE* (*MS Excel*) (див. рис. 4.1).

Далі обчислимо *F*-критерії за формулою 4.5.

$$F_1 = (c_{11} - 1)\frac{n-m}{m-1} = (1.15 - 1)\frac{55-3}{3-1} = 3.915.$$
$$F_2 = (c_{22} - 1)\frac{n-m}{m-1} = (1.476 - 1)\frac{55-3}{3-1} = 12.39.$$
$$F_3 = (c_{33} - 1)\frac{n-m}{m-1} = (1.656 - 1)\frac{55-3}{3-1} = 17.07.$$

Фактичні значення критеріїв порівнюються з критичним значенням *F*-критерію $F(\alpha;\nu_1;\nu_2)$ при $\nu_1 = m - 1$=3-1=2 та $\nu_2 = n - m$=55-3=52 ступенях свободи та рівні значущості α . Якщо $F_j > F(\alpha;\nu_1;\nu_2)$, то відповідна *j*-та незалежна змінна мультиколінеарна з іншими.

F(0.05;2; *52*) =3.175140971

Оскільки всі отриманні значення більше табличного, то всі незалежні змінні мультиколінеарні з іншими.

Якщо коефіцієнт детермінації для кожної змінної $R_j^2 = 1 - \frac{1}{c_{jj}}$ має значення, яке близьке до 1, то можна стверджувати про наявність мультиколінеарності.

$$R_1^2 = 0.13, R_2^2 = 0.32, R_3^2 = 0.396.$$

На третьому етапі дослідження за алгоритмом Феррара–Глобера перевіримо мультиколінеарність кожної пари пояснювальних змінних. За формулою 4.7:

$$r_{12} = \frac{-c_1}{\sqrt{c_{11} * c_{22}}} = \frac{-(-0.10)}{\sqrt{1.15 * 0.46}} = 0.08.$$

Аналогічно для $r_{13} = -0.339$ та $r_{23} = 0.557$

• по-друге, обчислення *t*-критеріїв (див. форм. 4.8):

$$t_{12} = 0.61, t_{13} = -2.6, t_{23} = 4.83$$

Фактичні значення критеріїв порівнюються з критичним значенням $t(\alpha;\nu)$ при ν ступенях свободи та рівні значущості α .

$t(0.05;55-3)=2.007$

Оскільки $|t_{13}| > t(\alpha;\nu)$, то між незалежними змінними x_1 і x_3 існує мультиколінеарність. Аналогічно $|t_{23}| > t(\alpha;\nu)$, тому між незалежними змінними x_2 і x_3 існує мультиколінеарність.

Розрахунок в Excel представлено нижче (рис.4.2).

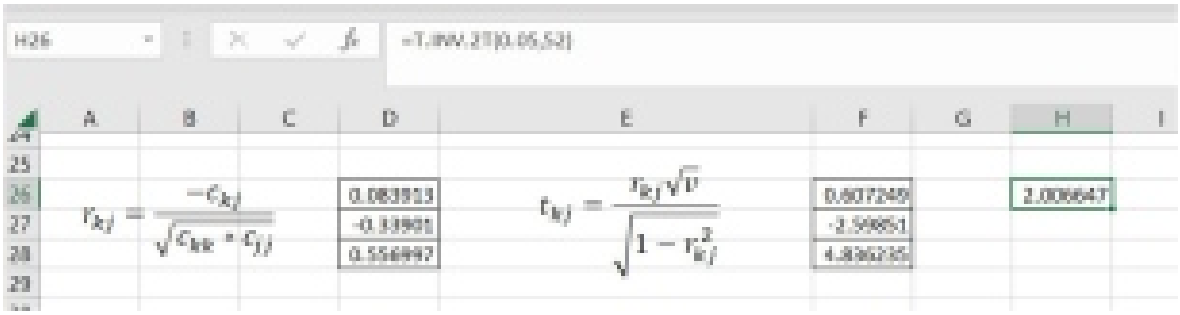


Рис.4.2. Виявлення мультиколінеарності між змінними

Таким чином, виявлена мультиколінеарність в даній моделі. Можна спробувати вилучити третій фактор, який мультиколінеарний з двома іншими.

4.8. Оцінка впливу окремих факторів на досліджувану змінну

Важливу роль при оцінці впливу окремих факторів грають коефіцієнти регресійної моделі a_j. Однак безпосередньо з їх допомогою не можна зіставити

фактори за ступенем їх впливу на залежну змінну через відмінності одиниць виміру і різного масштабу коливань.

Коефіцієнти еластичності показує, на скільки відсотків змінюється досліджувана змінна при зміні факторної змінної на 1 відсоток.

$$E_j = a_j \times \frac{x_{j\ p}}{y_{cp}} \tag{4.11}$$

Бета–коефіцієнт показує, на яку частину величини середньоквадратичного відхилення зміниться змінна у зі зміною відповідної незалежної змінної xj на величину свого середньоквадратичного відхилення при фіксованому рівні значень інших факторних змінних.

$$\beta_j = a_j \times \frac{S_{x_j}}{S_y} \tag{4.12}$$

$$S_{x_j} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - x_{j\ p}\right)^2} \tag{4.13}$$

$$S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - y_{cp}\right)^2} \tag{4.14}$$

Дельта – коефіцієнт показує частку впливу фактору в сумарному впливі всіх факторів

$$\Delta_j = r(x_j, y) \times \frac{\beta_j}{R^2} \tag{4.15}$$

4.9. Побудова прогнозів на основі моделі множинної регресії

Точковий прогноз – це розрахункове значення залежної змінної, отримане підстановкою в рівняння множинної лінійної регресії прогнозних (заданих

дослідником) значень незалежних змінних. Якщо задані значення $x_1^{\text{пр}}$, $x_2^{\text{пр}}, \dots, x_1^{\text{пр}}$, то прогнозне значення залежної змінної (точковий прогноз) дорівнюватиме

$$y_{\text{прогн}}^{\text{точ}} = a_0 + a_1 x_{\text{прогн1}} + a_2 x_{\text{прогн2}} + \cdots + a_m x_{\text{прогн}m}. \quad (4.16)$$

$$Y_{\text{прог}}^{\text{точ}} = X_{\text{прогн}} A \quad (4.17)$$

Інтервальний прогноз – це мінімальне і максимальне значення залежної змінної, в проміжок між якими вона потрапляє із заданою ймовірністю і при заданих значеннях незалежних змінних.

Інтервальний прогноз для лінійної функції обчислюється за формулою

$$V = X_{\text{прогн}} \times (X_{\text{в}}^T \times X_{\text{в}})^{-1} \times \text{X}_{\text{прогн}}^{\text{T}}, \quad (4.18)$$

$$L_i = S_{\text{ст}} \times t_{\alpha,n-m-1} \times \sqrt{1 + v_{ii}}, \quad (4.19)$$

$$S_{\text{ст}} = \sqrt{\frac{\sum_{i\ 1}^{n} \varepsilon_i^2}{n-m-1}}, \quad (4.20)$$

$$v_{ii} = \frac{m \prod_{j=1}^{m} \left(x_{i\ \text{прогн}} - x_{j\text{ср}}\right)^2}{\prod_{j=1}^{m} \sum_{i\ 1}^{n} \left(x_{i\ } - x_{j\ \text{р}}\right)^2}. \quad (4.21)$$

$$y_{i\ \text{прогн}} \in \left[y_{i\ \text{прогн}}^{\text{точ}} - L_i;\ y_{i\ \text{прогн}}^{\text{точ}} + L_i \right] \quad (4.22)$$

Або через матриці

$$\Delta Y_{\text{прогн}} = S_{\text{ст}} * t_{\alpha,n-m-1} \sqrt{X_{\text{пр}}^{\text{T}} B^{-1} X_{\text{пр}}}, \quad X_{\text{пр}}^T = \left(1, X_{1\text{пр}}, \text{X}_{2\text{пр}}, \dots, \text{X}_{m\text{пр}}\right) \quad (4.23)$$

де

$$B = X^T X. \quad (4.24)$$

$$Y_{\text{прогн}} - \Delta Y_{\text{прогн}} \le Y_{\text{прогн}}^{\text{точ}} \le Y_{\text{прогн}} + \Delta Y_{\text{прогн}} \quad (4.25)$$

4.10. Нелінійні моделі регресії

Нелінійні моделі регресії - це клас статистичних моделей, де залежність між змінними не може бути адекватно описана лінійною функцією. Вони використовуються тоді, коли взаємозв'язок між змінними є більш складним та нелінійним. Ось деякі ключові аспекти нелінійних моделей регресії:

1. Функціональна форма.

У нелінійних моделях регресії залежність між пояснювальними та поясновуваними змінними може бути виражена за допомогою нелінійних функцій, таких як показникові, логарифмічні, поліноміальні та інші.

2. Параметризація.

Нелінійні моделі мають більше параметрів, що може призвести до складнішого аналізу та оцінки. Відбір підходящих параметрів може бути важливим завданням.

3. Добір функцій.

Важливим аспектом нелінійних моделей є вибір відповідних нелінійних функцій для опису залежності між змінними. Це може вимагати експериментів та аналізу.

4. Перевірка адекватності.

Важливо перевірити адекватність обраної нелінійної моделі до даних, використовуючи статистичні методи та показники.

5. Інтерпретація.

Інтерпретація параметрів у нелінійних моделях може бути більш складною, оскільки зміни в пояснювальних змінних можуть впливати на поясновану змінну за складною закономірністю.

6. Прогнозування.

Нелінійні моделі можуть бути потужними для прогнозування, оскільки вони можуть зближати модель з реальними залежностями.

При використанні нелінійних моделей регресії важливо враховувати їхні особливості та вимоги до обробки даних, оцінки параметрів та інтерпретації результатів.

У результаті порівняння декількох регресійних залежностей, коефіцієнт детермінації R^2 допомагає визначити ту модель, яка краще відображає зв'язок між ознаками та має більшу пояснювальну силу. Високе значення R^2 свідчить про кращу відповідність моделі досліджуваним даним, що робить її більш ефективною для прогнозування та аналізу економічних явищ.

Розрізняють такі класи нелінійних регресійних моделей:

1. нелінійні щодо пояснювальних чинників**,** але лінійні за параметрами моделі (до них відносять, наприклад,

   поліноми різних ступенів – $y = a_0 + a_1x^1 + a_2x^2 + \cdots + a_n x^n + \varepsilon$**;**

   гіперболічні (зворотні) функції – $y = a + \frac{b}{x} + \varepsilon, y = a_0 + \frac{b_1}{x_1} + \frac{b_2}{x_2} + \cdots + \frac{b_n}{x_n} + \varepsilon.$

2. нелінійні за оцінюваними параметрами модел***:***

   степеневі функції – $y = a_0 \cdot x^{a_1} \cdot \varepsilon, y = a_0 x_1^{a_1} * x_2^{a_2} \dots x_n^{a_n} * \varepsilon,$ **;**

   показникові –$y = a_0 a_1^x \varepsilon,\ y = a_0 a_1^{x_1} a_2^{x_2} \dots a_n^{x_n} \varepsilon;$

   експоненційні – $y = a_0 e^{a\ b}\ \varepsilon$,

   функції модифікованої експоненти – $y = a_0 e^{a\ b}\ + k$, тощо.

Нелінійні щодо пояснювальних чинників моделі можуть бути приведені до лінійного вигляду шляхом переходу до нових змінних. Наприклад, для полінома n-го степеня $y = a_0 + a_1x^1 + a_2x^2 + \cdots + a_n x^n + \varepsilon$, перехід до нових змінних $x_1' = x^1$, $x_2' = x^2$,…, $x_n' = x^n$, дозволяє подати залежність у вигляді лінійної моделі від нових змінних – $Y = a_0 + a_1x'_1 + a_2x_2' + \cdots + a_nx_n' + \varepsilon$, параметри якої $(a_0, a_1, a_2, \dots a_n)$, відповідають параметрам початкової нелінійної моделі.

Більш складна ситуація з нелінійними за параметрами моделями. Серед них розрізняють внутрішньолінійні та внутрішньонелінійні моделі (рис. 4.3)

Рис. 4.3. Види регресійних моделей

Внутрішньолінійні моделі можуть бути приведені до лінійного вигляду за допомогою процедури логарифмування та подальшої заміни змінних. Так, наприклад, степенева функція $y = a_0 x^b \varepsilon$, де ε – залишки моделі, шляхом логарифмування ($lny = lna + b\ lnx + ln\varepsilon$) і найпростішої заміни ($y' = lny$, $a' = lna$, $x' = lnx$) приводиться до $y' = a' + b\ ' + \varepsilon$ і є внутрішньолінійною.

Внутрішньонелінійні моделі до лінійного виду приведені бути не можуть. Наприклад, якщо в ту ж степеневу модель включити ε вже як доданок, а не множник ($y = a_0 x^b + \varepsilon$), модель стане внутрішньонелінійною: логарифмування за моделлю нічого не дає – $y = \ln(ax^b + \varepsilon)$

Будь-яка квазілінійна модель у загальному випадку може бути представлена в такому вигляді:

$$\hat{y} = a_0 + a_1 f(x_1) + a_2 f(x_2) + \cdots + a_m f(x_m), \qquad (4.26)$$

де $f(x_j)$– нелінійні функції від x_j, j =1..m.

Квазілінійні моделі є більш простим і привабливим варіантом нелінійних моделей. Їх привабливість пояснюється тим, що оцінки параметрів таких моделей можуть бути отримані методами лінійного регресійного аналізу. Прикладом є модель попиту, для якої функція попиту має вигляд наступний вигляд:

$$Q = a_0 + a_1 P + a_2 P^2 \tag{4.27}$$

де Q – попит; P – ціна; a_0, a_1, a_2 – параметри моделі.

Нелінійні за факторами і за параметрами економетричні моделі є більш складними, тому оцінювання їх параметрів, як правило, виконується методами нелінійного регресійного аналізу. Прикладом може бути модель, що описує залежність між обсягом надходжень до бюджету і податковою ставкою на основі кривої Лаффера:

$$\hat{y} = a_0 e^{a_1 (x - \ _2)^2} \tag{4.28}$$

де $\hat{y}$ – податкові надходження; x – податкова ставка; a_0, a_1, a_2 – параметри моделі.

Якщо порівнювати лінійні та нелінійні економетричні моделі, то можна зробити висновок, що лінійна регресійна модель та відповідні методи оцінювання її параметрів, тестування і прогнозування в цілому теоретично краще обґрунтовані, ніж нелінійні, та мають відносно простий обчислювальний апарат. За необхідності застосування в аналізі нелінійних моделей практикують такі підходи:

1) замість складної нелінійної функціональної залежності навіть з невеликою кількістю факторів застосувати лінійну регресійну функцію з великою кількістю факторів;

2) за допомогою математичних співвідношень перетворити нелінійну функцію на лінійну.

Нелінійні функції, що можуть бути лінеарізовані

Процес приведення нелінійної регресійної моделі до лінійного вигляду називається лінеаризацією, а сама модель – лінеаризованою. Для отриманих моделей методологія економетричного дослідження така ж, як і для загальної лінійної регресійної моделі. Як вже було вказано вище, можливість лінеаризації моделі залежить від типу нелінійної моделі.

Для внутрішньолінійних моделей і моделей нелінійних щодо пояснювальних факторів параметри зазвичай намагаються отримати шляхом приведення їх до лінійного виду (цей процес називають процедурою лінеаризації). Лінеаризувавши залежність, її параметри можна оцінити за допомогою МНК.

Нелінійні щодо факторів, але лінійні за параметрами моделі

I. Степеневі адитивні залежності (поліноміальні функції). Загальний вид: $y = a_0 + \sum_{i=1}^{m}\left(a_{1i}x_i^1 + a_{2i}x_i^2 + \cdots + a_{ni}x_i^n\right) + \varepsilon$

В економіці часто використовують поліноміальні функції другого та третього степеня. Функції більш вищих степенів мають занадто велику кількість перегинів, які неможливо інтерпретувати за економічною природою явищ. Функції другого та третього степеня необхідно використовувати, коли виходячи з теоретичних передумов можна зробити висновок про зміну на певному інтервалі значень факторної ознаки характеру зв'язку між ним і результівною ознакою.

Кубічною функцією моделюють залежність загальних витрат від обсягу випуску продукції.

Такі функції можна лінеаризувати за допомогою переходу до нових змінних. Розглянемо процедуру лінеаризації на прикладі полінома другого степеня. Регресійна модель має вигляд:

$$y = a_0 + a_1 x^1 + a_2 x^2 + \varepsilon \quad (4.29)$$

Позначивши $x_1' = x^1$, $x_2' = x^2$ і підставивши нові змінні в модель (4.29), отримаємо $y = a_0 + a_1 x_1' + a_2 x_2' + \cdots + a_n x_n' + \varepsilon$. Параметри моделі легко отримати за МНК.

Квадратичні моделі використовуються для опису дуже широкого спектра економічних процесів завдяки їхнім універсальним можливостям. У загальному випадку квадратична модель має вигляд:

$$\hat{y} = a_0 + a_1 x + a_2 x^2 \quad (4.30)$$

де a_0, a_1, a_2– параметри моделі. Залежно від знаків і значень параметрів моделі вона може відображати еволюцію – дуже різну на різних проміжках інтервалу зміни пояснювальної ознаки х: зростання, спад, зростання з подальшим спадом, спад з подальшим зростанням. Так, квадратична модель може описувати кількісну залежність між обсягом випуску і середніми або граничними видатками, між витратами на рекламу та прибутками тощо.

II. Гіперболічні залежності (зворотні функції)

Загальний вид:

$$y = a_0 + \frac{b_1}{x_1} + \frac{b_2}{x_2} + \cdots + \frac{b_n}{x_n} + \varepsilon. \quad (4.31)$$

Гіперболи (або зворотні функції) зазвичай застосовують, коли необмежене збільшення пояснювальної змінної асимптотично наближає залежну змінну до деякої межі.

В економіці використовуються такі зворотні залежності:

- крива Філліпса описує взаємозв'язок ВВП і зайнятості і моделюється функцією $\boldsymbol{y = a_0 + \frac{a_1}{x} + \varepsilon}$;

- крива Енгеля описує зниження частки витрат на продовольчі товари зі збільшенням доходу домогосподарства, створюється за допомогою моделі

$$\boldsymbol{y = a_0 - \frac{a_1}{x} + \varepsilon.}$$

Обернено пропорційна однофакторна функція регресії має вигляд:

$$\hat{y} = a_0 + a_1 \frac{1}{x} \qquad \textit{(4.32)}$$

де a_0 , a_1 – параметри моделі.

Обернено пропорційна модель відноситься до квазілінійних моделей. У зв'язку з цим її лінеаризація здійснюється простою заміною змінної: $z = \frac{1}{x}$. Після підстановки в (4.32) нової змінної отримуємо таку парну лінійну регресію

$$\hat{y} = a_0 + a_1 z \qquad (4.33)$$

III. ***Лог-лінійні залежності лінійно-логаріфмічні моделі***

Прикладом застосування лог-лінійної залежності ($lny = a_0 + a_1\ x + \varepsilon$) в економіці є завдання аналізу банківських вкладів за початковим внеском і відсотковою ставкою. Лінеаризацію можна провести шляхом переходу до нової змінної $y' = lny$, що дозволить подати модель у виді: $y'=a_0+a_1x+\varepsilon$.

Лінійно-логарифмічні моделі ($y = a_0 + a_1 lnx + \varepsilon$) застосовуються, наприклад, для аналізу темпу зростання інфляції у результаті збільшення обсягу грошової маси в обігу. Заміна $x' = lnx$ дозволяє привести модель до лінійного виду.

Нелінійні за оцінюваними параметрами моделі

IV. Показникові залежності

Загальний вид:

$$y = a_0 \cdot a_1^{x_1} \cdot a_2^{x_2} \cdot \ldots \cdot a_n^{x_n} \cdot \varepsilon \qquad (4.34)$$

Лінеаризація моделі $y = a_0 a_1^{x_1} a_2^{x_2} *** a_n^{x_n} \varepsilon$ можлива шляхом логарифмування рівняння ($lny = lna_0 + x_1 lna_1 + x_2 lna_2 + \cdots + x_n lna_n + \varepsilon$) та заміни $y' = lny$, $a_i' = lna_i$ $j = \overline{1,n}$. $\rightarrow y' = a_0' + a_1' x_1 + a_2' x_2 + \cdots + a_n' x_n + \varepsilon$.

Приклад застосування в економіці: за допомогою експоненційних функцій (показникова функція з основою, що дорівнює основі нормального логарифма – $y = a_0 e^b \ \varepsilon$) моделюють вплив на виробництво науково-технічного прогресу, зростання чисельності населення території за часом тощо.

Показникова однофакторна модель регресії має декілька форм:

$$\hat{y} = a_0 a_1^x \tag{4.35}$$

$$\hat{y} = a_0 e^{a_1 x} \tag{4.36}$$

$$\hat{y} = a_0 (1 - a_1)^x \tag{4.37}$$

$$\hat{y} = e^{a_0 + \ _1 x} \tag{4.38}$$

Експоненціальна модель (4.36) використовується для опису швидкозростаючих або швидкоспадних економічних процесів. Найбільш типовим її застосуванням є ситуація, коли аналізується зміна результату у з постійним темпом приросту в часі. Лінеаризація показникових фукцій (4.35)–(4.38) виконується шляхом застосування операції логарифмування та подальшої заміни змінних, а саме:

$$\hat{y} = a_0 a_1^x \quad \rightarrow \quad ln\hat{y} = lna_0 + x \cdot lna_1 \tag{4.39}$$

$$\hat{y} = a_0 e^{a_1 x} \quad \rightarrow \quad ln\hat{y} = lna_0 + a_1 \cdot x \tag{4.40}$$

$$\hat{y} = a_0 (1 - a_1)^x \quad \rightarrow \quad ln\hat{y} = lna_0 + x \cdot \ln(1 - a_1) \tag{4.41}$$

$$\hat{y} = e^{a_0 + \ _1 x} \quad \rightarrow \quad ln\hat{y} = a_0 + a_1 \cdot x \tag{4.42}$$

V. Степеневі мультиплікативні залежності

Загальний вид:

$$y = a_0 x_1^{a_1} x_2^{a_2} * \ldots * x_n^{a_n} \varepsilon \tag{4.43}$$

Щоб лінеаризувати, необхідно прологарифмувати обидві частини рівняння - $lny = lna_0 + a_1 \cdot lnx_1 + a_2 \cdot lnx_1 + \cdots + a_n lnx_n + \varepsilon$ та провести заміну $y' = lny$, $a'_0 = lna_0$, $x'_i = lnx_i$, $i = \overline{1,n}$. Модифікована модель має лінійний вид $y' = a'_0 + a_1 x'_1 + a_2 x'_2 + \cdots + a_n x'_n + \varepsilon$.

Приклад застосування – виробнича функція Кобба – Дугласа $Y = a \cdot K^{a_1} L^{a_2}$.

Розглянкио степеневу однофакторну модель, вона описується функцією регресії такого вигляду

$$\hat{y} = a_0 x^{a_1} \tag{4.44}$$

де a_0, a_1 – параметри моделі.

Нелінійна модель на основі степеневої функції регресії є однією з найпоширеніших у практиці моделей і описує достатньо широкий спектр економічних явищ і процесів, таких як процес виробництва (виробничі функції), попит на товари різних категорій (криві Енгеля), для опису кривих байдужості тощо.

Лінеаризація функції (4.44), як було вже описано, здійснюється у два кроки. Спочатку виконується логарифмування її лівої та правої частин: $ln\hat{y} = lna_0 + a_1 lnx$, а потім здійснюється така заміна:$y_1 = ln\hat{y}$,$b_0 = lna_0$, $b_1 = a_1$,$x_1 = lnx$

У результаті нелінійна функція (4.44) зводиться до лінійної форми:

$$y_1 = b_0 + b_1 x_1 \tag{4.45}$$

Параметр a_1 у степеневій моделі характеризує еластичність змінної у за змінною x, тобто цей параметр фактично дорівнює коефіцієнту еластичності у за x. Тому часто степеневу модель ще називають моделлю сталої еластичності, що вказує на можливі напрями її застосування.

Крім описаних функцій також використовуються:

1. ***Модифіковані показникові функції.***

   Загальний вид: $y = k + a_0 a_1^{x_1} \varepsilon$.

2. ***Криві Гомперца.***

   Загальний вид: $y = a_0 a_1{}^{a_2^x} \varepsilon$.

3. ***Логістичні криві (криві Перла – Рида*).**

   Загальний вид: $y = \frac{k}{1 + {}_1 a_2^x}$.

4.11. Система лінійних одночасних рівнянь

Системи лінійних одночасних рівнянь є важливим інструментом в економетриці та економічному аналізі для вивчення взаємозв'язків між декількома змінними в одному і тому ж часовому періоді. Ці системи використовуються для моделювання складних економічних процесів та встановлення взаємозв'язків між різними змінними, які впливають одна на одну.

Основні риси систем лінійних одночасних рівнянь:

1. Взаємодія змінних. В системах лінійних одночасних рівнянь різні змінні можуть взаємодіяти одна з одною, впливаючи на свої значення через спільний вплив.
2. Матрична форма. Системи лінійних одночасних рівнянь можуть бути представлені у матричній формі, де кожна рівність представлена у вигляді лінійної комбінації змінних та коефіцієнтів.
3. Ідентифікація. Важливою властивістю є ідентифікація, тобто можливість однозначно визначити значення параметрів моделі за наявності даних.
4. Ендогенні та екзогенні змінні. В системах лінійних одночасних рівнянь зазвичай виділяються ендогенні (змінні, які пояснюються в межах моделі) та екзогенні (змінні, які вважаються відомими та не залежать від інших змінних).

5. Методи оцінки. Для оцінки параметрів систем лінійних одночасних рівнянь застосовують методи, такі як метод найменших квадратів, метод інструментальних змінних та інші.
6. Економічні застосування. Системи лінійних одночасних рівнянь застосовуються для моделювання економічних систем, таких як системи виробництва, споживання, обміну, фінансів тощо.

Ці системи грають важливу роль у розв'язанні економічних питань, де декілька змінних взаємодіють та впливають на динаміку системи, та допомагають зрозуміти складні ефекти та взаємозалежності в економіці.

Система нелінійних незалежних рівнянь:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1 \\ y_2 = a_{20} + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2 \\ \cdots\cdots\cdots\cdots\cdots \\ y_k = a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{km}x_m + \varepsilon_k \end{cases}$$

Система лінійних рекурсивних рівнянь:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1 \\ y_2 = b_{21}y_1 + a_{20} + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2 \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{30} + a_{31}x_1 + a_{32}x_2 + \cdots + a_{3m}x_m + \varepsilon_3 \\ \cdots\cdots\cdots\cdots\cdots \\ y_k = b_{k1}y_1 + b_{k2}y_2 + \ldots + b_{kk-1}y_{k-1} + a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \cdots \\ \qquad + a_{km}x_m + \varepsilon_k \end{cases}$$

Система одночасних (взаємозалежних) рівнянь

$$\begin{cases} y_1 = a_{10} + b_{12}y_2 + b_{13}y_3 + \cdots + b_{1k}y_k + a_{11}x_1 + a_{12}x_2 + \cdots \\ \qquad + a_{1m}x_m + \varepsilon_1 \\ y_2 = a_{20} + b_{21}y_1 + b_{23}y_3 + \cdots + b_{2k}y_k + a_{21}x_1 + a_{22}x_2 + \cdots \\ \qquad + a_{2m}x_m + \varepsilon_2 \\ y_3 = a_{30} + b_{31}y_1 + b_{32}y_2 + \cdots + b_{3k}y_k + a_{31}x_1 + a_{32}x_2 + \cdots \end{cases}$$

$$+a_{3m}x_m + \varepsilon_3$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$y_k = a_{k0} + b_{k1}y_1 + b_{k2}y_2 + \cdots + b_{kk-1}y_{k-1} + a_{k1}x_1 + a_{k2}x_2 + \cdots$$

$$+a_{km}x_m + \varepsilon_k$$

Приведена форма моделі

$$\begin{cases} \boldsymbol{y_1 = \delta_{10} + \delta_{11}x_1 + \delta_{12}x_2 + \cdots + \delta_1 \ x_m + \varepsilon_1} \\ \boldsymbol{y_2 = \delta_{20} + \delta_{21}x_1 + \delta_{22}x_2 + \cdots + \delta_2 \ x_m + \varepsilon_2} \\ \boldsymbol{\cdots\cdots} \\ \boldsymbol{y_k = \delta_{k0} + \delta_{k1}x_1 + \delta_{k2}x_2 + \cdots + \delta_k \ x_m + \varepsilon_k} \end{cases}$$

4.12. Проблема ідентифікації

Для того, щоб модель була такою, що ідентифікується, необхідно, щоб кожне рівняння моделі було ідентифіковане. Якщо хоча б одне рівняння СФМ неідентифіковане, то вся модель вважається неідентифікованою.

Необхідною умовою ідентифікації окремого рівняння моделі є рахункове правило. Якщо позначити через H число досліджуваних змінних $\boldsymbol{y_i}$, присутніх в і-му рівнянні, а через D позначити число факторних змінних x_j, відсутніх в і-му рівнянні, то рахункове правило має вигляд:

- якщо D + 1 < H, то рівняння неідентифіковані;
- якщо D + 1 = H, то рівняння ідентифіковані;
- якщо D + 1> H, то рівняння над ідентифіковані.

Достатня умова ідентифікованих окремого рівняння моделі виконується, якщо визначник матриці, складеної з коефіцієнтів в інших рівняннях при змінних (як досліджуваних у, так і факторних х), відсутніх в даному і-му рівнянні не дорівнює нулю, а ранг цієї матриці, водночас, не менше, ніж кількість всіх досліджуваних змінних в системі рівнянь за виключенням одного.

4.13. Індивідуальне завдання № 9

Засвоєння методики визначення коефіцієнтів нелінійної однофакторної моделі

Мета роботи: Набути навичок з уміння розрахувати коефіцієнтіи нелінійної однофакторної моделі

Порядок виконання:

1. Знайти коефіцієнти моделей Y1= f(X1), Y1= f(X2), Y1= f(X3), Y2= f(X1), Y2= f(X2), Y2= f(X3) для гіперболічної, степеневої та показової моделей.
2. Оцінити якість рівняння регресій (R^2) для всіх формул та значущість коефіцієнтів.

Таблиця 4.10. Вихідні дані

X1, тис. грн	Y1, тис. грн	X2, тис. грн	X3, тис. грн	Y1, тис. грн
100+20*N	70+5*N	75+15*N	105+12*N	170+7*N
130+10*N	90+5*N	80+15*N	90+12*N	180+7*N
150+10*N	105+5*N	115+15*N	145+12*N	165+7*N
160+10*N	95+5*N	100+15*N	170+12*N	175+7*N
175+10*N	130+5*N	120+15*N	175+12*N	185+7*N
180+10*N	120+5*N	110+15*N	170+12*N	190+7*N
185+10*N	140+5*N	100+15*N	165+12*N	200+7*N
200+10*N	150+5*N	140+15*N	210+12*N	195+7*N
215+10*N	150+5*N	160+15*N	220+12*N	200+7*N
220+10*N	145+5*N	150+15*N	215+12*N	205+7*N
240+10*N	165+5*N	185+15*N	210+12*N	210+7*N
250+10*N	172+5*N	152+15*N	220+12*N	220+7*N
260+10*N	155+5*N	110+15*N	225+12*N	215+7*N
270+10*N	143+5*N	120+15*N	230+12*N	225+7*N
280+10*N	167+5*N	140+15*N	280+12*N	230+7*N

Методичні вказівки

Для однофакторної моделі $y = a_0 + \frac{a_1}{x} + \varepsilon$ заміна $x' = \frac{1}{x}$ дасть таку залежність від нових змінних: $y = a_0 + a_1 x' + \varepsilon$. Таким чином, будемо шукати параметри регресії рівняння

$$\hat{y} = a_0 + a_1 x' \quad (4.46)$$

Степенева однофакторна модель описується функцією регресії такого вигляду

$$\hat{y} = a_0 x^{a_1} \quad (4.47)$$

де a_0, a_1 – параметри моделі.

Лінеаризація функції (4.47) здійснюється у два кроки. Спочатку виконується логарифмування її лівої та правої частин:

$$ln\hat{y} = lna_0 + a_1 lnx$$

а потім здійснюється така заміна

$y_1 = ln\hat{y}, b_0 = lna_0, b_1 = a_1, x_1 = lnx$

У результаті нелінійна функція (4.47) зводиться до лінійної форми:

$$y_1 = b_0 + b_1 x_1$$

Для однофакторної показникової моделі $\boldsymbol{y = a_0 a_1^{x_1} \varepsilon}$ лінеаризація буде наступним чином проводитись $lny = ln\boldsymbol{a_0} + \boldsymbol{x_1} ln\boldsymbol{a_1}$

$y_1 = ln\hat{y}, b_0 = lna_0, b_1 = lna_1, x_1 = x$

$$y_1 = b_0 + b_1 x_1$$

Розрахуємо для y=f(x) гіперболічну, степеневу та показову моделі.

Для гіперболічної моделі робимо заміну $\boldsymbol{x' = \frac{1}{x}}$.

Для степеневої однофакторної потрібні заміни $\boldsymbol{y_1 = l\ \hat{y}, x_1 = l\ x}$.

Для показової моделі необхідна заміна $\boldsymbol{y_1 = l\ \hat{y}}$.

Значення x, y та необхідних замін представлено у таблиці 4.11.

Таблиця 4.11 – Вихідні дані прикладу та заміни

x	y	x'=1/x	lnx	lny
3	26	0.333	1.099	3.258
7	30	0.143	1.946	3.401
10	33	0.100	2.303	3.497
20	49	0.050	2.996	3.892
40	78	0.025	3.689	4.357

x	y	x'=1/x	lnx	lny
52	81	0.019	3.951	4.394
70	88	0.014	4.248	4.477
75	92	0.013	4.317	4.522
80	118	0.013	4.382	4.771
86	138	0.012	4.454	4.927

За допомогою Regression (Регресія) розрахуємо для гіперболічної функції параметри регресії. В якості y беремо y, в якості x беремо $\boldsymbol{x' = \frac{1}{x}}$. Результати в таблиці 4.12.

Таблиця 4.12 – Результати застосування регресії для визначення гіперболічної функції

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.717219
R Square	0.514403
Adjusted R Square	0.453703
Standard Error	28.26114
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	6768.564442	6768.56444	8.47456204	0.01955580
Residual	8	6389.535558	798.691944		
Total	9	13158.1			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	92.72675	11.15357871	8.3136323	3.30748E-05	67.006554	118.446951
X Variable 1	-269.006	92.4066484	-2.9111101	0.0195558	-482.0960	-55.91581

Як видно з розрахунків коефіцієнт детермінації 0.514. Виходячи з того, p-значення (significance-F) для F менше 0.05, модель значуща.

Для коефіцієнтів p-значення теж менше 0.05, тому вони є значущими.

$\boldsymbol{a_0 = 9\ .7\ \ 7, a_1 = -2\ \ 9.00\ .}$

$$\hat{y} = 92.727 - 269.006\frac{1}{x}.$$

За допомогою Regression (Регресія) розрахуємо для степеневої функції параметри регресії. В якості у беремо $\boldsymbol{l\ \ \hat{y}}$, в якості х беремо $\boldsymbol{l\ \ x}$. Результати в таблиці 4.13.

Таблиця 4.13 – Результати застосування регресії для визначення степеневої функції

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.97279488
R Square	0.94632988
Adjusted R Square	0.93962111
Standard Error	0.14622501
Observations	8
	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3.0160833	3.0160833	141.05873	2.31916E-06
Residual	8	0.171054	0.0213818		
Total	9	3.1871373			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.52682882	0.14424471	17.517652 3	1.151E-07	2.19419990 5	2.85945773 6
X Variable 1	0.48606858	0.04092584	11.876814 6	2.319E-06	0.39169343 2	0.58044373 1

Отже, коефіцієнт детермінації 0.946. Виходячи з того, р-значення (significance-F) для F менше 0.05, модель значуща.

Для коефіцієнтів р-значення теж менше 0.05, тому вони є значущими.

$\boldsymbol{l\ \ a_0 = 2.5\ \ 7, a_0 = e^{2.5\ \ 7} = 12.5\ \ 3, a_1 = 0.486}$. Отже,

$$\boldsymbol{\hat{y} = 12.5\ \ 3 * x^{0.486}}.$$

За допомогою Regression (Регресія) розрахуємо для показникової функції параметри регресії. В якості y беремо lny, в якості x беремо x. Результати в таблиці 4.14.

Таблиця 4.14 – Результати застосування регресії для визначення показникової функції

Regression Statistics	
Multiple R	0.967387
R Square	0.935838
Adjusted R Square	0.927817
Standard Error	0.159881
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1.00000	2.98264	2.98264	116.68359	0.00000
Residual	8.00000	0.20449	0.02556		
Total	9.00000	3.18714			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3.36702	0.08834	38.11282	0.00000	3.16330	3.57074
X Variable 1	0.01767	0.00164	10.80202	0.00000	0.01389	0.02144

Отже, коефіцієнт детермінації 0.936. Виходячи з того, p-значення (significance-F) для F менше 0.05, модель значуща.

Для коефіцієнтів p-значення теж менше 0.05, тому вони є значущими.

Отже, $\boldsymbol{l\ a_0 = 3.3\ 7, a_0 = e^{3.3\ 7} = 2\ .99,}$ $\boldsymbol{l\ a_1 = 0.01\ 7, a_1 = e^{0.01\ 7} = 1.01\ .}$

$$\boldsymbol{\hat{y} = 2\ .99 * 0.01\ 7^x.}$$

Аналізуючи отримані рівняння, найбільший коефіцієнт детермінації має степенева однофакторна модель, отже вона краще описує заданий процес.

4.14. Індивідуальне завдання № 10

Засвоєння методики розрахунку коефіцієнтів моделей множинної регресії

Мета роботи: Набути навичок з уміння розраховувати коефіцієнти моделей множинної регресії

Порядок виконання:

1. Знайти коефіцієнти моделей Y1= f(X1, X2, X3), Y2= f(X1, X2, X3) для лінійних та гіперболічної, степеневої та показової моделей за допомогою додатку Regression (Регресія) електронних таблиць Microsoft Excel.

2. Оцінити якість апроксимації та оцінити значущість коефіцієнтів моделей.

Методичні вказівки

Гіперболічні залежності (зворотні функції) мають загальний вид: $y = a_0 + \frac{b_1}{x_1} + \frac{b_2}{x_2} + \cdots + \frac{b_n}{x_n} + \varepsilon$. Щоб лінеаризувати зворотню функцію, достатньо зробити заміну $x_i' = \frac{1}{x_i}$.

Таким чином, будемо шукати параметри регресії рівняння $\hat{y} = a_0 + a_1 x_1' + a_2 x'_2 + \cdots + a_n x'_n$.

Степеневі мультиплікативні залежності мають загальний вид:$y = a_0 x_1^{a_1} x_2^{a_2} * \ldots * x_n^{a_n} \varepsilon$

Для лінеаризації слід прологарифмувати обидві частини рівняння - $lny = lna_0 + a_1 \cdot lnx_1 + a_2 \cdot lnx_1 + \cdots + a_n lnx_n + \varepsilon$ та провести заміну $y' = lny$, $a'_0 = lna_0$, $x'_i = lnx_i$, $i = \overline{1,n}$. Модифікована модель має лінійний вид $y' = a'_0 + a_1 x'_1 + a_2 x'_2 + \cdots + a_n x'_n + \varepsilon$.

Лінеаризація показникової моделі $y = a_0 a_1^{x_1} a_2^{x_2} \cdots a_n^{x_n} \varepsilon$ можлива шляхом логарифмування рівняння ($lny = lna_0 + x_1 lna_1 + x_2 lna_2 + \cdots + x_n lna_n + \varepsilon$) та заміни $y' = lny$, $a_i' = lna_i$ $j = \overline{1,n}$. $\rightarrow y' = a_0' + a_1' x_1 + a_2' x_2 + \cdots + a_n' x_n + \varepsilon$.

Розрахуємо для y=f(x) гіперболічну, степеневу та показникову моделі.

Для гіперболічної моделі робимо заміну $x_1{}' = \frac{1}{x_1}, x_2{}' = \frac{1}{x_2}, x_3{}' = \frac{1}{x_3}$.

Для степеневої однофакторної потрібні заміни $y_1 = ln\hat{y}, x_1' = lnx_1, x_2' = lnx_2, x_3' = lnx_3$.

Для показникової моделі необхідна заміна $y_1 = ln\hat{y}$.

Значення x, y та необхідних замін представлено у таблиці 4.15.

Таблиця 4.15. Вихідні дані прикладу та заміни

x1	x2	x3	y	1/x1	1/x2	1/x3	lnx1	lnx2	lnx3	lny
1	3	5	26	1.0000	0.3333	0.2000	0.000	1.099	1.609	3.258
2	7	7	30	0.5000	0.1429	0.1429	0.693	1.946	1.946	3.401
3	10	9	33	0.3333	0.1000	0.1111	1.099	2.303	2.197	3.497
4	20	11	19	0.2500	0.0500	0.0909	1.386	2.996	2.398	2.944
5	4	14	78	0.2000	0.2500	0.0714	1.609	1.386	2.639	4.357
6	9	17	81	0.1667	0.1111	0.0588	1.792	2.197	2.833	4.394
7	11	23	88	0.1429	0.0909	0.0435	1.946	2.398	3.135	4.477
8	15	24	92	0.1250	0.0667	0.0417	2.079	2.708	3.178	4.522
9	13	28	118	0.1111	0.0769	0.0357	2.197	2.565	3.332	4.771
10	11	33	138	0.1000	0.0909	0.0303	2.303	2.398	3.497	4.927

За допомогою Regression (Регресія) розрахуємо для лінійної функції параметри регресії. Результати представлено у таблиці 4.16.

Таблиця 4.16. Результати застосування регресії для визначення лінійної функції

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.998904
R Square	0.997809
Adjusted R Square	0.996714
Standard Error	2.373332
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	15392.3	5130.768	910.8886	2.29841E-08
Residual	6	33.79624	5.632706		
Total	9	15426.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	17.57273	2.006213	8.759152	0.000123	12.6636988	22.48175
X Variable 1	11.9686	1.956502	6.117344	0.000871	7.181207861	16.75598
X Variable 2	-2.8499	0.184633	-15.4355	4.68E-06	-3.30168483	-2.39813
X Variable 3	0.950528	0.607574	1.564465	0.168743	-0.5361514	2.437207

Отже, коефіцієнт детермінації 0.998. Виходячи з того, р-значення (significance-F) для F менше 0.05, модель значуща.

Для коефіцієнтів р-значення менше 0.05, для зсуву, $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ А коефіцієнт при $\boldsymbol{x_3}$ за критерієм Стьюдента на рівні значущості 0.05 незначущий. Рівняння регресії:

$$\hat{\boldsymbol{y}} = \mathbf{1}\ .\mathbf{57} + \mathbf{11.97}\boldsymbol{x_1} - \mathbf{2.85}\boldsymbol{x_2} + \mathbf{0.95}\boldsymbol{x_3}$$

Далі за допомогою Regression (Регресія) розрахуємо для гіперболічної функції параметри регресії, використовуючі заміни $x_i' = \frac{1}{x_i}$. Результати в таблиці 4.17.

Таблиця 4.17 – Результати застосування регресії для визначення гіперболічної функції

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.941976
R Square	0.887319
Adjusted R Square	0.830979
Standard Error	17.02069
Observations	10

ANOVA

	df	SS	MS	F	Significance F
Regression	3	13687.88	4562.625	15.74927	0.002994
Residual	6	1738.224	289.7039		
Total	9	15426.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	139.0634	15.6053	8.91129	0.000111	100.8786	177.2482
X Variable 1	195.3082	87.29442	2.23735	0.066589	-18.2936	408.9099
X Variable 2	96.97072	102.3649	0.947304	0.380057	-153.507	347.4486
X Variable 3	-1678.56	389.6485	-4.30788	0.005049	-2632	-725.124

Отже, коефіцієнт детермінації 0.8873. Виходячи з того, р-значення (significance-F) для F менше 0.05, модель значуща.

Для коефіцієнтів р-значення менше 0.05, тільки для зсуву та $\boldsymbol{x_3}$. Тому коефіцієнти при $\boldsymbol{x_1}$ та $\boldsymbol{x_2}$ за критерієм Стьюдента на рівні значущості 0.05 незначущі.

Отже, $\hat{\boldsymbol{y}} = \mathbf{13}\ .\mathbf{0} + \frac{\mathbf{1}\ \mathbf{5.3}}{\boldsymbol{x_1}} + \frac{\mathbf{96.97}}{\boldsymbol{x_2}} - \frac{\mathbf{1}\ \mathbf{78.56}}{\boldsymbol{x_3}}$

Далі за допомогою Regression (Регресія) розрахуємо для степеневої функції параметри регресії, використовуючі заміни $y' = lny$, $\boldsymbol{a'_0} = ln\boldsymbol{a_0}$, $\boldsymbol{x'_i} = ln\boldsymbol{x_i}$, $i = \overline{\mathbf{1},\boldsymbol{n}}$. Результати представлено у таблиці 4.18.

Таблиця 4.18. Результати застосування регресії для визначення степеневої функції

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.979619
R Square	0.959653
Adjusted R Square	0.939479
Standard Error	0.173632
Observations	10

ANOVA

	df	SS	MS	F	Significance F
Regression	3	4.302429	1.434143	47.56994	0.000141
Residual	6	0.180889	0.030148		
Total	9	4.483317			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.79328	0.767901	2.335301	0.058217	-0.08571	3.672266
X Variable 1	-0.0452	0.434319	-0.10407	0.920507	-1.10794	1.017542

X Variable 2	-0.5957	0.136321	-4.36982	0.00472	-0.92926	-0.26213
X Variable 3	1.360016	0.474589	2.865674	0.028588	0.19874	2.521292

Отже, коефіцієнт детермінації 0.9597. Виходячи з того, р-значення (significance-F) для F менше 0.05, модель значуща.

Для коефіцієнтів р-значення менше 0.05, тому тільки для $\boldsymbol{x_2}$ та $\boldsymbol{x_3}$, тому параметр зсуву та коефіцієнти при $\boldsymbol{x_1}$ за критерієм Стьюдента на рівні значущості 0.05 незначущі.

$\boldsymbol{l\ \ a_0 = 1.79\ \ , a_0 = e^{1.79}\ \ = 6.00\ \ , a_1 = -0.0\ \ 5\ \ , a_2 = -0.5957. a_3 = 1.3}$ Отже,

$\hat{y} = 6.009 * {x_1}^{-0.045} x_2^{-0.5\ \ 6} x_3^{1.3}$.

Нарешті, за допомогою Regression (Регресія) розрахуємо для показникової функції параметри регресії. В якості у беремо lny, в якості х беремо х. Результати в таблиці 4.19.

Таблиця 4.19. Результати застосування регресії для визначення показникової функції

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.990516
R Square	0.981122
Adjusted R Square	0.971683
Standard Error	0.118769
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	4.39868	1.466227	103.9421	1.46E-05
Residual	6	0.084637	0.014106		
Total	9	4.483317			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3.355231	0.100398	33.41944	4.78E-08	3.109567	3.600895
X Variable 1	0.415128	0.09791	4.239898	0.00544	0.175551	0.654705
X Variable 2	-0.0697	0.00924	-7.54405	0.000281	-0.09231	-0.0471

X Variable 3	-0.05062	0.030405	-1.66492	0.146983	-0.12502	0.023776

Отже, коефіцієнт детермінації 0.981. Виходячи з того, р-значення (significance-F) для F менше 0.05, модель значуща.

Для коефіцієнтів р-значення менше 0.05, для всіх крім коефіцієнту при $\boldsymbol{x_3}$, тому за критерієм Стьюдента на рівні значущості 0.05 тільки він незначущий, а інші значущі.

$\boldsymbol{a_0 = e^{3.3\ 5} = 2\ .65, a_1 = e^{0.4\ 5} = 1.5\ 5,\quad a_2 = e^{-\ .0\ 97} = 0.9\ , a_3 = e^{-\ .0\ } = 0.95}$. Отже,

$$\hat{y} = 28.65 * 1.515^{x_1} * 0.93^{x_2} * 0.95^{x_3}.$$

Висновок

Не зважаючи на те, що всі види моделей статистично значущі, немає моделі, де б усі параметри теж були статистично значущі. Найбільший коефіцієнт детермінації має лінійна модель і тільки один незначущий коефіцієнт при $\boldsymbol{x_3}$.

4.15. Індивідуальне завдання № 11

Засвоєння методики визначення впливу факторів у моделях множинної регресії

Мета роботи: Набути навичок з уміння визначення вплив факторів у моделях множинної регресії

Порядок виконання:

Для моделей моделей Y1= f(X1, X2, X3), Y2= f(X1, X2, X3), для лінійних та гіперболічної, степеневої та показової моделей, отриманих у попередній роботі за допомогою додатку Regression (Регресія) електронних таблиць Microsoft Excel визначити:

1. Наявність колінеарності у тому числі й методом Феррара-Глобера.

Дані для розрахунку у таблиці 4.10. Приклад розв'язання представлено у пункті 4.7.

4.16. Індивідуальне завдання № 12

Засвоєння методики визначення прогнозів у моделях множинної регресії

Мета роботи: Набути навичок з уміння визначення прогнозів у моделях множинної регресії

Порядок виконання:

Для моделей моделей Y1= f(X1, X2, X3), Y2= f(X1, X2, X3), для лінійних та гіперболічної, степеневої та показової моделей, отриманих у попередній роботі

1. Для всіх лінійних та нелінійних моделей знайти точковий прогноз для точок Xi > 1,2Xmax та Xi < 1,2Xmin.
2. Для всіх лінійних та нелінійних моделей моделей знайти довірчий інтервал для всіх точкових прогнозів для рівнів значущості 0,15; 0,1 та 0,05.

Методичні матеріали у пункті 4.6. Розглянемо приклад.

На рисунку 4.4 представлені дані, для яких була побудована лінійна модель

$$Y = 17.6 + 12 * x_1 - 2.85 * x_2 + 0.95 * x_3$$

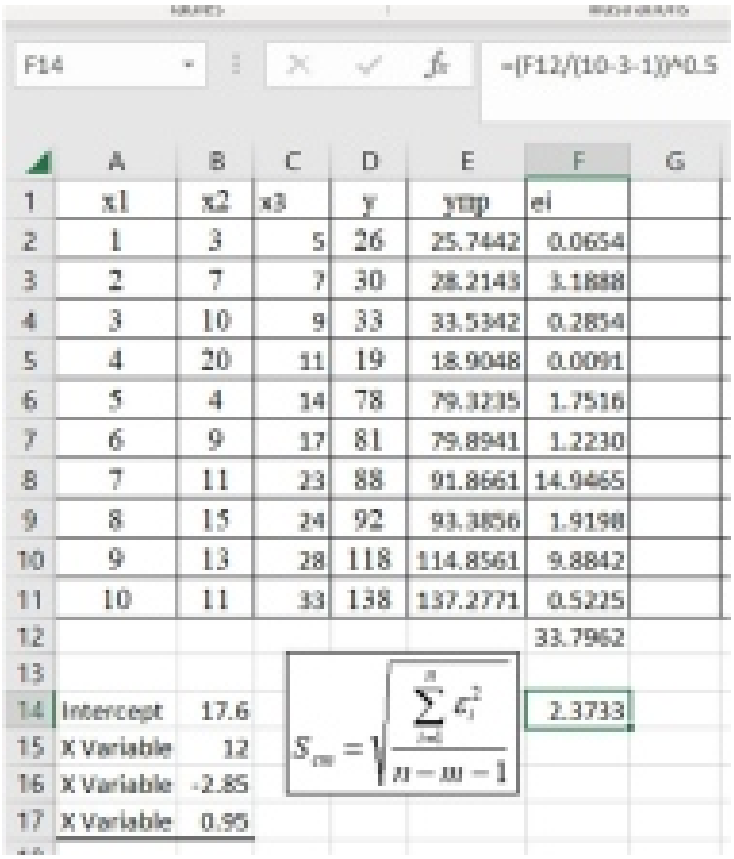


Рис. 4.4. Дані та розрахунок

$S_{\text{ст}} = 2.37.$

Побудуємо прогноз для $x_1 = 11, x_2 = 12, x_3 = 32$

Точковий прогноз:

$Y = 17.6 + 12 * 11 - 2.85 * 12 + 0.95 * 32 = 145,45$

Для інтервально прогнозу розрахуємо спочатку $X_{\text{пр}}^{\text{T}} B^{-1} X_{\text{пр}}$.

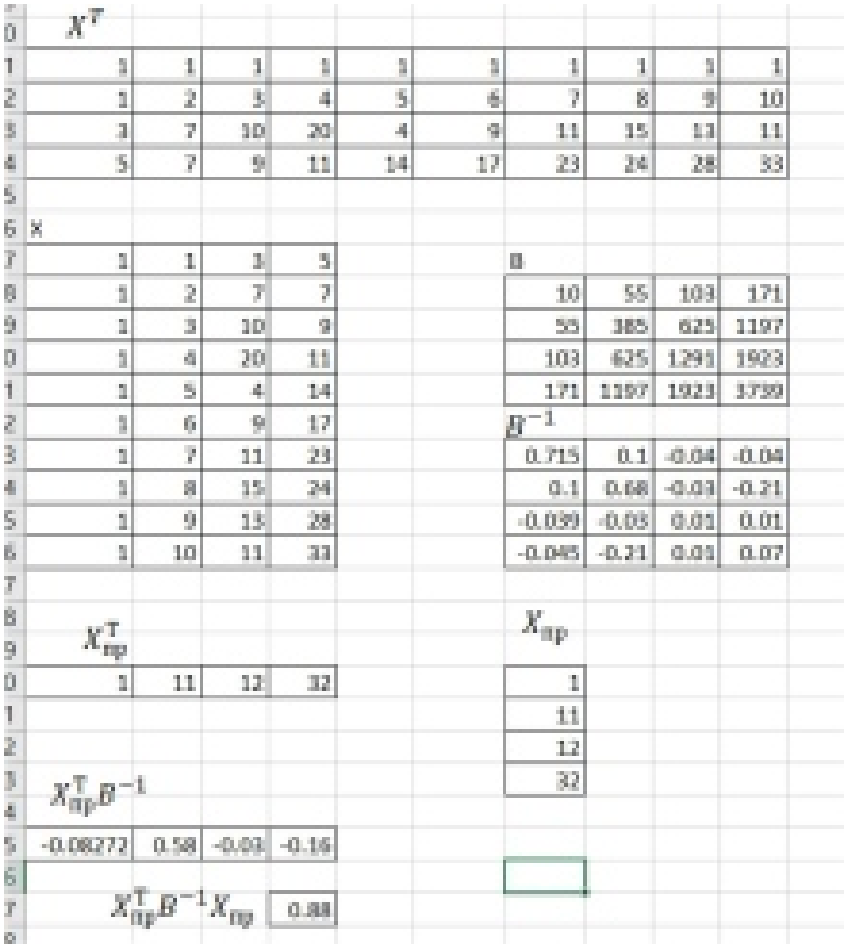
Нижче приведено розрахунок в Excel (див.рис.4.5).

Далі розрахуємо довірчий інтервал на рівні значущості 0.5.

$$t_{0.05,10-3-1} = 2.4469.$$

$$\Delta_{\text{пр}} = 2.37 * 2.4469\sqrt{0.88} = 5.45$$

Тому довірчий інтервал на рівні значущості 0.05 становить [145.45-5.45, 145.45+5.45] тобто [140, 150.9].



X^T

1	1	1	1	1	1	1	1	1	1
2	1	2	3	4	5	6	7	8	9	10
3	1	7	10	20	4	9	11	15	13	11
4	5	7	9	11	14	17	23	24	28	33

X

1	1	1	5
1	2	7	7
1	3	10	9
1	4	20	11
1	5	4	14
1	6	9	17
1	7	11	23
1	8	15	24
1	9	13	28
1	10	11	33

B

10	55	103	171
55	385	625	1197
103	625	1291	1923
171	1197	1923	3799

B^{-1}

0.715	0.1	-0.04	-0.04
0.1	0.68	-0.03	-0.21
-0.039	-0.03	0.01	0.01
-0.045	-0.21	0.01	0.07

$X_{пр}^T$

1	11	12	32

$X_{пр}$

1
11
12
32

$X_{пр}^T B^{-1}$

-0.08272	0.58	-0.03	-0.16

$X_{пр}^T B^{-1} X_{пр}$ 0.88

Рис. 4.5. Розрахунок прогнозу

Аналогічно будуємо прогноз на рівні значущості 0.1 та 0.15.

4.17. Індивідуальне завдання № 13

Засвоєння методики визначення впливу окремих факторів на змінну

Мета роботи: Набути навичок з уміння оцінки впливу окремих факторів на досліджувану змінну за коефіцієнтами еластичності, бета та дельта.

Порядок виконання:

Для моделей моделей Y1= f(X1, X2, X3), Y2= f(X1, X2, X3), для лінійних та гіперболічної, степеневої та показової моделей, отриманих у попередній роботі за допомогою додатку Regression (Регресія) електронних таблиць Microsoft Excel визначити:

Оцінити вплив окремих факторів на досліджувану змінну за коефіцієнтами еластичності, бета та дельта.

Методичні матеріали

Оцінку впливу окремих факторів на досліджувану змінну розглянуто в пункті 4.8. За формулами 4.11-4.15 розрахуємо коефіцієнти еластичності, бета–коефіцієнти та дельта – коефіцієнти. Вихідні дані представлено у таблиці 4.20.

Таблиця 4.20. Вихідні дані

	X1	X2	X3	Y1
	272	239.3	226.7	44.2
	261	387.2	208.3	48.0
	152	382.8	140.8	48.0
	220	320.5	201.7	38.0
	130	378.5	198.3	40.2
	186	314.7	152.5	48.7
	261	339.3	100.8	42.2
	216	366.9	112.5	60.2
	204	249.4	190.8	59.3
	249	211.7	174.2	43.3
	130	374.1	116.7	35.6
	123	207.4	110.0	39.6
	172	368.3	195.0	58.4
	136	381.4	135.0	27.3
	169	229.1	146.7	31.6
	245	321.9	110.0	61.6
	274	223.3	115.0	41.6
	178	316.1	119.2	51.8
	225	268.3	111.7	43.6
	150	192.9	231.7	46.0
	211	278.4	195.8	47.8
	250	234.9	135.8	34.9
	242	287.1	177.5	51.6
	163	332.1	221.7	46.4
	186	226.2	204.2	39.3
Середнє =	200.2	297.25	161.3	45.164

Для лінійної моделі регресії були знайдені коефіцієнти за допомогою Регрессії (табл.4.21).

Таблиця 4.21. Коефіцієнти регресії

Коефіцієнти	
a1	0.054674584
a2	0.033733799
a3	0.021975424

Далі представлено проміжні розрахунки в таблиці 4.22.

Таблиця 4.22. Проміжні розрахунки

X1-X1ср	X2-X2ср	X3-X3ср	Y1-Y1ср	(X1-X1ср)^2	(X2-X2ср)^2	(X3-X3ср)^2	(Y1-Y1ср)^2
71.8	-58.0	65.4	-0.9	5155.24	3364.0	4272.8	0.9
60.8	89.9	47.0	2.8	3696.64	8082.0	2212.1	8.0
-48.2	85.6	-20.5	2.8	2323.24	7318.8	418.9	8.0
19.8	23.2	40.4	-7.2	392.04	538.2	1629.5	51.3
-70.2	81.2	37.0	-4.9	4928.04	6593.4	1371.5	24.4
-14.2	17.4	-8.8	3.5	201.64	302.8	77.4	12.3
60.8	42.1	-60.5	-2.9	3696.64	1768.2	3656.2	8.7
15.8	69.6	-48.8	15.1	249.64	4844.2	2381.4	226.7
3.8	-47.9	29.5	14.2	14.44	2289.6	872.2	200.8
48.8	-85.6	12.9	-1.8	2381.44	7318.8	165.6	3.4
-70.2	76.8	-44.6	-9.6	4928.04	5905.9	1992.1	92.3
-77.2	-89.9	-51.3	-5.6	5959.84	8082.0	2631.7	31.5
-28.2	71.1	33.7	13.3	795.24	5048.1	1135.7	176.4
-64.2	84.1	-26.3	-17.8	4121.64	7072.8	691.7	317.9
-31.2	-68.2	-14.6	-13.6	973.44	4644.4	214.1	185.2
44.8	24.6	-51.3	16.4	2007.04	607.6	2631.7	268.7
73.8	-74.0	-46.3	-3.6	5446.44	5468.6	2143.7	13.0
-22.2	18.8	-42.1	6.6	492.84	355.3	1775.2	43.7
24.8	-29.0	-49.6	-1.6	615.04	841.0	2463.5	2.6
-50.2	-104.4	70.4	0.8	2520.04	10899.4	4951.5	0.7
10.8	-18.9	34.5	2.6	116.64	355.3	1192.6	6.8
49.8	-62.4	-25.5	-10.3	2480.04	3887.5	648.6	105.6
41.8	-10.2	16.2	6.4	1747.24	103.0	262.4	40.8
-37.2	34.8	60.4	1.3	1383.84	1211.0	3644.1	1.6
-14.2	-71.1	42.9	-5.8	201.64	5048.1	1837.6	34.0
				56828	101950.2	45273.7	1865.4

Розрахунок коефіцієнтів за формулами представлено нижче (див. табл. 4.23-4.24).

Таблиця 4.23. Розрахунок коефіцієнту еластичності

Коефіцієнти еластичності $E_j = a_j \times \frac{x_{j\ p}}{y_{cp}}$	
Э1	0.242355503
Э2	0.222019153
Э3	0.078482886

Таблиця 4.24. Розрахунко бета–коефіцієнтів та дельта – коефіцієнтів

	$S_{x_j} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - x_{j\ p}\right)^2}$	$S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - y_{cp}\right)^2}$	$\beta_j = a_j \times \frac{S_{x_j}}{S_y}$
1	48.66038772	8.81620282	0.30177237
2	65.17611558		0.24938605
3	43.4327652		0.10826129

$\Delta_j = r(x_j, y) \times \frac{\beta_j}{R^2}$

Δ1	0.596467024
Δ2	0.341393988
Δ3	0.062138987

R^2= 0.123216119

За розрахованими коефіцієнтами еластичності, бета-коефіцієнтом та дельта-коефіцієнтом, ми бачимо, що найбільший вплив має незалежна змінна X1.

4.18. Індивідуальне завдання № 14

Засвоєння методики визначення ідентифікації систем моделей

Мета роботи: Набути навичок з уміння ідентифікації системи моделей

Порядок виконання:

Варіанти

1. $\begin{cases} y_1 = 1 + 2y_2 + 1.5y_3 + 3x_1 \\ y_2 = 2 + 1.2y_3 + 2x_1 + 1.2x_2 \\ y_3 = 3 + 0.5y_1 + 1.3x_1 + 0.8x_3 \end{cases}$

2. $\begin{cases} y_1 = 1.5 + 2y_2 + 1.5x_1 + 3x_2 \\ y_2 = 2 + 1.2y_3 + 2x_1 + 2.2x_3 \\ y_3 = 3 + 0.5y_1 + 1.3x_1 + 0.7x_3 \end{cases}$

3. $\begin{cases} y_1 = 1.8 + 2.2y_2 + 3x_1 \\ y_2 = 2 + 1.2y_1 + 2x_1 + 1.2x_2 \\ y_3 = 2.2 + 0.5y_1 + 2.3x_1 + 0.8x_3 \end{cases}$

4. $\begin{cases} y_1 = 0.5 + 2y_3 + 1.5x_1 + 2x_2 \\ y_2 = 2 + 1.5x_1 + 2x_2 + 0.2x_3 \\ y_3 = 3.2 + 0.5y_2 + 1.3x_2 + 0.8x_3 \end{cases}$

5. $\begin{cases} y_1 = 2.5 + 2y_2 + 1.5y_3 + 3.5x_3 \\ y_2 = 4.2 + 1.2y_3 + 2x_1 + 1.2x_2 \\ y_3 = 3 + 0.5y_2 + 1.3x_1 + 1.8x_2 \end{cases}$

6. $\begin{cases} y_1 = 1 + 2y_2 + 1.5y_3 + 3x_1 \\ y_2 = 2.2 + 1.2y_3 + 2x_1 + 1.2x_2 \\ y_3 = 2.5 + 0.5y_1 + 1.5x_1 + 0.5x_3 \end{cases}$

7. $\begin{cases} y_1 = 5 + 1.5y_3 + 3x_1 \\ y_2 = 2 + 1.2y_1 + 2x_1 + 1.2x_2 \\ y_3 = 3.5 + 0.5y_1 + 1.3y_2 + 0.8x_3 \end{cases}$

8. $\begin{cases} y_1 = 1 + 2y_2 + 1.5x_1 + 3x_2 \\ y_2 = 2.6 + 2x_1 + 1.2x_2 \\ y_3 = 3 + 0.5y_1 + 1.3x_1 + 0.7x_3 \end{cases}$

9. $\begin{cases} y_1 = 1 + 2y_2 + 1.5y_3 + 3x_1 \\ y_2 = 2 + 1.5y_1 + 2.5x_1 + 1.2x_2 \\ y_3 = 3 + 0.5y_1 + 1.3x_1 + 2.8x_3 \end{cases}$

10. $\begin{cases} y_1 = 1 + 2y_2 + 1.5y_3 + 3x_1 \\ y_2 = 2.2 + 1.2y_3 + 0.5x_1 + 1.2x_2 \\ y_3 = 0.9 + 0.5y_1 + 1.3x_1 \end{cases}$

1. Ідентифікувати всі рівняння та систему в цілому.

Приклад виконання

$$\begin{cases} y_1 = 1 + 2y_2 + 1.5y_3 + 3x_1 \\ y_2 = 0.5 + 2.1y_3 + 1.8x_1 + 3.2x_2 \\ y_3 = 1 + 4y_1 + 1.2x_1 + 2x_2 \\ y_4 = 1.5 + 3y_1 + 1.5y_2 + 3.3x_3 \end{cases}$$

Таблиця 4.25. Коефіцієнти системи рівнянь

Рівняння	1	y_1	y_2	y_3	y_4	x_1	x_2	x_3
1	-1	1	-2	-1.5	0	-3	0	0
2	-0.5	0	1	-2.1	0	-1.8	-3.2	0
3	-1	-4	0	1	0	-1.2	-2	0
4	-1.5	-3	-1.5	0	1	0	0	-3.3

Таким чином, для даної задачі: $n = 4$ – загальна кількість ендогенних змінних у системі; $m = 3$ – загальна кількість екзогенних змінних у системі.

Необхідною умовою ідентифікації окремого рівняння моделі є рахункове правило. Якщо позначити через H число досліджуваних змінних $\boldsymbol{y_i}$, присутніх в і-му рівнянні, а через D позначити число факторних змінних x_j, відсутніх в і-му рівнянні, то рахункове правило має вигляд:

- якщо D + 1 < H, то рівняння неідентифіковані;
- якщо D + 1 = H, то рівняння ідентифіковані;
- якщо D + 1> H, то рівняння над ідентифіковані.

$$D_1 = 2, H_1 = 3 \rightarrow 2 + 1 = 3 .$$

$$D_2 = 1, H_2 = 2 \rightarrow 1 + 1 = 2 .$$

$$D_3 = 1, H_3 = 2 \rightarrow 1 + 1 = 2.$$

$$D_4 = 2, H_4 = 3 \rightarrow 2 + 1 = 3 .$$

Отже, всі рівняння мають необхідну умову ідентифікації.

Перевіримо достатню умову ідентифікації окремого рівняння моделі, яка виконується, якщо визначник матриці, складеної з коефіцієнтів в інших рівняннях при змінних (як досліджуваних y, так і факторних x), відсутніх в даному і-му рівнянні не дорівнює нулю, а ранг цієї матриці, водночас, не менше, ніж кількість всіх досліджуваних змінних в системі рівнянь за виключенням одного.

Розглянемо перше рівняння, в яке входять змінні y_1, y_2, y_3 та x_1. Для перевірки цього рівняння необхідно побудувати відповідну матрицю коефіцієнтів для змінних y_4, x_2, x_3, які включені в інші рівняння моделі, крім першого.

Матриця буде мати вигляд:

$$\begin{pmatrix} 0 & -3.2 & 0 \\ 0 & -2 & 0 \\ 1 & 0 & -3.3 \end{pmatrix}$$

Визначник цієї матриці дорівнює 0.

Оскільки визначник матриці дорівнює нулю, то ранг матриці A < 3. Тому перше рівняння системи не задовільнює умову рангу, тобто не є строгоідентифікованим.

Аналогічно дрозглянемо друге рівняння, в яке входять змінні y_2, y_3, x_1 та x_2. Для перевірки цього рівняння необхідно побудувати відповідну матрицю коефіцієнтів для змінних y_1, y_4, x_3, які включені в інші рівняння моделі, крім другого.

$$\begin{pmatrix} 1 & 0 & 0 \\ -4 & 0 & 0 \\ -3 & 1 & -3.3 \end{pmatrix}$$

Визначник цієї матриці теж дорівнює 0. Оскільки визначник матриці дорівнює нулю, то ранг матриці A < 3. Тому перше рівняння системи не задовільнює умову рангу, тобто не є строгоідентифікованим.

Аналогічно слід перевірити третє та четверте рівняння.

Контрольні запитання

1. Що називається множинною регресією?
2. В чому різниця між простою та множинною регресією?
3. Які дані називають спостереженням?
4. Що визначає зрушення?
5. Що визначаєють коефіцієнти регресії при кожній змінній?
6. Що показує стандартна помилка оцінки?
7. Що враховує коефіцієнт детермінації у разі множинної регресії?
8. З чого починається статистичний висновок?
9. На що вказують стандартні помилки коефіцієнтів?
10. Про що скаже p-значення, величина якого більше, ніж 0.05?
11. Що показують приватні коефіцієнти еластичності?

12. Коли виникає мультиколінеарність?

13. На що впливає мультиколінеарність змінних?

14. Яким чином проводиться класифікація переліку Х-змінних за пріоритетами?

15. В чому полягає алгоритм Феррара-Глобера?

16. Що показують коефіцієнти еластичності, бета–коефіцієнт, дельта – коефіцієнт?

17. Які нелінійні регресійні моделі можна привести до лінійного виду?

18. Що таке ідентифікація системи рівнянь?

У цьому розділі студенти навчилися визначати параметри лінійної множинної регресії та нелінійних регресій, виявляти вплив змінних на результат

Розділ 5.
КУРСОВА РОБОТА

У цьому розділі студенти ознайомляться з темами курсових робіть, принципами оформлення роботи, порядком захисту роботи.

Курсова робота виконується студентами спеціальності 051 Економіка. Ця робота є одним із елементів підготовки бакалавра, яка поглиблює знання студентів в ораному напрямку.

5.1. Мета і завдання курсової роботи

Курсова робота є одним з найбільш важливих елементів навчального процесу здобувачів вищої освіти. Курсова робота є самостійним науково-практичним дослідженням студента. Її метою є закріплення та систематизація отриманих теоретичних знань, а також набуття практичних навичок з дисципліни.

Метою даної курсової роботи є:

- побудова економетричних моделей, що кількісно описують взаємозв'язки між економічними змінними;

- отримання практичних навичок дослідження соціально-економічних процесів, що протікають в економічній системі;

- оволодінні навичками використання сучасних інформаційних технологій;

- формуванні здатності розв’язання складних задач у межах дисципліни;

- формуванні навичок публічного захисту результатів виконаного дослідження.

При виконанні курсової роботи необхідно використовувати матеріали лекційних та практичних занять, а також усіх доступних джерел інформації, включаючи самостійний їх пошук.

Завданнями курсової роботи є:

– дослідження та критичний аналіз теоретичних положень щодо теми курсової роботи;

– пошук вихідних даних для побудови моделі за темою дослідження у відкритих джерелах статистичної інформації;

– обробка даних, їх аналіз, оцінка впливу та взаємозв’язку та придатності до використанні в економетричному дослідженні;

– побудова моделі соціально-економічного процесу за темою дослідження, обґрунтування виду моделі;

– аналіз змодельованого процесу, отриманих результатів моделювання та їх інтерпретація.

Курсова робота складається з трьох розділів: теоретичного, аналітичного і практичного. Робота повинна містити самостійно виконані розрахунки за реальними даними, власні судження автора щодо проблем обраної теми, логічність матеріалу викладення та ілюстрації текстового матеріалу.

Для виконання курсової роботи студенту потрібно самостійно знайти необхідну інформацію в доступних джерелах, а також провести дослідження за обраною темою.

Результати виконання та захисту курсової роботи мають свідчити про ґрунтовні знання автором теми дослідження та дисципліни в цілому. Матеріал повинен відображати ставлення студента до тих точок зору, з якими він зустрівся при вивченні літератури (виклавши різні думки зі спірного питання, студент повинен указати, яку з названих точок зору він підтримує і чому, або висловити і мотивувати свою точку зору на розглянуту тему). Об’єктом дослідження курсової роботи є соціально-економічний процес відповідно до обраної тематики.

5.2. Тематика курсових робіт

Тематика курсових робіт охоплює коло проблем, що становить зміст дисципліни та відповідає програмі її вивчення. Для написання курсової роботи студент обирає одну із тем із наступного переліку, узгодивши її з викладачем:

1. Тестування адекватності моделі множинної регресії згідно загальної схеми.

2. Вплив темпів економічного зростання на демографічну ситуацію в країні (в регіоні).

3. Інвестиції як фактор економічного зростання національної (регіональної) економіки.

4. Аналіз впливу валютного курсу на темпи інфляції у національній економіці.

5. Дослідження попиту на товари індивідуального споживання.

6. Вплив рівня життя населення на демографічну ситуацію у країні (у регіоні).

7. Дослідження взаємозв’язку темпів економічного росту та рівня безробіття у країні (у регіоні).

8. Дослідження ринкової рівноваги на ринку товарів.

9. Дослідження зайнятості в країні (у регіоні).

10. Дослідження продуктивності праці на підприємстві.

11. Фактори, що впливають на купівельну спроможність населення у регіоні.

12. Аналіз факторів, що впливають на рівень заробітної плати на підприємстві.

13. Фактори, що впливають на кількість браків у країні (у регіоні).

14. Модель попиту та пропозиції на ринку товарів.

15. Модель попиту та пропозиції на ринку праці.

16. Фактори, що впливають на рівень зайнятості у країні (у регіоні).

17. Дослідження процесу виробництва на підприємстві на основі виробничої функції.

18. Дослідження процесу виробництва у галузі на основі виробничої функції.

19. Дослідження проблеми гетероскедастичності в економетричній моделі.

20. Дослідження проблеми автокореляції залишків в економетричній моделі.

21. Дослідження проблеми мультиколінеарності в економетричній моделі.

22. Аналіз факторів, що впливають на ринок нерухомості.

23. Економетричний аналіз в маркетингу.

24. Економетричний аналіз у рекламі.

25. Аналіз смертності у країні (у регіоні).

26. Економетричний аналіз бідності у країні (у регіоні).

27. Економетричні моделі аналізу кредитного портфеля.

28. Економетричні моделі оцінки фінансового стану підприємств.

29. Економетричні моделі прогнозування попиту на продукцію підприємства.

30. Економетрична оцінка прибутковості підприємств.

31. Економетричний аналіз туристичної привабливості регіонів України.

32. Економетричні моделі оцінки зайнятості населення.

33. Моделювання циклічності фондового ринку України.

34. Економетричні моделі оцінки вартості бізнесу.

35. Економетричні моделі оцінки рівня соціально-економічного розвитку регіонів України.

36. Економетричне моделювання динаміки галузевого фондового індексу.

37. Економетричні моделі оцінки ефективності корпоративних інвестицій.

38. Прогнозування фінансових показників діяльності підприємств.

39. Економетричні моделі оцінки інвестиційної привабливості підприємств.

40. Економетричні моделі оцінки кредитоспроможності позичальника.

41. Економетричні моделі аналізу трудових ресурсів України.

42. Модель прогнозування обсягу продажів продукції підприємства.

43. Економетричне моделювання в дослідженні економічної безпеки підприємства.

44. Економетричні моделі оцінки вартості нерухомості.

45. Економетричні моделі інфляційних процесів.

46. Економетричне моделювання рейтингу комерційних банків.

5.3. Порядок видачі завдання на курсову роботу

Тему курсової роботи обирає студент. Одна тема дослідження може бути обрана лише одним студентом групи.

Керівник курсової роботи за необхідності рекомендує студенту літературу, нормативні і довідкові матеріали, типові проекти та інші джерела за темою; проводить консультації з питань виконання розділів курсової роботи; перевіряє виконання курсової роботи; готує студента до захисту курсової роботи.

5.4. Зміст курсової роботи

Курсова робота складається з наступних елементів:

титульна сторінка (додаток);

зміст курсової роботи, який розташовують безпосередньо після титульної сторінки, починаючи з нової сторінки;

вступ, який відображає актуальність проблеми, її наукову і практичну цінність, об'єкт і предмет дослідження, основну мету і задачі дослідження. Вступ починають з нової сторінки після змісту. Обсяг – 1-2 сторінки.

основна частина роботи, яка складається з трьох розділів. Кожен розділ починають з нової сторінки. Розділи можуть поділятися на підрозділи. Кожен підрозділ повинен містити закінчену інформацію.

У першому розділі «Теоретичне та статистичне обґрунтування економетричної моделі» подається теоретичне та економічне обґрунтування економетричної моделі, що пропонується як інструмент економетричного дослідження. У розділі дається стислий опис об'єкта дослідження, звертається особлива увага на його властивості, та принципи моделювання даних процесів. Також у даному розділі необхідно представити формальну постановку задачі з описом усіх змінних та зв'язків між ними, а також обґрунтувати вибір типу економетричної моделі.

Приблизний обсяг розділу – 7-10 сторінок.

Другий розділ «Побудова та аналіз економетричної моделі» представляє результати побудови та аналізу однієї або декількох регресійних моделей, кількість яких залежить від завдання курсової роботи. Мінімальною вимогою до економетричної моделі є вимога наявності у моделі двох екзогенних змінних. Даний розділ поєднує наступні два етапи типового економетричного дослідження, а саме:

- Параметризація моделі (Оцінювання параметрів моделі).
- Верифікація моделі.

Приблизний обсяг розділу – 8-10 сторінок.

Третій розділ «Прогнозування та економіко-математичний аналіз» наводяться результати прогнозування та економіко-математичного аналізу економічного процесу, який виступав у якості об’єкту дослідження у курсовій роботі. Основну частину цього розділу складає обґрунтування результатів рішення моделі, побудова прогнозів, їх інтерпретація і напрямки використання.

В процесі побудови прогнозів студент самостійно визначає прогнозні значення пояснюючих змінних моделі. Бажаним також є попереднє оцінювання прогнозних якостей побудованої моделі.

Економіко-математичний аналіз зводиться до економічної інтерпретації параметрів оціненої моделі, у визначені граничного та відносного впливу пояснюючих змінних на залежну, а також при можливості – до оцінки силу впливу кожної пояснюючої змінної на залежну.

Приблизний обсяг розділу – 5-7 сторінок.

висновок, який розташовують безпосередньо після викладу основної частини роботи та містять стисле резюме отриманих результатів. Обсяг – 1-2 сторінки.

список використаних джерел, які були використанні при написанні курсової роботи, повинен бути приведений після висновку з наступної сторінки;

додатки, які містять матеріал, що є необхідним для повноти курсової роботи, але включення його в основну частину може змінити упорядковане і логічне уявлення про роботу; через великий обсяг не може бути послідовно розміщений в основній частині роботи.

При виконанні курсової роботи необхідно дотримуватись нормативно встановлених правил оформлення тексту, таблиць, формул, розрахунків, схем, рисунків, діаграм тощо.

Закінчену курсову роботу необхідно зброшурувати і здати на кафедру за два дні до початку екзаменаційної сесії.

5.5. Вимоги до оформлення курсової роботи

5.5.1. Загальні вимоги

Курсову роботу друкують з одного боку аркушів білого паперу формату А4 через 1,5 міжрядкові інтервали до сорока рядків на сторінці.

Текст роботи необхідно оформляти, залишаючи поля, мм: ліворуч – не менш як 30, праворуч – не менш як 10, угорі – не менш як 20, внизу – не менш як 20.

Шрифт друку має бути чіткий, чорного кольору середньої жирності, щільність тексту роботи – однакова.

Вписувати в текст роботи окремі іншомовні слова, формули, умовні знаки можна чорнилом, тушшю, пастою тільки чорного кольору, при цьому щільність вписаного тексту повинна бути наближеною до щільності основного.

Заголовки структурних частин курсової роботи: «ЗМІСТ», «ВСТУП», «РОЗДІЛ», «ВИСНОВКИ», «СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ», «ДОДАТКИ» друкують великими літерами симетрично до тексту. Заголовки підрозділів друкують маленькими літерами (крім першої великої) з абзацного відступу. Крапку в кінці заголовка не ставлять. Якщо заголовок складається з двох або більше речень, їх розділяють крапкою. Заголовки пунктів друкують маленькими літерами (крім першої великої) з абзацного відступу в розрядці в підбір до тексту. В кінці заголовка, надрукованого в підбір до тексту, ставиться крапка. Відстань між заголовком (за винятком заголовка пункту) та текстом має дорівнювати 3-4 інтервалам. Кожну структурну частину курсової роботи треба починати з нової сторінки.

5.5.2. Нумерація

Нумерацію сторінок, розділів, підрозділів, пунктів, підпунктів, рисунків, таблиць, формул подають арабськими літерами без знаку «№».

Першою сторінкою курсової роботи є титульна сторінка, яку враховують у загальній нумерації. На титульній сторінки номер сторінки не ставлять, сторінка змісту і вступу також не нумерується, на наступних – номер проставляють у правому верхньому куті сторінки без сторінки без крапки в кінці.

Номер розділу ставлять після слова «РОЗДІЛ», після номеру крапку не ставлять, потім з нового рядка друкують заголовок розділу.

Підрозділи нумерують у межах кожного розділу. Номер підрозділу складається з номеру розділу і порядкового номера підрозділу, між якими ставлять крапку. В кінці номера підрозділу має стояти крапка, наприклад, 2.3.

(третій підрозділ другого розділу). Потім у тому самому рядку друкують заголовок пункту. Пункт може не мати заголовка. Підпункти нумеруються у межах кожного пункту за тими самими правилами, що й пункти.

Ілюстрації (фотографії, креслення, схеми, графіки, карти) і таблиці необхідно подавати в роботи безпосередньо після тексту, де вони згадані вперше, або на наступній сторінці.

Ілюстрації позначають словом «Рис.» і нумерують послідовно в межах розділу, за винятком ілюстрацій, поданих у додатках. Номер ілюстрації має складатися з номера розділу і порядкового номера ілюстрації, між якими ставиться крапка. Наприклад, 1.2. (другий рисунок першого розділу). Номер ілюстрації, її назву і пояснювальні підписи розміщують послідовно під ілюстрацією. Якщо в роботі наведено одну ілюстрацію, то її нумерують за загальними правилами.

Таблиці нумерують послідовно (за винятком таблиць, поданих у додатках) в межах розділу. В правову верхньому куті над відповідним заголовком таблиці розміщують напис «Таблиця» із зазначенням її номера. Номер таблиці має складатися з номера розділу і порядкового номера таблиці, між якими ставиться крапка, наприклад «Таблиця 1.2» (друга таблиця першого розділу).

При переносі частини таблиці на іншу сторінку слово «Таблиці» і номер її вказують один раз праворуч над першою частиною таблиці, над перенесеними частинами пишуть слова «Продовження табл.» і вказують номер, наприклад, «Продовження табл. 1.2».

Формули в курсовій роботі (якщо їх більше однієї) нумерують у межах розділу. Номер формули складається з номера розділу і порядкового номера формули в розділі, між якими ставлять крапку. Номер формули пишуть праворуч на рівні самої формули в круглих дужках, наприклад, (3.1) (перша формула третього розділу).

Примітки до тексту і таблиць, в яких вказують довідкові та пояснювальні дані, нумерують послідовно в межах однієї сторінки. Якщо приміток на одному аркуші кілько, то після сова “Примітки” ставлять двокрапку, наприклад:

Примітки:

1…

2…

3…

Якщо примітка одна її не нумерують і після слова «Примітка» ставлять крапку.

5.5.3. Таблиці

Цифровий матеріал, як правило, має оформлюватися у вигляді таблиць. Кожна таблиця повинна мати назву, яку розміщують над таблицею і друкують симетрично до тексту. Слово «Таблиця» та її назву починають з великої літери. Назву не підкреслюють.

Заголовки граф треба починати з великих літер, підзаголовки – з маленьких, якщо вони складають одне речення із заголовком, і з великих, якщо вони є самостійними. Висота рядків має бути не менш як 8 мм.

Таблицю розміщують після першого згадування про неї в тексті так, щоб її можна було читати без повороту переплетеного блоку роботи або з поворотом за годинниковою стрілкою. Таблицю з великою кількістю граф можна ділити на частини і розміщувати одну частину під другою в межах однієї сторінки.

Приклад побудови таблиці:

Таблиця (номер)

Назва таблиці

5.5.4. Формули

Пояснення значень символів і числових коефіцієнтів треба подавати безпосередньо під формулою в тій послідовності, в якій вони наведені у формулі. Значення кожного символу і числового коефіцієнта треба подавати з нового рядка. Перший порядок пояснення починають зі слова «де».

Рівняння і формули треба виділяти в тексті вільними рядками. Вище і нижче кожної формули потрібно залишати не менше одного вільного рядка. Якщо рівняння не вміщується в один рядок, його слід перенеси після знака рівності (=) або після математичних знаків (+, -, ×).

5.5.5. Посилання

У процесі написання курсової роботи необхідно давати посилання на джерела, матеріали або окремі результати, з яких наводяться в роботі, або на ідеях і висновках яких розробляються проблеми, завдання, питання, дослідження. Такі посилання дають змогу відшукати документи і перевірити достовірність відомостей про цитування документа, дають необхідну інформацію щодо нього, допомагають з’ясувати зміст, мову тексту, обсяг.

Посилатися в тексті курсової роботи на джерела слід зазначити порядковим номером за переліком посилань, виділеним двома квадратними дужками, наприклад, «…у працях [1-7]…».

5.5.6. Список використаних джерел

Джерела можна розміщувати в списку одним із таких способів:

- у порядку появи посилань у тексті;
- в алфавітному порядку прізвищ перших авторів або заголовків;
- у хронологічному порядку.

Відомості про джерела, занесені до списку, необхідно давати згідно з вимогами державного стандарту з обов’язковим наведенням назв праць.

5.5.7. Додатки

Додатки оформляють як продовження курсової роботи на наступних його сторінках або у вигляді окремої частини (книги), розміщуючи їх у порядку появи посилань у тексті роботи.

Додатки слід позначити послідовно великими літерами української абетки, за винятком літер Г, Є, І, Ї, Й, О, Ч.

Текст кожного додатка в разі потреби може бути поділений на розділи і підрозділи, які нумеруються у межах кожного додатка, наприклад, А.2.

5.6. Порядок захисту курсової роботи

Студенти захищають курсову роботу публічно.

До захисту курсових робіт допускаються студенти, які виконали всі вимоги. Процедура захисту передбачає стисле викладення студентом результатів дослідження і відповідей на запитання. В ході захисту оцінюються не тільки виконання роботи, але й якість самого захисту.

Якщо захист курсової роботи оцінено незадовільно, то студент може повторно захищати ту саму роботу з доопрацюваннями, визначеними комісією або ж зобов’язаний розробити нову тему, яку визначає керівник.

Виконавши курсову роботу за наведеним вище порядком, студент поглибить свої знання в напрямку освіти за спеціальність 051 Економіка.

ПІДСУМКИ

Матеріал цього посібника розкриває всю сутність економетрики – це не тільки розрахунок коефіцієнтів лінійної моделі.

Повний комплекс заходів включає в себе статистичній збір даних, розділення їх на екзогенні та ендогенні, визначення їх статистичних характеристик, розрахунок довірчих інтервалів цих характеристик, уникнення мультиколінеарності, вибір виду функції апроксимації, розрахунок коефіцієнтів моделі, визначення статистичної достовірності цих коефіцієнтів, розрахунок якості апроксимації, розрахунок впливу кожного екзогенного фактору на ендогенний, визначення якості прогнозування.

Такий перелік вимагає від дослідника, а всі сучасні економісти є дослідниками, вільного володіння розрахунковими можливостями програми Excel, прикладам розрахунку на якій у посібнику приділена значна частка обсягу викладення матеріалу.

Виконання індивідуальних завдань поглиблює знання студентів не тільки в області економетричних розрахунків, воно заставляє глибше зрозуміти основи теорії ймовірності та математичної статистики.

Створення моделей не самоціллю економетрики, розроблені за всіма правилами, статистично достовірні моделі готові для подальшого використання при оптимізаційних розрахунках. Наприклад таких, як оптимальний план випуску продукції, оптимальні плани перевезень, оптимальні ціни, що враховують інтереси виробників і споживачів, тощо.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Білоцерківський О.Б., Ширяєва Н.В. Економетрія. Навчально-методичний посібник. — Харків: НТУ "ХПІ", 2008. — 80 с.
2. Бобровнича Н.С., Борисевич Є.Г. Економетрія. Навчальний посібник. — Одеса: ОНАЗ ім. О.С. Попова, 2010. — 180 с.
3. Доля В.Т. Економетрія. Навч. посібник. — Х.: ХНАМГ, 2010. — 171 с.
4. Економетрика в електронних таблицях: навч. посіб./Васильєва Н.К., Мироненко О.А., Самарець Н.М., Чорна Н.О.; за заг. Ред. Васильєвої Н.К. Дніпро, Біла К.О, 2017. 149 с.
5. Економетрика: метод. рекомендації до виконання курсової роботи для студентів спеціальності 051 «Економіка» / уклад. Т.Ю. Яковенко. Дніпро, НТУ «ДП», 2019. – 12 с.
6. Економетрика: метод. рекомендації до виконання лабораторних робіт для студентів спеціальності 051 «Економіка» / уклад. І.М.Пістунов. – Дніпро: НТУ «ДП», 2021. – 16 с.
7. Пістунов І.М. Економетрика: навч. наоч. посіб. Дніпро: НТУ «ДП», 2024. 36 с.
8. Лещинський О.Л., Рязанцева В.В., Юнькова О.О., Юртин І.І. Практикум з економетрії. К. : Персонал, 2009. — 256 с.
9. Пістунов І.М. Економічна кібернетика: Навч. посібник. Видання 2-ге, виправлене і доповн. ДНІПРО, НГУ, 2014. 215 с.

10. Пістунов І.М., Турчанінова І.Ю. Теорія ймовірності та математична статистика для економістів. З елементами електронних таблиць: Навч. Посібн. Дніпро, НТУ «ДП», 2023. 171 с.
11. Прикладна економетрика : навч. посіб. : у двох частинах. Частина 1 : [Електронне видання] / Л. С. Гур'янова, Т. С. Клебанова, С. В. Прокопович та ін. – Харків : ХНЕУ ім. С. Кузнеця, 2016. – 235 с

ДОДАТКИ

СЛОВНИК СПЕЦІАЛЬНИХ ТЕРМІНІВ

Абстрактні системи – системи з об'єктів, що існують лише в думці людини: поняття, ідеї, гіпотези, плани і тому подібне.

Автокореляцією залишків – вплив фактора часу, який виражений у кореляційній залежності між значеннями залишків e_i за поточний і попередній моменти часу.

Адаптивні системи – системи, які реагують зміни зовнішнього середовища та здатні пристосовуватися до них.

Алгебраїчна система (алгебраїчна структура) – в математиці це не порожня множина з заданим на ній набором операцій та відношень, що задовольняють деякій системі аксіом.

Асиметрія визначає зміщення ознаки в сукупності відносно її середньої величини. Додатна асиметрія – це зрушення розподілу у бік позитивних відхилень, від’ємна, у бік негативних.

Вибірка – це підмножина об'єктів або одиниць, взятих з певної популяції, що досліджується в рамках статистичного дослідження.

Вибіркове середнє або **середнє арифметичне** – це є один з основних показників центральної тенденції в статистиці. Воно використовується для опису типового значення вибірки або підмножини даних. Вибіркове середнє обчислюється шляхом додавання всіх значень у вибірці і поділення на кількість спостережень.

Відкриті (незамкнуті) системи – ті системи, які з певною регулярністю і певним чином обмінюються з навколишнім середовищем речовиною,

енергією або інформацією. Вони називаються відкритими (незамкнутими) по якомусь з цих ознак (відкриті для енергії, відкриті для інформації, для речовини).

Генеральна сукупність – це сукупність об’єктів, відносно яких проводиться дослідження та з яких можна зробити вибірку.

Гетероскедастичність – це статистичне поняття, яке відноситься до нерівномірності дисперсії помилок в регресійній моделі. В ідеальному випадку, в регресійній моделі дисперсія помилок (різниця між спостережуваними значеннями та прогнозованими значеннями) повинна бути постійною (гомоскедастичною). Однак, в реальних даних може спостерігатися гетероскедастичність, коли дисперсія помилок змінюється залежно від значень змінних.

Детермінована (визначена) система – система, складові якої, об'єкти або частини, і зв'язки між ними функціонують таким чином, який точно передбачений.

Динамічні ситеми – системи, в яких показники, що характеризують стан елементів, змінні в часі, а зв'язки між елементами – динамічні.

Дискретна модель – математична чи імітаційна модель, змінні якої приймають тільки дискретні значення, тобто змінюються від одного значення до іншого і не приймають проміжних значень (наприклад, модель, що прогнозує рівні запасів організації, ґрунтуючись на відвантаженнях, які змінюються, і платежах). Протилежність: неперервна модель.

Дисперсія – показник, що характеризує розсіювання значень ознаки щодо його середньої величини $D_X = \frac{1}{N-1}\sum_{i=1}^{N} X_i^2 - M_X^2$, де X_i – окреме значення ознаки; M_X – середня арифметична ознаки; N – число значень ознаки.

Екзогенні змінні – це змінні, які включені в економетричну модель і не залежать від інших змінних в цій моделі. Вони є зовнішніми

факторами або умовами, які впливають на ендогенні змінні, але самі не залежать від них.

Економетрика – це не те ж саме, що економічна статистика. Вона не ідентична і тому, що ми називаємо економічної теорією, хоча значна частина цієї теорії носить кількісний характер. Економетрика не є синонімом додатків математики до економіки. Як показує досвід, кожна з трьох відправних точок – статистика, економічна теорія та математика – необхідна, але не достатня умова для розуміння кількісних співвідношень в сучасному економічному житті. Це єдність всіх трьох складових. І це єдність утворює економетрику».

Економіка – наука, що вивчає поведінку і взаємодію людей, фірм, урядів та інших економічних агентів в процесі виробництва, розподілу та споживання ресурсів з метою задоволення їхніх потреб та досягнення економічних цілей.

Ексцес (excess) - це статистичний показник, який відноситься до моментів розподілу випадкової величини і використовується для характеристики форми розподілу даних. Він вимірює ступінь відхилення від нормального розподілу. Ексцес вказує на хвостатість розподілу та наявність важких хвостів (тобто довших або важчих хвостів, ніж у нормального розподілу). Він визначається відносно нормального розподілу, де нормальний розподіл має ексцес рівний нулю.

Емерджентність – тобто наявність таких специфічних властивостей системи, які не випливають з властивостей, притаманних її елементам, а виникають у процесі їхньої взаємодії як наслідок відповідних кооперативних ефектів.

Ендогенні змінні – це змінні, які включені в економетричну модель і є об'єктом дослідження або пояснюються іншими змінними в цій моделі. Вони є результатом внутрішніх процесів або взаємодій у системі, що вивчається. Ендогенні змінні можуть бути як залежними

змінними, які потрібно пояснити, так і пояснюють змінні, які впливають на них.

Закриті (замкнуті) системи – системи, в яких обмін речовиною, енергією або інформацією з навколишнім середовищем не спостерігається або ж він незначний і їм нехтують.

Зв'язок – або кореляція, статистичний термін, який використовується для вимірювання ступеня залежності між двома змінними. У контексті статистики та економетрії, зв'язок є важливим поняттям, яке допомагає зрозуміти, як зміна однієї змінної впливає на іншу.

Змінні - показники, які змінюються в часі істотно і ці зміни враховуються в дослідженнях.

Коефіцієнт детерміації (часто також використовують термін «квадрат множинної кореляції»), R^2, показує, який відсоток варіації Y пояснюється впливом всіх X- змінних.

Коефіцієнт кореляції – параметр, який характеризує ступінь лінійного взаємозв'язку між двома вибірками, розраховується за формулою $r_{xy} = \frac{\sum(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\cdot\sum(y_i-\bar{y})^2}} = \frac{cov(x,y)}{\sigma_x\sigma_y}$, де σ_x, σ_y – середнє квадратичне відхилення вибірки x та y відповідно.

Кореляційна матриця - таблицею, яка містить коефіцієнти кореляції для кожної пари змінних з вашої багатовимірної сукупності даних.

Кореляція, або точніше **лінійний коефіцієнт кореляції (r)**, є безрозмірне (що не має одиниць вимірювання) число в діапазоні від **-1** до **1**, яке характеризує силу взаємозв'язку.

Людино-машинні системи – сучасного виробництва і для інших областей нашого життя характерна наявність сукупності машин, що виконують певні складні функції за участю людини, причому основне навантаження лягає не на його м'язи, а на психічні процеси сприйняття, запам'ятовування, мислення.

Макетна модель – це реально існуюча модель, що відтворює модельовану систему у деякому масштабі.

Математична модель – система математичних співвідношень, які описують досліджуваний процес або явище.

Математичне моделювання – метод дослідження процесів або явищ шляхом створення їхніх математичних моделей, дослідження цих моделей.

Математичний стандарт – дивись середнє квадратичне відхилення.

Машинні системи – системи, що складаються з одних тільки машин і механізмів і що діють автоматично, без участі людей.

Медіáна (англ. median) – в статистиці це величина ознаки, що розташована посередині ранжованого ряду вибірки, тобто — це величина, що розташована в середині ряду величин, розташованих у зростальному або спадному порядку.

Метод найменших квадратів полягає у відшуканні параметрів моделі тренда, яка краще всього описує тенденцію розвитку якого-небудь випадкового явища в часі або в просторі

Мóда – значення випадкової величини, що трапляється найчастіше в сукупності спостережень.

Модель – речова, знакова або уявна (мислена) система, що відтворює, імітує, відображає принципи внутрішньої організації або функціонування, певні властивості, ознаки та/або/і характеристики об’єкта дослідження (оригіналу). Розрізняють фізичні, математичні та ін. моделі. Слово «модель» походить від латинського modulus, що означає міра, такт, ритм, величина.

Мультиколінеарність – це статистичний термін, що описує високу кореляцію між двома або більше залежними змінними у регресійній моделі. Це означає, що одна або кілька змінних можуть бути добре передбачені або пояснені за допомогою інших змінних у моделі.

Неадаптивні системи – системи, які пасивні до змін зовнішнього середовища.

Нестабільна ситема – система з властивостями, що змінюються, і функціями та без циклів, що строго повторюються, в цих змінах.

Об'єктом прийнято називати деяке явище, підприємство, механізм, технологічний процес, які є предметом вивчення дослідника.

Параметр – у контексті статистики та економетрії, параметр – це характеристика популяції, яка використовується для опису та моделювання статистичних даних.

Помилкова кореляція - високу кореляцію, яка виникає завдяки дії деякого третього чинника.

Постійні системи - системи, що існують тривалий час в порівнянні з обмеженим періодом діяльності людей в них.

Проста регресія – прогнозування єдиної змінної Y на підставі єдиної змінної X.

Прості системи – системи, що мають просту структуру і виконують якісь нескладні функції.

Результативними ознаки – а ознаки, які характеризують наслідки зв'язку, *у*.

Розмах (англ. range) – в статистиці різниця між найбільшим та найменшим із сукупності числових значень.

Середнє (вибіркове середнє) – ви

Середнє квадратичне відхилення (або математичний стандарт чи просто стандарт) – це узагальнююча характеристика абсолютних розмірів варіації ознаки в сукупності. Середнє квадратичне відхилення є мірилом надійності середньої. Чим менше середнє квадратичне відхилення, тим краще середня арифметична відбиває собою всю вибірку. Середнє квадратичне відхилення – це квадратний корінь з дисперсії. $\sigma_X = \sqrt{D_X}$, де D_X – дисперсія ознаки.

Системою є сукупність об'єктів і процесів, званих компонентами або елементами, взаємозв'язаних і таких, що взаємодіють між собою, які

утворюють єдине ціле, таке, що володіє властивостями, не властивими складовим його компонентам, узятим окремо.

Складні системи – системи, що складаються з великої кількості взаємозв'язаних і взаємодіючих між собою частин (мають складну структуру) і виконують якусь достатньо складну функцію.

Соціальні системи – системи, що складаються з людей. Природно, в такі системи входять і різні об'єкти, з яких складаються фізичні неживі системи – машини, будівлі, споруди.

Соціально-економічна система - це цілісна сукупність взаємозв'язаних і взаємодіючих соціальних і економічних інститутів (суб'єктів) і відносин з приводу розподілу і споживання матеріальних і нематеріальних ресурсів, виробництва, розподілу, обміну і споживання товарів і послуг.

Стабільна система – система, в якій її властивості і функції протягом тривалого часу істотно не змінюються або змінюються у формі циклів, що повторюються.

Статистика – це наука, яка вивчає методи збору, аналізу, інтерпретації, представлення та організації даних, які виникають у великому обсязі і є випадковими. Статистика знаходить широке застосування в різних галузях, включаючи науку, бізнес, медицину, економіку, соціологію, психологію, інженерію та інші.

Статичні системи - системи, в яких показники, що характеризують стан елементів, постійні в часі (параметри), а зв'язки між елементами жорсткі.

Стохастичні системи - системи, в яких складові їх об'єкти (частини) і зв'язки між ними функціонують так, що не можна точно затверджувати про послідовність їх станів, детально передбачати їх поведінку. Також їх називають випадковими, імовірнісними або такими, що нерегулярно функціонують.

Тимчасові системи - системи, які створюються на заданий період часу, а потім ліквідовуються.

Тренд - це лінія, яка і характеризує тенденцію розвитку.

Фактор – у контексті економетрії, фактор (англ. factor) - це змінна або змінні, які впливають на залежну змінну в регресійній моделі. Фактори також називають незалежними змінними, предикторами або змінними впливу. Вони використовуються для пояснення та передбачення залежної змінної в рамках економетричного дослідження.

Факторні ознаки – ознаки, що характеризують причини і умови зв'язку, *x*.

Фізичні системи – системи, які складаються з реально існуючих (природних або штучних) об'єктів: машин, виробів, устаткування, працівників і так далі

Часовий ряд – це сукупність значень будь-якого показника за декілька послідовних моментів (періодів) часу.

Додаток Б

ТАБЛИЦЯ ЗНАЧЕНЬ ФУНКЦІЙ, НЕОБХІДНИХ ДЛЯ РОЗРАХУНКІВ

Фрагмент таблиці значень функції $\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

для квантиля $z = \frac{\beta - m}{\sigma}$

x	Φ(x)		x	Φ(x)		x	Φ(x)
0,00	0,0000		0,95	0,8209		1,90	0,9928
0,05	0,0564		1,00	0,8427		1,95	0,9942
0,10	0,1125		1,05	0,8624		2,00	0,9953
0,15	0,1680		1,10	0,8802		2,05	0,9936
0,20	0,2227		1,15	0,8961		2,10	0,9970
0,25	0,2763		1,20	0,9103		2,15	0,9976
0,30	0,3286		1,25	0,9229		2,20	0,9981
0,35	0,3794		1,30	0,9340		2,25	0,9985
0,40	0,4284		1,35	0,9438		2,30	0,9988
0,45	0,4755		1,40	0,9523		2,35	0,9991
0,50	0,5205		1,45	0,9597		2,40	0,9993
0,55	0,5633		1,50	0,9661		2,45	0,9995
0,60	0,6069		1,55	0,9716		2,50	0,9996
0,65	0,6420		1,60	0,9736		2,55	0,9997
0,70	0,6778		1,65	0,9804		2,60	0,9998
0,75	0,7112		1,70	0,9838		2,65	0,9998
0,80	0,7421		1,75	0,9867		2,70	0,9999
0,85	0,7707		1,80	0,9891		2,75	0,9999
0,90	0,7969		1,85	0,9911		2,80	0,9999
0,95	0,8209		1,90	0,9928		3,00	1,0000

Додаток В

Фрагмент таблиці значень $\chi 2(r,p)$

r p	1	2	3	4	5	6	7	8	9	10	11	12	13
0,99	0	0,02	0,115	0,297	0,55	0,87	1,24	1,65	2,09	2,56	3,053	3,571	4,107
0,98	0	0,04	0,185	0,429	0,75	1,13	1,56	2,03	2,53	3,06	3,609	4,178	4,765
0,97	0,001	0,061	0,245	0,535	0,9	1,33	1,8	2,31	2,85	3,41	3,997	4,601	5,221
0,96	0,003	0,082	0,3	0,627	1,03	1,49	2	2,54	3,1	3,7	4,309	4,939	5,584
0,95	0,004	0,103	0,352	0,711	1,15	1,64	2,17	2,73	3,33	3,94	4,575	5,226	5,892
0,94	0,006	0,124	0,401	0,788	1,25	1,76	2,32	2,91	3,52	4,16	4,81	5,48	6,163
0,93	0,008	0,145	0,449	0,862	1,35	1,88	2,46	3,07	3,7	4,35	5,024	5,71	6,409
0,92	0,01	0,167	0,495	0,931	1,44	2	2,59	3,22	3,87	4,54	5,221	5,921	6,634
0,91	0,013	0,189	0,54	0,999	1,53	2,1	2,72	3,36	4,02	4,7	5,405	6,118	6,844
0,9	0,016	0,211	0,584	1,064	1,61	2,2	2,83	3,49	4,17	4,87	5,578	6,304	7,041
0,89	0,019	0,233	0,628	1,127	1,69	2,3	2,95	3,62	4,31	5,02	5,742	6,48	7,228
0,88	0,023	0,256	0,671	1,188	1,77	2,4	3,05	3,74	4,44	5,16	5,899	6,648	7,407
0,87	0,027	0,279	0,714	1,249	1,85	2,49	3,16	3,85	4,57	5,3	6,05	6,809	7,577
0,86	0,031	0,302	0,756	1,308	1,92	2,57	3,26	3,97	4,7	5,44	6,196	6,964	7,742
0,85	0,036	0,325	0,798	1,366	1,99	2,66	3,36	4,08	4,82	5,57	6,336	7,114	7,901
0,84	0,041	0,349	0,839	1,424	2,07	2,75	3,45	4,19	4,93	5,7	6,473	7,259	8,055
0,83	0,046	0,373	0,881	1,481	2,14	2,83	3,55	4,29	5,05	5,82	6,606	7,401	8,205
0,82	0,052	0,397	0,922	1,537	2,21	2,91	3,64	4,39	5,16	5,94	6,737	7,54	8,351
0,81	0,058	0,421	0,964	1,593	2,27	2,99	3,73	4,49	5,27	6,06	6,864	7,675	8,494
0,8	0,064	0,446	1,005	1,649	2,34	3,07	3,82	4,59	5,38	6,18	6,989	7,807	8,634

Навчальне видання

Пістунов Ігор Миколайович
Приходченко Оксана Юріївна

ЕКОНОМЕТРИКА.
З РОЗРАХУНКАМИ НА EXCEL

Навчальний посібник

Електронне видання

У редакції авторів

Підписано до друку 18.01.2024. Формат 30 х 42/4.
Папір офсетний. Ризографія. Умовн. друк. арк. 11,77.
Обліково-видавн. арк. 12,03. Тираж 150 пр. Зам. № 96/12

Підготовлено у НТУ «Дніпровська політехніка».
Свідоцтво про внесення до державного реєстру ДК №1842.
49005, м. Дніпропетровськ, просп. Д. Яворницького, 19.

ВІДОМОСТІ ПРО АВТОРІВ

ПІСТУНОВ ІГОР МИКОЛАЙОВИЧ – доктор технічних наук, професор кафедри економіки та економічної кібернетики НТУ «ДП». Автор 9 монографій, 88 статей, 62 тез доповідей та 65 навчальних матеріалів з економіки. Випустив 2 кандидатів і консультував трьох кандидатів та двох докторів економічних наук.

ПРИХОДЧЕКНО ОКСАНА ЮРІЇВНА – кандидат економічних наук, доцент кафедри економіки та економічної кібернетики НТУ «ДП»