Міністерство освіти і науки України
Вінницький національний технічний університет

Е К О Н О М Е Т Р І Я

Навчальний посібник

Вінниця
ВНТУ
2014

УДК 330.43(075)
ББК 65в6я73
А35

Автори:
А. О. Азарова, Н. В. Сачанюк-Кавецька, О. М. Роїк, Ю. В. Міронова

Рекомендовано Міністерством освіти і науки України як навчальний посібник для студентів вищих навчальних закладів, які навчаються за освітньо-професійною програмою бакалавра з галузі знань "Економіка та підприємництво" та "Менеджмент". Лист № 1/11-7160 від 17.04.2013 р.

Рецензенти:
І. Г. Лук'яненко, доктор економічних наук, професор
В. В. Вітлінський, доктор економічних наук, професор
О. М. Ляшенко, доктор економічних наук, професор

А35 **Економетрія** : навчальний посібник / А. О. Азарова, Н. В. Сачанюк-Кавецька, О. М. Роїк, Ю. В. Міронова. – Вінниця : ВНТУ, 2014. – 304 с.
ISBN

У посібнику розглянуто фундаментальні засади економетричного моделювання, що посідає чільне місце у системі підготовки економістів нового покоління. Запропоновані авторами теоретичні засади та практичні економетричні аспекти дозволять студентам моделювати різноманітні аспекти господарської діяльності, засновуючись на сучасних методах системного та економетричного аналізу, а також комп'ютерних технологіях. Автори посібника показали перспективи застосування економетричного моделювання, що уможливлює отримання якісно нових результатів управління.

Посібник розроблено згідно з навчальною програмою МОНУ з дисципліни "Економетрія".

УДК 330.43(075)
ББК 65в6я73

ISBN

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ЕМ – економіко-математична модель
ЕММ – економіко-математичні методи
МНК – метод найменших квадратів
ВВ – випадкова величина
РВВ – розподіл випадкової величини

ВСТУП

Тенденції розвитку економічної науки за умов безперервно зростаючої складності соціально-економічних процесів, удосконалення форм господарювання, ускладнення управлінських функцій на всіх рівнях ієрархії економічних систем зумовили необхідність широкого застосування в економічних дослідженнях та у практиці управління економікою математичних методів. Застосування математичного моделювання та комп’ютерних технологій в економіці дає змогу якісно й кількісно оновлювати засоби теоретичного аналізу та методи прийняття рішень в усіх галузях економіки, забезпечує можливість підтверджувати або відкидати теоретичні та абстрактно-розрахункові припущення в нових напрямах економічної теорії, прогнозувати та регулювати економічні процеси, не здійснюючи реальних експериментів, які можуть призвести до небажаних соціально-економічних наслідків.

Останнє десятиліття економетрія як навчальна дисципліна стрімко розвивається. Підтвердженням всесвітнього визнання економетрії є присудження за найбільш видатні розробки у цій галузі Нобелівських премій в економіці, авторами яких є:

– Ян Тінберген (1903-1994) – нідерландський економіст та Рангар Фріш (1885-1973) – норвезький економіст – лаурети Нобелівської премії 1969 року за створення та застосування динамічних моделей для аналізу економічних процесів;

– Пол А. Самуельсон (1915-2009) – американський економіст – лауреат Нобелівської премії 1970 року за розроблення статистичної та динамічної економічної теорії, що сприяло зростанню рівня аналізу економічної науки;

– Василь Леонтьєв (1912-1986) – американський економіст – лауреат Нобелівської премії 1973 року за розроблення балансових моделей для моделювання взаємозв’язків великої кількості змінних (метод “витрати – випуск”);

– Леонід Канторович (1912-1986) – російський економіст, Тьяллінг С. Кумпанс (1910-1985) – американський економіст – лауреати Нобелівської премії 1975 року за вклад у теорію оптимального розподілу ресурсів. Т. С. Кумпанс також зробив значний внесок у розвиток статистичних методів в економетрії та створення лінійних економетричних моделей;

– Лоуренс Р. Клейн (нар. 1920) – американський економіст – лауреат Нобелівської премії 1980 року за створення економетричних моделей та їх застосування під час аналізу економічних коливань й економічної політики;

– Трюгве Хаавельмо (1911-1999) – норвезький економіст – лауреат Нобелівської премії 1989 року за перетворення основ теорії ймовірності в межах економетрії та аналіз залежних економічних структур;

– Джеймс Дж. Хекман (нар. 1944) – американський економіст; Даніел Д. Макфадден (нар. 1944), американський економіст, – лауреати Нобелівської премії 2000 року за розроблення мікроеконометрії та методів статистичного аналізу;

– Леонід Гурвіц (1917-2008) – американський економіст, Ерік Маскін (нар. 1950) – американський економіст, Роджер Маерсон (нар. 1951) – американський економіст – лауреати Нобелівської премії 2007 року за створення основ теорії оптимальних механізмів (механізмів поділу).

Крім того, процес прийняття науково обґрунтованих рішень в економіці щільно пов'язаний із визначенням кількісних співвідношень між економічними показниками. Так, наприклад, для з'ясування доцільності інвестування придбання нового обладнання (або розроблення нової технології) потрібно знати, який додатковий дохід можна отримати на кожну одиницю капітальних вкладень у разі реалізації різних проектів інвестування.

Мова економіки все більше стає мовою математики, а саму економіку все частіше називають однією з найбільш математизованих наук. Завдяки математичній формалізації в економічній теорії здійснюється пошук змістовних економічних аналогів абстрактних математичних величин та відношень, які фігурують у математичних моделях.

Досягнення сучасної економічної науки висувають новітні вимоги до професійної освіти економістів та менеджерів. Якщо в період планової економіки акцент робився на балансові та оптимізаційні методи дослідження, то в період пізньоіндустріальної економіки суттєво зростає роль економетричних методів. Математика настільки глибоко ввійшла в економіку, що інколи складно виокремити економічні знання в математичних. Тому сьогодні коректніше вести мову не про використання математики в економіці, а про взаємодію економічної та математичної наук, яка піднімає економічну теорію на якісно новий рівень.

Досить часто розглядають економетрію як науку, що встановлює та досліджує кількісні закономірності та взаємозв'язки в економічному житті за допомогою математичних і статистичних методів.

Метою економетрії є емпіричне виведення економічних законів.

Основні результати економічної теорії носять якісний характер, а економетрія вносить в них емпіричний вміст. Математична економіка зображує економічні закони у вигляді математичних співвідношень, а економетрія проводить експериментальну перевірку цих законів. Економічна статистика дає інформаційне забезпечення досліджуваного процесу у вигляді вихідних (оброблених) статистичних даних та економічних показників, а економетрія, використовуючи традиційні математико-статистичні та спеціально розроблені методи, проводить аналіз кількісних взаємозв'язків між цими показниками.

Економетрія – економіко-математична наука, яка поєднує економічну теорію, математику та статистику, забезпечує ефективний взаємозв'язок теоретичного та прикладного в економічній науці.

На основі положень економічної теорії та статистичних даних про соціально-економічні системи економетрія за допомогою економіко-математичних моделей досліджує закономірності розвитку цих систем з метою прогнозування, аналізу їх взаємного впливу та прийняття оптимальних рішень щодо управління ними.

Економетрія посідає чільне місце у системі підготовки економістів нового покоління, спільно з іншими математичними та економічними дисциплінами формує нове економічне мислення майбутніх фахівців.

Студенти, які здобудуть навички моделювання різноманітних економічних ситуацій, стануть носіями нової ідеології економічного мислення, що ґрунтується на глибоких знаннях сучасних методів системного та економетричного аналізу і комп'ютерних технологій, які застосовуються для дослідження реальних економічних об'єктів, процесів, явищ.

Автори посібника намагалися показати, що економетричне моделювання суттєво розширює можливості економічного аналізу, дає змогу отримати якісно нові результати дослідження соціально-економічних систем.

ТЕМА 1 ОСНОВНІ АСПЕКТИ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ

1.1 Вступ до економетричного моделювання

Розглянемо таку ситуацію. Припустимо, ми хочемо продати автомобіль. Перед нами постає запитання: яку суму ми можемо очікувати за наше авто? Зрозуміло, що ми будемо керуватися інформацією про ціни на подібні авто. Що означає “подібні автомобілі”? Зрозуміло, що це авто, які мають достатньо близькі до нашого автомобіля показники (марка, рік випуску, потужність двигуна, кілометраж та ін.).

Відвідавши автосалони, автомобільні ринки, переглянувши газети з оголошеннями, ми формуємо своє уявлення про можливу ціну.

На цьому елементарному прикладі можна відслідкувати основні моменти економетричного моделювання. Розглянемо наші дії більш формалізовано.

Ми ставимо задачу визначення ціни – величини, яка формується під впливом деяких факторів (марка, рік випуску, потужність двигуна, кілометраж та ін.). Такі залежні величини називають *пояснюваними, залежними (ендогенними)* змінними, а фактори, від яких вони залежать – *пояснювальними (екзогенними)*. Формуючи загальне уявлення про стан ринку, ми отримуємо *очікуване* значення залежної змінної при заданих значеннях екзогенних змінних.

На вказану конкретну ціну, *спостережуване* значення залежної змінної, впливають також *випадкові* явища – характер продавця, можливі терміни продажу автомобіля та ін.

Зрозуміло, що менеджер великого салону, де спеціалізуються на торгівлі автомобілями на вторинному ринку, скоріш за все, захоче мати більш точне уявлення щодо очікуваної ціни та можливої поведінки випадкової складової. Наступним кроком і буде економетричне моделювання.

Спільним моментом для довільної економетричної моделі є розбиття залежної змінної на дві частини – *пояснювальну* та *випадкову*. Очевидно, що задачею моделювання є *на основі експериментальних даних визначити пояснювану частину та, розглядаючи випадкову складову як випадкову величину, отримати оцінки параметрів її розподілу.*

Таким чином, економетрична модель цього процесу є такою:

Спостережува не значення залежної змінної	Пояснювана частина, яка залежить від значень екзогенних змінних	+	Випадкова складова	
Y	$f(X)$	+	ε.	(1.1)

Припустимо, що в нашому випадку отримано такий вираз для визначення ціни:

$$y = 35000 - 1500x_1 - 0{,}3x_2,$$

де y – очікувана ціна автомобіля (ум. гр. од.);

x_1 – термін експлуатації автомобіля (в роках);

x_2 – пробіг (тис. км).

Яке ж практичне значення одержаного результату? Зрозуміло, що дана формула дає уявлення про формування ціни на автомобіль. З іншого боку, дана залежність дозволяє з’ясувати вплив кожної пояснюваної змінної на ціну. Зокрема, ціна нового ($x_1 = x_2 = 0$) авто 35000 ум. гр. од., при цьому тільки за рахунок збільшення терміну експлуатації на один рік ціна авто зменшується в середньому на 1500 ум. гр. од. Слід відмітити, що даний результат дозволяє прогнозувати ціну на авто, якщо відомі його основні параметри. Тепер менеджер зможе визначити ціну авто, навіть якщо його рік випуску та пробіг раніше не зустрічалися в даному автосалоні.

1.2 Історія розвитку економетрії як науки

Методи математичної статистики вперше застосували в біології. Наприкінці XIX ст. англійський біолог К. Пірсон досліджував криві розподілу деяких числових показників людського організму.

Дещо пізніше він та його школа почали вивчати кореляції в біології та будувати лінійні регресії. Запропоновані біологами підходи були застосовані у економіці.

У 1897 р. вийшла у світ праця В. Парето, в якій було проведено дослідження доходів населення в різних країнах. У ній вперше була використана так звана крива Парето, параметри якої було отримано статистичними методами.

Була запропонована крива Парето:

$$y = A(x - a)^{-\alpha}, \tag{1.2}$$

де y – чисельність населення, що має дохід, більший за x;

x – величина доходу;

a – мінімальний дохід;

$A,\ \alpha$ – параметри залежності, які було отримано статистичними методами.

На початку XX ст. вийшло кілька праць англійського статистика

Д. Гукера, у яких за допомогою кореляційно-регресійних методів, що були започатковані школою К. Пірсона, вивчалися взаємозв’язки між економічними показниками, зокрема вплив банкрутств на товарній біржі на ціну зерна.

У подальшому почали з’являтися численні праці як з розвитку теорії математичної статистики та її прикладних елементів, так і з практичного застосування цих методів в економічному аналізі. Серед них були і праці Г. Мура, які вийшли друком протягом 1914 − 1917 рр.

Необхідно зауважити, що термін “економетрія” вперше запровадив львівський учений П. Чомпа, який опублікував у Львові в 1910 р. книгу “Нариси економетрії і природної теорії бухгалтерії, яка ґрунтується на політичній економії”. Проте поняття не набуло належного визнання, оскільки на той час не було фундаментальних праць у цій галузі науки.

Як самостійна дисципліна економетрія сформувалася у 20 − 30 рр. ХХ століття завдяки працям Г. Мура і Г. Шульца. У перших працях розроблялися аналітико-статистичні моделі. Здебільшого це були рівняння лінійної регресії з параметрами, оцінюваними за методом найменших квадратів. Такі рівняння дозволяли описувати функції попиту та їх залежність від прибутків, обсягів випуску продукції, цін, податків, інших чинників, а також функції пропозиції, виробничі функції, які відображали технологічну залежність випуску продукції від витрат праці та засобів виробництва.

У 1928 р. було опубліковано дослідження американських вчених математика Ч. Кобба та економіста П. Дугласа виробничої функції, яке ввійшло до економетрії як класичний приклад і досі є важливим інструментом економетричного аналізу:

$$Q = A \cdot K^{\alpha} \cdot L^{\beta},$$

де Q – обсяг виробництва;

A – постійний коефіцієнт;

α і β – показники, що характеризують віддачу, використання кожного з двох основних видів ресурсів;

K – витрати капіталу;

L – витрати праці.

Завдяки своїй простоті та раціональності ця функція широко застосовується досі й отримала подальші узагальнення в різних напрямках.

Праці цих вчених можна вважати основою сучасної економетрії, її методів та принципів.

Економетрія як окрема галузь науки відома під такою назвою, починаючи з 1930 р. Саме тоді було засновано економетричне товариство, яке на той час визначало себе так: “Міжнародне товариство для розвитку економічної теорії і її зв’язку зі статистикою та математикою”.

У 1933 році Р. Фріш проголосив синтез економічної теорії, статистики та математики. Економетристи спробували поєднати позитивні ознаки математичної школи в політичній економії та у статистичному напрямі, тобто підтримали ідею поєднання в економічному дослідженні абстрактно-теоретичного аналізу, емпіричних даних і математичних методів.

Одними із засновників економетрії вважають Р. Фріша, Е. Шумпетера, Я. Тінбергена – послідовників неокласичної економічної школи та кейнсіанства. Вони одними з перших науковців цілеспрямовано намагалися поєднати економічну теорію з математичними та статистичними методами.

Спочатку вчені обмежувалися вивченням деяких моделей попиту і пропозиції. Але після Другої світової війни вони почали вивчати комплексні економетричні моделі на макрорівні, в яких основна увага приділялася таким явищам, як попит, фінансовий стан, податки, прибуток, ціна тощо.

У процесі розвитку економетрії з'явилися ознаки її розшарування, відходу від триєдиної формули Р. Фріша. Вони розвивалися в двох напрямах:

- одночасно з економіко-теоретичними дослідженнями на основі застосування математичних і статистичних методів все більшого значення набували прикладні математичні та емпірико-статистичні розробки, які не належали безпосередньо до економічної теорії;
- виокремлювалися абстрактно-теоретичні дослідження математичних моделей економіки, в яких не використовувалися емпіричні дані.

У наш час набувають впровадження у вітчизняну практику економетричні підходи з використанням програмних комплексів ПК. За сучасних умов зростає роль економіко-математичних методів як одного із засобів розвитку динамічно розвиненої та стійкої економіки з науково обґрунтованими шляхами розвитку за умов кризи та прогнозами на майбутнє.

Окремо слід окреслити внесок українських вчених у розвиток економіко-математичних методів.

В останній третині XIX століття в Україні виник особливий напрям у розвитку економічної думки, відомий під назвою "київська наукова школа в політичній економії". Одним із засновників цієї школи був професор Київського університету М. Х. Бунге (1823-1895). Його послідовником став Д. І. Піхно (1853-1909), що опублікував низку праць, перша з яких називалася "Закон попиту і пропозиції". Платіжну здатність покупця у цій праці автор намагався подати математично. Повніше своє розуміння економічної теорії Д. І. Піхно виклав в "Основах політичної економії". Політична економія, за визначенням вченого, ставить собі за мету вивчення закономірностей економічних явищ. При цьому використовують різні методи, зокрема й математичні. Проте Д. І. Піхно вважав, що значимість математичних методів для з’ясування законів політичної економії не варто перебільшувати.

Світоглядної позиції київської школи політичної економії дотримувався і Р. М. Орженцький (1863-1923), який у праці “Основні закони цінності та їх пряме застосування” спробував поєднати математичні методи дослідження з психологічною теорією цінності.

О. Д. Білімович (1876-1963) у монографії “До питання про розцінку господарських благ” (1914) написав: “Намагання подати свої теоретичні надбудови у формі можливих точніших схем зближує нас з економістами математичної школи”.

В. Ф. Арнольд опублікував у 1904 р. в Одесі “Політично-економічні етюди”, в яких намагався використати засоби елементарної алгебри для обґрунтування теорії граничної корисності. Він писав: “Завданням нашим було … показати, що в результаті введення математичних прийомів оброблення економічних даних – хоч би ці прийоми були в найелементарнішому вигляді – багато політекономічних проблем дістають можливості строгого і точного, хоч і абстрактного розв’язання”. В. Ф. Арнольд усіляко відстоював застосування математичних методів в економічній науці, зокрема політичній економії. Він стверджував, що політична економія може мати шанси на успіх лише тоді, коли буде використовувати методи природничих наук. Спроби викласти економічні закони математичними формулами В. Ф. Арнольд вважав своїм основним завданням.

На початку XIX століття, коли статистика стала університетською наукою і необхідним інструментом державного управління, розпочався якісно новий етап в історії економіко-статистичної думки в Україні. Видатними вченими, які розвивали економіко-статистичну науку в цей період, були: Д. П. Журавський (1810-1856), В. М. Навроцький (1847-1882), О. О. Русов (1847-1915), В. Є. Варзар (1851-1940) та ін.

Українську економіко-статистичну думку кінця XIX – першої половини XX ст. не можна належно оцінити без урахування великої наукової спадщини Ф. А. Щербини (1849-1936) та М. Б. Птухи (1884-1961), які зробили значний внесок у розвиток методики оброблення статистичних даних.

Яскравим представником українських економістів-новаторів, які торували нові шляхи у розвитку світової економічної науки, був Є. Є. Слуцький (1880-1948). Першою його працею в економічні науці стало його студентське дослідження “Теорія граничної корисності”, написане у 1910 р. У цій праці він широко використав математичний апарат, без якого, за його словами, не міг би “… показати справжню взаємодію теорії попиту і пропозиції (в її сучасній конструкції) та теорії затрат виробництва”.

У 1912 році Є. Є. Слуцький підготував і видав посібник "Теорія кореляції і елементи вчення про криві розподілу", в якому використав найновіші здобутки математичної статистики.

Знаменитим дослідженням Є. Є. Слуцького початку XX ст. є праця "До теорії збалансованості бюджету споживача", опублікована в Італії в 1915 р. Це дослідження привернуло увагу зарубіжних учених-економістів лише через двадцять років після опублікування у зв'язку з появою у світовій науці економетрії. У 1935 американський економіст Г. Шульц зазначив, що Є. Є. Слуцький значно розширив, поглибив, конкретизував теорію споживчого попиту.

Англійський економіст-математик Р. Аллен, автор відомої книги "Математична економія", у середині 30-х років минулого століття опублікував статтю "Теорія споживчого вибору професора Слуцького", в якій відзначив наукове новаторство українського вченого. Пізніше, у 1958 році в журналі "Econometrica" Р. Аллен знову опублікував статтю про Є. Є. Слуцького, в якій він наголосив, що роботи українського вченого мали "великий і міцний вплив" на розвиток економетрії.

Про запозичення з робіт Є. Є. Слуцького зізнався у своїй книзі "Вартість і капітал" (1939) лауреат Нобелівської премії у галузі економіки Дж. Р. Гікс. Відзначивши високу математизованість праці українського вченого, Дж. Р. Гікс визнав, що саме Є. Є. Слуцький є фундатором ефекту доходу і, відповідно, ефекту заміщення, що сприяло введенню у зарубіжні підручники з економічної теорії терміна "ефект Слуцького". У своїй книзі англійський вчений писав, що Є. Є. Слуцький був першим економістом, який зробив значний крок уперед порівняно з класиками математичної школи (в економетричному напрямку).

1.3 Основні математичні передумови економетричного моделювання

Нехай маємо p пояснювальних змінних та залежну змінну Y. Змінна Y є *випадковою величиною,* що має при заданих значеннях факторів деякий розподіл. Якщо випадкова величина y неперервна, то можна вважати, що її розподіл для кожного припустимого набору значень факторів (x_1, …, x_p) має умовну щільність $f_{x_1,\ldots,x_p}(y)$.

Зазвичай роблять припущення щодо розподілу y (нормальний розподіл). Пояснювальні змінні X_j, $j=\overline{1,p}$ можуть вважатися як випадковими, так і детермінованими, тобто такими, що набувають певних значень. Із попереднього прикладу про продаж авто ми можемо заздалегідь визначити для себе параметри автомобіля та шукати оголошення про продаж авто з такими параметрами, тоді випадковою величиною залишається тільки ціна; а можемо випадковим чином обирати оголошення про продаж, тоді екзогенні змінні будуть випадковими величинами.

Класична економетрична модель розглядає екзогенні змінні x_j як детерміновані. Пояснювана частина Y_e є функцією від екзогенних змінних:

$$Y_e = f(X_1, \dots, X_p).$$

Таким чином, економетрична модель є такою:

$$Y = f(X_1, \dots, X_p) + \varepsilon.$$

Найбільш природним вибором залежної частини є її середнє значення – умовне математичне сподівання $M_{x_1,\dots,x_p}(Y)$, одержане для даного набору значень екзогенних змінних (x_1, …, x_p). (У подальшому математичне сподівання будемо позначати $M_x[Y]$).

Рівняння $M_x[Y] = f(x_1,\dots,x_p)$ називають *рівнянням регресії*. При такому виборі залежної частини економетрична модель є такою:

$$Y = M_x[Y] + \varepsilon, \tag{1.3}$$

де ε – випадкова величина, яку називають *збуренням* або *похибкою*.

Рівняння (1.3) називають *рівнянням регресійної моделі*. Однак відмітимо, що економетрична модель не обов’язково є регресійною, тобто залежна частина не завжди є умовним математичним сподіванням залежної змінної. З математичної точки зору регресійні моделі є більш простими об’єктами, ніж економетричні моделі загального типу. Відзначимо деякі властивості регресійної моделі.

Розглянемо рівність (1.3) та обчислимо від обох частин математичне сподівання при заданому наборі значень екзогенних змінних X. У цьому випадку $M_x[Y]$ є числовою величиною, що дорівнює своєму математичному сподіванню, і можна отримати рівність

$$M_x[\varepsilon] = 0, \tag{1.4}$$

тобто, в регресійній моделі очікуване значення випадкової похибки дорівнює нулю. Звідси випливає некорельованість випадкових помилок та екзогенних змінних X.

1.4 Економетрична модель та експериментальні дані

Щоб одержати достатньо достовірні та інформативні дані щодо розподілу деякої випадкової величини, необхідно мати вибірку її спостережень достатньо великого об’єму. Вибірка спостережень залежної змінної та екзогенних змінних X_j, $j=\overline{1,p}$, є відправною точкою будь-якого економетричного дослідження.

Такі вибірки є наборами значень X_j, Y_i, $j=\overline{1,p}$, $i=\overline{1,n}$; n – число спостережень, p – кількість екзогенних змінних. Як правило, число спостережень достатньо велике (десятки, сотні) та значно перевищує кількість екзогенних змінних. Однак проблема в тому, що випадкові величини Y_i, одержувані при різних наборах значень X_j, строго кажучи, мають різні розподіли. А це означає, що для кожної випадкової величини Y_i ми маємо лише одне спостереження.

У класичному курсі економетрії розглядають два типи вибіркових даних:

1. *Просторові дані (cross-sectional data)*. В економіці під просторовою вибіркою розуміють набір економічних змінних, одержаний на даний момент часу. Однак таке означення не дуже зручне для економетриста через неоднозначність поняття “момент часу”. Зрозуміло, що говорити про просторові вибірки є сенс у випадку одержання всіх спостережень за незмінних умов (набір спостережень є набором незалежних вибіркових даних з деякої генеральної сукупності).

Таким чином, ми будемо називати *просторовою вибіркою* серію n незалежних спостережень (p+1)-вимірної випадкової величини (X_{i1}, X_{i2}, …, X_{ip}; Y_i).

Як визначити, чи є вибірка серією незалежних спостережень? На це запитання немає однозначної відповіді. Зазвичай за незалежні приймають величини, не пов’язані причинно. Однак на практиці далеко не завжди питання незалежності розв’язується просто.

Повернемося до прикладу про продаж автомобілів.

Нехай Y – ціна автомобіля, X – рік випуску, а (x_1, y_1), … , (x_n, y_n) – серія даних, одержаних з газети “20 хвилин”. Чи можна вважати ці спостереження незалежними?

Різні продавці не знайомі між собою, вони дають свої оголошення незалежно один від одного, тому пропозиція щодо незалежності спостережень виглядає досить розумною. З іншого боку, людина, яка призначає ціну на свій автомобіль, керується цінами попередніх оголошень, тому і заперечення незалежності спостережень також має право на існування.

Із цього можна зробити висновок, що рішення щодо просторового характеру вибірки певною мірою є суб’єктивним. Таким чином, модель, побудована на основі просторової вибірки експериментальних даних (x_i, y_i), є такою:

$$y_i = f(x_i) + \varepsilon_i, \quad i = \overline{1,n}, \tag{1.5}$$

де похибки регресії мають задовольняти умови:

$$M(\varepsilon_i) = 0, \tag{1.6}$$

$$r(\varepsilon_i, \varepsilon_j) = 0, \tag{1.7}$$

$$D(\varepsilon_i) = \sigma_i^2, \tag{1.8}$$

де $D(\varepsilon_i)$ – умовна дисперсія випадкової величини ε_i.

Що стосується умови (1.8), то тут можливі два випадки:

а) $\sigma_i^2 = \sigma_j^2$ при всіх i та j. Властивість сталості дисперсій похибок регресії називається *гомоскедастичністю*. У цьому випадку розподіли випадкових величин Y_i відрізняються лише значенням математичного сподівання (пояснюваної частини);

б) $\sigma_i^2 \neq \sigma_j^2$. У цьому випадку має місце *гетероскедастичність* моделі. Гетероскедастичність потрібно усунути.

Перевірку моделі на гомоскедастичність можна проілюструвати на такому прикладі. Наприклад, ціна автомобіля, якому 15 років, навряд чи може піднятися вище 2000 ум. гр. од., тому стандартна похибка ціни в цьому випадку не перевищить 300 − 400 ум. гр. од. Разом із тим, автомобіль якому два роки, може коштувати і 7000, і 17000 ум. гр. од., тобто стандартна похибка не менше 1500 − 2000 ум. гр. од.

Однак у багатьох випадках гетероскедастичність моделі не є очевидною, і потрібно застосовувати методи математичної статистики для прийняття рішення про те, який тип моделі буде розглядатися. Це питання буде більш ретельно розглянуто у розділі 5;

2. *Часовий (динамічний) ряд (time-series data).* Часовим (динамічним) рядом називають вибірку спостережень, в якій важливі не лише самі спостережувані значення випадкових величин, а й порядок їхнього слідування. Найчастіше впорядкованість зумовлена тим, що експериментальні дані є серією спостережень однієї і тієї ж випадкової величини в послідовні моменти часу. У цьому випадку динамічний ряд називають *часовим рядом*.

1.5 Поняття економіко-математичної моделі. Класифікація економіко-математичних моделей

Вплив математичного моделювання на економічну теорію є різнобічним. Викладення багатьох економічних проблем формалізованою мовою дає можливість запобігти двозначності міркувань, значною мірою прояснює суть проблеми, яскраво інтерпретує теоретичні положення. Окрім того, застосування мови математики сприяє уточненню багатьох економічних категорій, кращій систематизації теоретичних знань, збагаченню понятійного апарату економічної науки.

Економіко-математичною моделлю називають сукупність пов'язаних між собою математичними залежностями величин − факторів, всі чи частина яких мають економічний характер. ***Моделювання*** є процесом побудови, реалізації та дослідження моделі, який здатний замінити реальну систему та дати інформацію про неї. Моделювання є важливим інструментом наукової абстракції, що допомагає виділити, уособити та проаналізувати суттєві для даного об'єкта характеристики (властивості, взаємозв'язки, структурні та функціональні параметри).

Дослідження з моделюванням економічних систем поділяються на три основні групи:

- теоретичні дослідження – розроблення проблем економічної теорії та теоретико-методологічних проблем управління економікою з використанням математичного моделювання;
- прикладні дослідження – розв'язання практичних завдань управління економічними системами;
- інструментальні дослідження – створення інструментальних засобів для проведення економічних досліджень.

Економіко-математична модель має пізнавальну і практичну цінність, якщо вона ***відповідає*** певним ***вимогам***:

- спирається на основні положення економічної теорії;
- адекватно відображає реальну економічну дійсність;
- враховує найбільш важливі фактори, які визначають рівень досліджуваних показників;
- відповідає встановленим критеріям;
- дозволяє отримати такі знання, які до її реалізації були невідомими;
- може бути достатньо абстрактною, щоб припустити варіювання великим числом змінних, але не настільки, щоб виникли сумніви в її надійності і практичній корисності отриманих результатів;
- задовольняти умови, які обмежують час розв'язування задачі;
- дозволяє реалізувати її існуючими засобами.

Економіко-математичні моделі дозволяють виявити особливості функціонування економічних об'єктів або явищ і на базі цього передбачити майбутню поведінку об'єкта при зміні будь-яких параметрів.

У моделях всі зв'язки між змінними можуть бути описані кількісно, що дозволяє отримати більш якісний та надійний прогноз. Можливості прогнозування полягають в тому, що можна отримати набагато кращі результати та позбутися зайвих витрат. Неповнота економічних моделей випливає з їх абстрактності.

Звичайна економіко-математична модель складається з цільової функції:

$$f\,(x_1, ..., x_n) \to extreme \qquad (1.9)$$

від шуканих величин $x_1, ..., x_n$ та обмежень на область використання цих величин:

$$g_k\,(x_1, ..., x_n) < b_k\ (k = 1,..,m). \qquad (1.10)$$

Цільова функція $f(x_1, ..., x_n)$ виражає значення критерію оптимальності, який зумовлений значеннями шуканих величин $x_1, ..., x_n$, а $g_k(x_1, ..., x_n)$ – техніко-економічні умови досліджуваного процесу.

Наприклад, $f\ (x_1,...,x_n)$ – прибуток підприємства залежно від обсягу виробленої продукції: першого виду – x_1, другого виду – x_2, ..., n-го виду – x_n. Тоді обмеженнями на таку функцію будуть $g_k\ (x_1, ..., x_n)$ – обсяги споживаних ресурсів k-го виду, а b_k – обсяги виділених ресурсів k-го виду ($k = 1, ..., m$).

Оскільки підприємству необхідно розробити план виробництва продукції, який забезпечує *максимізацію прибутку* при використанні лише виділених ресурсів b_k, то економіко-математична модель такої задачі полягає у пошуку таких $x_1,..., x_n > 0$, за яких: $f\,(x_1, ..., x_n) \to max$.

Якщо в задачі про пошук оптимального плану виробництва продукції підприємство використовує критерій оптимальності мінімуму витрат за умов виконання плану, то цільова функція $f\ (x_1, ..., x_n)$ буде виражати витрати, а обмеження $g_i(x_1, ..., x_n) \geq b_i$ – умови виконання плану виробництва продукції за видами ($i = 1, 2, ..., n$).

Тоді математична модель набуде вигляду:

$$f\,(x_1,..., x_n) \to min,\ g_i\,(x_1, ..., x_n) \geq b_i\ (i = \overline{1,n}). \qquad (1.11)$$

Економіко-математичні моделі поділяють на класи за рядом ознак. *Залежно від об'єкта моделювання* та *математичного апарату* виділяють такі моделі: макро- та мікроекономічні, теоретичні та прикладні, статичні та динамічні, детерміновані та стохастичні, оптимізаційні та моделі рівноваги тощо.

Макроекономічні моделі описують економіку загалом, пов'язуючи між собою узагальнені показники матеріального та фінансового плану:

ВВП, споживання, інвестиції, зайнятість, процентну ставку, кількість грошей тощо.

Мікроекономічні моделі описують взаємодію структурних і функціональних складових економіки або поведінку окремої складової, зокрема фірми, підприємства, банку і т. п. у ринковому середовищі.

Статичні моделі описують стан економічного об'єкта в певний момент чи період часу, а ***динамічні моделі*** вивчають взаємозв'язки економічних змінних у часі. Ті змінні, що вивчаються в динаміці, у статичних моделях мають фіксоване значення. Однак динамічна модель не зводиться до простої суми статичних моделей, а описує взаємодію сил, що рухають економіку.

У динамічних моделях вхідні фактори та результат прямо залежать від часу, а у статичних залежність від часу є або зовсім відсутньою, або дуже слабкою.

У ***детермінованих моделях*** передбачаються жорсткі функціональні зв'язки між змінними, а ***стохастичні*** припускають наявність випадкових впливів на досліджувані змінні.

Моделі рівноваги описують такий стан економіки, коли всі сили, що намагаються вивести її з рівноваги, мають нульову сумарну дію.

Оптимізаційні моделі застосовують для пошуку найкращих управлінських рішень за певним критерієм оптимізації при дотриманні низки обмежень. Серед основних критеріїв можна зазначити: максимізацію прибутку, мінімізацію грошових або часових витрат та ін.

Необхідно зауважити, що предметом економетричного дослідження є прикладні стохастичні економічні моделі, тобто загальні економічні моделі, у яких модельні коефіцієнти набувають конкретних числових значень залежно від використаної статистичної інформації, що була попередньо підготовлена.

Також економіко-математичні моделі можна класифікувати за такими ознаками: *призначенням, ступенем ймовірності, способом врахування змінювання процесу у часі, точністю математичного відображення досліджуваних процесів та явищ*.

За призначенням моделі поділяються на чотири класи: імітаційні, балансові, оптимізаційні, сітьові.

За ступенем імовірності моделі поділяють на два класи: ймовірнісні (стохастичні), параметри та зовнішні зміни яких носять випадковий характер; детерміновані, в яких випадковий характер зміни параметрів не береться до уваги.

Серед економіко-математичних моделей виділяють також такі типи:

1) ймовірнісно-статистичні моделі. Моделі вартості та розширеного відтворення є ймовірнісно-статистичними, тому дослідження їх здійснюється за допомогою методів економічної та математичної статистики з використанням апарату теорії ймовірностей. Насамперед тут

використовується вибірковий (репрезентативний метод), який дозволяє за обмеженими статистичними даними визначити окремі економічні показники, а також оцінити ступінь точності отриманих результатів. Найбільш важливими в економетричній статистиці є методи *кореляційно-регресійного аналізу*.

Одним із напрямків кореляційно-регресійного аналізу в економічних дослідженнях є моделювання залежності між обсягом виробленої продукції, собівартістю одиниці продукції, капітальними витратами, продуктивністю праці та використанням основних виробничих фондів. Ці ж методи використовують при дослідженні залежності попиту від ціни, пропозиції від ціни та майбутніх впливів пропозиції на ціну, тобто аналізується завершений цикл послідовних впливів одного чинника на інший.

Методами математичної статистики досліджується кореляція між окремими елементами в суспільному виробництві.

Окремим підприємствам побудовані математичні криві слугують не лише для вивчення товарного ринку. Їх використовують для встановлення та аналізу функціональної залежності між величиною випуску продукції та витратами виробництва у формі так званих виробничих функцій, параметри яких отримують також в результаті оброблення емпіричних даних методами кореляційного аналізу.

Статистичне моделювання (метод Монте-Карло) використовується в тих випадках, коли за умов складних взаємозв'язків факторів аналітичні методи є безсилими. Застосування методу статистичних випробувань безпосередньо пов'язано з ПК. Суть статистичного моделювання полягає у чисельному відтворенні (імітації) випадкових процесів за заздалегідь відомими параметрами, а також у визначенні невідомих параметрів моделі.

Цей метод успішно використовується для розв'язання питань підвищення ефективності використання обладнання та полегшення пошуку раціональної організації складних виробничих та технологічних процесів. Універсальність моделей, простота алгоритмів, які їх реалізують, а також наявність персонального комп'ютера роблять статистичне моделювання незамінним апаратом в економічних дослідженнях;

2) матричні моделі. У цьому класі моделей одержує строгий математичний опис один із важливих методів планування – *балансовий метод*. Ці моделі призначені для аналізу та планування виробництва і розподілу продукції на різних рівнях ієрархії (від окремого підприємства до цілого народного господарства). Назву *матричні* ці моделі одержали завдяки тому, що для їх реалізації використовується математичний апарат матричної алгебри. До простіших матричних моделей належать моделі, що описують поведінку економічної системи у вигляді *однопродуктових* схем

виробництва продукції та її розподілу. В цих моделях використовують деякий однорідний продукт як предмет праці і предмет споживання.

Загальність матричних моделей різних рівнів ієрархії полягає у тому, що їх можна розглядати за змістом і структурою на прикладі однієї з них.

Залежно від того, чи береться до уваги при складанні моделей фактор часу, вони можуть бути відповідно *статичними* чи *динамічними*.

Матричні статичні моделі розробляються для окремо взятих періодів. Зв'язок між попередніми або наступними періодами в рамках цих моделей не досліджується.

Динамічні матричні моделі відображають не стан, а процес розвитку економіки, встановлюють безпосередній зв'язок між минулими і майбутніми періодами розвитку.

Важливість матричних моделей в тому, що вони дозволяють формалізувати розрахунки та реалізувати ці моделі на ПК, а також забезпечують організацію інформації в найбільш економній формі;

3) моделі оптимального планування. Вони відрізняються тим, що, на відміну від балансових методів, враховують кілька способів виробництва (споживання). Крім того, змінні (вхідні та вихідні) задаються не ззовні моделі, а визначаються за умов оптимальності цільової функції. При цьому моделі оптимального планування дозволяють розв'язувати задачі не лише міжгалузевих балансів, але й також розміщення виробничих сил, спеціалізації та кооперування підприємств.

До *моделей оптимального планування* належать:

- планування на підприємствах та будівництві;
- планування постачання та перевезення;
- управління запасами;
- сумішно-розкрійні задачі.

Планування на підприємствах та будівництві. До цих моделей належать задачі оптимізації виробничої програми підприємства та її розподілу за календарними періодами, оптимальне завантаження виробничих агрегатів і машин, розрахунок виробничих потужностей підприємства, складання оптимальних графіків запуску виробництва, випуску виробів й ін.

Планування постачання та перевезення. Основна мета цього напрямку – мінімізація транспортних витрат при перевезенні різних вантажів від постачальників до споживачів. При цьому можуть використовуватися різні обмеження: пропускна здатність окремих ланцюгів транспортної мережі, взаємозамінність деяких видів вантажу, першочерговість перевезення найбільш важливих вантажів і т. п. Сюди ж відносять і задачу комівояжера та перевезення дрібних партій вантажу.

Управління запасами. Методи оптимального планування застосовують при розв'язанні різних проблем постачання та збуту, раціонального розміщення оптових і роздрібних баз, а також планування роботи товарної мережі, оптимального керування запасами.

Сумішно-розкрійні задачі. До цієї категорії належать задачі оптимального складу сумішей та сполук. Ці задачі зустрічаються на підприємствах, де продукцію одержують в результаті змішування, сплавлювання або сполучення деяких видів компонентів сировини чи матеріалів. Методи оптимального планування дозволяють знайти набір компонентів суміші, при якому продукція даного складу та якості буде отримана за мінімальних витрат. У задачі щодо оптимального розкрою матеріалів критерієм оптимальності є мінімальні сумарні витрати (відходи) матеріалів після розрізання їх на заготівки необхідної величини та форми.

Усі перелічені моделі досліджуються засобами *лінійного програмування*.

Типовими прикладами економічних задач управління запасами є задачі виробництва та збереження продукції, розподілу капіталовкладень, календарного виробничого планування, складання графіків запуску деталей у виробництво, визначення найкоротшої відстані між пунктами на транспортній мережі. Розв'язання таких задач досягається засобами *динамічного програмування*.

1.6 Етапи економіко-математичного моделювання

Можна виділити шість основних етапів економіко-математичного моделювання: постановочний, апріорний, етап параметризації, інформаційний етап, етапи ідентифікації та верифікації моделі.

1-ий етап (постановочний). Формується мета дослідження, набір економічних змінних, що будуть брати участь у побудові моделі.

Як мету зазвичай розглядають аналіз досліджуваного економічного процесу, прогнозування його економічних показників, імітацію розвитку об'єкта при різних значеннях екзогенних змінних, розроблення управлінських рішень.

При виборі економічних змінних необхідне теоретичне обґрунтування кожної змінної. Для вибору змінних можуть бути використані різні методи, а для оцінювання впливу якісних ознак використовують навіть фіктивні змінні.

2-ий етап (апріорний). Проводиться аналіз суті об'єкта, що вивчається, формування та формалізація апріорної інформації (відомої до початку моделювання).

Однією з основних проблем використання математичного моделювання в економічних дослідженнях є наявність якісної інформації. Точність і повнота первинної інформації, реальні можливості її одержання та оброблення визначають вибір типу математичної моделі і можливості її використання.

Якість інформації можна визначити як сукупність властивостей, що зумовлюють можливість її використання для задоволення певних потреб, залежно від її призначення. Виділимо суттєві показники якості інформації:

- *репрезентативність* – правильність відбору та формування інформації з метою адекватного відображення необхідних властивостей економічних систем;
- *змістовність* – відношення кількості семантичної інформації до обсягу даних;
- *повнота* – інформація містить мінімальний, але достатній для прийняття необхідного рішення набір економічних показників;
- *доступність* – інформація повинна бути зрозумілою та зручною у використанні;
- *актуальність* – ступінь збереження цінності інформації у момент її використання, який залежить від динаміки зміни параметрів економічних систем;
- *стійкість* – властивість інформації реагувати на зміни вхідних даних, зберігаючи при цьому необхідну точність;
- *точність* – ступінь близькості статистичного значення показника до його істинного значення. Для кількісних економічних показників використовують чотири класифікаційні поняття точності: формальна точність (вимірюється значенням наймолодшого розряду числа, яким поданий показник); реальна точність (розраховується значенням останнього розряду числа); досяжна точність (максимальна точність, яку можна одержати за даних конкретних умов функціонування економічної системи); необхідна точність (визначається функціональним призначенням економічного показника);
- *достовірність* – властивість інформації відображати реальні значення параметрів економічної системи з необхідною точністю;
- *цінність* – міра інформації на прагматичному рівні.

Відповідність інформації об’єкту дослідження – *адекватність* – виражають у трьох формах: синтаксичній, семантичній, прагматичній. Відповідно до цих форм адекватності здійснюють і вимірювання інформації.

Синтаксичною мірою інформації є ентропія системи, яку визначають за формулою Шеннона:

$$H = -\sum_{i=1}^{n} p_i \log_a p_i ,$$

де H – ентропія (невизначеність стану) системи;

p_i, $i = \overline{1,n}$ – ймовірність того, що система перебуває в i-му стані;

n – кількість станів системи.

Щоб виміряти кількість інформації на семантичному рівні, тобто змістовний вміст інформації, найчастіше використовують тезаурусну міру інформації (тезаурус – сукупність знань, які має користувач). Кількість семантичної інформації – це її корисність, цінність для користувача. Ця міра інформації є відносною величиною, зумовленою особливостями використання інформації. Наприклад, якщо інформацію використовують для моделювання управління економічною системою, то цінність інформації доцільно вимірювати в тих самих одиницях, у яких вимірюють цільову функцію управління системою.

Прагматичною мірою інформації в системах управління економічними об'єктами буде економічна ефективність функціонування системи управління.

3-ій етап (параметризація). Здійснюється безпосереднє моделювання, тобто вибір загального вигляду моделі.

Основна задача, яка розв'язується на цьому етапі, – вибір вигляду функції $f(X)$ в економетричній моделі (1.1). Дуже серйозною проблемою на даному етапі є проблема специфікації моделі, зокрема запис у математичній формі виявлених зв'язків та взаємовідношень.

4-ий етап (інформаційний). В економетричному дослідженні важливе значення має інформація, на підставі якої будують модель. Тому на даному етапі збирають необхідну статистичну інформацію. Тут можуть розглядатися дані, одержані як за участі дослідника, так і без неї (за умов пасивного чи активного експерименту). Статистичну інформацію можна поділити на два види:

- апріорна інформація, яка має якісний характер; джерелом апріорної інформації є економічна теорія;
- апостеріорна інформація – має кількісний характер; джерелом цієї інформації є спостереження, досліди (статистичні дані).

5-ий етап (ідентифікація моделі). Здійснюється математико-статистичний аналіз моделі та оцінювання її параметрів.

6-ий етап (верифікація моделі) Здійснюється перевірка істинності, адекватності моделі. З'ясовують, наскільки вдало розв'язано проблеми специфікації, ідентифікації, точності розрахунків за даною моделлю.

Для верифікації економіко-математичних моделей їх часто порівнюють з іншими моделями, які вже знайшли своє практичне застосування і довели свою ефективність, а також застосовують теоретичні методи перевірки адекватності моделей.

До причин порушення адекватності слід віднести: відсутність суттєвих зв'язків, хибну структуру досліджуваного об'єкта, неточну оцінку змінних, спрощення функціональних залежностей, використання нераціональних методів оцінювання параметрів.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Що таке економетрія?
2. Які основні етапи розвитку економіко-математичних досліджень?
3. Які вчені входили до математичної школи політичної економії?
4. Яка офіційна дата народження економетрії як науки?
5. Які вчені сформували теоретичні та практичні основи економетрії?
6. Хто і коли вперше використав термін "економетрія"?
7. Що таке модель?
8. Охарактеризуйте основні етапи моделювання.
9. Як класифікують моделі?
10. Що таке економіко-математична модель?
11. Як класифікують економіко-математичні моделі?
12. Які є показники якості інформації?
13. Які існують форми вираження адекватності інформації?
14. Які наукові теорії пропагували представники київської наукової школи в політичній економії?
15. Які вчені розбудовували київську наукову школу політичної економії?
16. Які вчені розвивали українську економіко-статистичну науку?
17. Визначте внесок Є. Є. Слуцького у розвиток економічної теорії.
18. Хто з економетристів є лауреатом Нобелівської премії в галузі економіки?
19. Що є метою економетрії?
20. Знання з яких наук поєднує в собі економетрія?
21. Які величини називають пояснюваними, а які – пояснювальними?
22. Що є задачею економетричного моделювання?
23. Охарактеризуйте криву В. Парето.
24. Охарактеризуйте основні математичні передумови економіко-математичного моделювання.
25. Які типи вибіркових даних розглядають в економетрії?
26. Що називають гомоскедастичністю, а що – гетероскедастичністю?
27. Що називають часовим рядом?
28. На які групи поділяють економетричні дослідження з моделюванням економічних систем?
29. Яким вимогам повинна відповідати економіко-математична модель?
30. Охарактеризуйте структуру економіко-математичної моделі.
31. Охарактеризуйте макро- та мікроекономічні моделі. Які моделі називають статичними?
32. Охарактеризуйте імовірнісно-статистичні та матричні моделі.
33. В яких випадках використовують статистичне моделювання?
34. Охарактеризуйте моделі оптимального планування.

ВПРАВИ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Розгадайте кросворд, зображений на рис. 1.1.

По горизонталі

2. Набір економічних змінних, одержаний на даний момент часу, називають вибіркою ... .
3. Моделі, призначені для аналізу та планування виробництва й розподілу продукції на різних рівнях ієрархії.
5. Властивість інформації відображати реальні значення параметрів економічної системи з необхідною точністю.
6. Процес побудови, реалізації та дослідження моделі, який здатний замінити реальну систему та дати інформацію про неї.
7. Львівський вчений, що вперше запровадив термін "економетрія".
8. Один із розробників теорії оптимального розподілу ресурсів, лауреат Нобелівської премії 1975 року.
10. Розробник кривої з дослідження доходів населення в різних країнах.
12. Наука, що встановлює та досліджує кількісні закономірності та взаємозв'язки в економічному житті за допомогою математичних і статистичних методів.
13. Норвезький економіст, що застосував основи теорії ймовірностей в межах економетрії та проаналізував залежні економічні структури.
14. Показник, що вимірює відношення кількості семантичної інформації до обсягу даних.
17. Функція, що виражає значення критерію оптимальності, який зумовлений значеннями шуканих величин.
20. Синтаксична міра інформації.
21. Ступінь близькості статистичного значення показника до його точного значення.
24. Автор праці "До теорії збалансованості бюджету споживача".
26. Моделі, що описують взаємодію структурних і функціональних складових економіки або поведінку окремої складової в ринковому середовищі.
27. Автор праці "Закон попиту і пропозиції".
29. Властивість сталості дисперсій похибок регресії.
30. Якісна інформація, яку почерпнуто з економічної теорії.

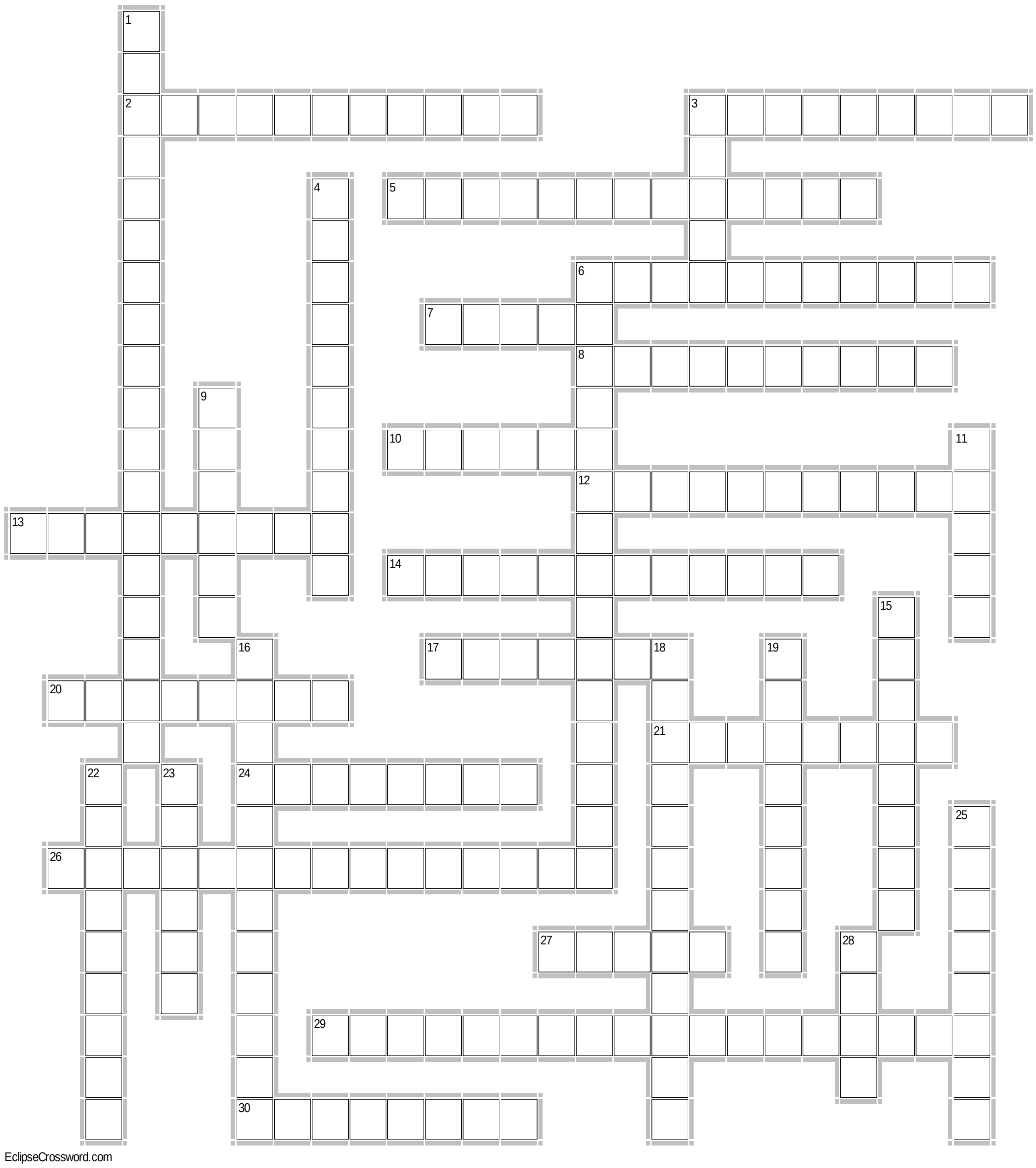


Рисунок 1.1 – Кросворд до розділу 1

По вертикалі

1. Показник якості інформації, що характеризує правильність відбору та формування інформації.
3. Один із засновників київської наукової школи політичної економіки.
4. Лауреат Нобелівської премії за розроблення статистичної та динамічної економічної теорії.
6. Моделі, які описують економіку загалом, пов’язуючи між собою узагальнені показники матеріального та фінансового плану.

9. Один із авторів класичної виробничої функції.

11. Вибірку спостережень, в якій важливі не лише спостережувані значення випадкових величин, а й порядок їх слідування, називають часовим ...

15. Моделі, які описують стан економічного об'єкта в певний момент чи період часу.

16. Кількісна інформація, яку одержано зі спостережень.

18. Ступінь збереження цінності інформації у момент її використання, який залежить від динаміки зміни параметрів економічних систем.

19. Міра інформації на прагматичному рівні.

22. Властивість інформації реагувати на зміни вхідних даних, зберігаючи при цьому необхідну точність.

23. Один із авторів теорії оптимальних механізмів, лауреат Нобелівської премії 2007 року.

25. Автор балансових моделей для моделювання взаємозв'язків великою кількістю змінних.

28. Один із творців економетрії, норвезький економіст, лауреат Нобелівської премії 1969 року.

ТЕСТИ

Оберіть правильну (правильні) відповіді

1. В економетричній моделі залежні величини називають:
   а) пояснюваними;
   б) ендогенними;
   в) пояснювальними;
   г) екзогенними.

2. У економетричній моделі залежна змінна розбивається на такі дві частини:
   а) пояснювану та випадкову;
   б) факторну та результативну;
   в) ендогенну та екзогенну;
   г) жодної правильної відповіді.

3. Засновником математичної статистики був:
   а) лейб-медик короля Людовіка XV Франсуа Кене;
   б) біолог К. Пірсон;
   в) Карл Маркс;
   г) жодної правильної відповіді.

4. Перша робота з економетрики була написана у 1897 р. Її автором був:

а) Хікс;
б) К. Пірсон;
в) В. Парето;
г) Піфагор.

5. Економетричні дослідження В. Парето були присвячені:

а) вивченню розвитку суспільних закономірностей;
б) вивченню монетарної теорії;
в) вивченню доходів населення в різних країнах;
г) жодної правильної відповіді.

6. Крива В. Парето має вигляд:

а) $y = A(x-a)^{-\alpha}$;
б) $y = (x-a)(x-b)$;
в) $y = (x-a)^{\alpha}$;
г) жодної правильної відповіді.

7. Хто на початку ХХ ст. застосував кореляційно-регресійні методи, розроблені К. Пірсоном, для вивчення взаємозв'язку між кількістю банкрутств на товарній біржі та ціною зерна?

а) Д. Гукер;
б) К. Пірсон;
в) Хікс;
г) Ф. Кене.

8. Вперше запровадив термін «економетрія»:

а) Г. Мур;
б) П. Чомпа;
в) Г. Шульц;
г) К. Пірсон.

9. Одна з перших виробничих функцій була побудована:

а) Ч. Коббом та П. Дугласом у 1928 р.;
б) Хіксом у 1913 р.;
в) К. Пірсоном;
г) Д. Гукером.

10. Як самостійна дисципліна економетрія сформувалася у 20-30 рр. ХХ ст. завдяки працям:

а) К. Пірсона;
б) Д. Гукера;
в) Г. Мура і Г. Шульца;

г) жодної правильної відповіді.

11. У перших працях Г. Мура і Г. Шульца розроблялися:
а) оптимізаційні моделі;
б) аналітико-статистичні моделі;
в) моделі штучного інтелекту;
г) моделі нечіткої логіки.

12. До засновників економетрії належать:
а) Р. Фріш;
б) Е. Шумпетер;
в) Я. Тінберген;
г) жодної правильної відповіді.

13. У 1933 році Р. Фріш проголосив синтез:
а) економічної теорії, політекономії та основ ринкової економіки;
б) економічної теорії, статистики та математики;
в) дискретної математики, теорії ймовірностей та економіки;
г) статистики, менеджменту та маркетингу.

14. Засновниками "київської наукової школи в політичній економії" в останній третині XIX ст. були:
а) М. Х. Бунге;
б) Д. І. Піхно;
в) Є. Є. Слуцький;
г) Дж. Р. Гікс.

15. Видатними вченими, які розвивали економіко-статистичну науку на початку XIX ст. слід вважати:
а) В. М. Навроцького;
б) Д. П. Журавського;
в) О. О. Русова;
г) жодної правильної відповіді.

16. До розвитку української економіко-статистичної думки кінця XIX – першої половини XX ст. доклали зусиль:
а) Ф. А. Щербина;
б) М. Б. Птуха;
в) Є. Є. Слуцький;
г) жодної правильної відповіді.

17. Серед основних наукових доробків Є. Є. Слуцького вважають розвиток:
а) теорії споживчого попиту;

б) теорії граничної корисності;
в) теорії рівноваги фірми на ринку;
г) жодної правильної відповіді.

18. Економетрична модель набуває вигляду:
а) $Y = f(X_1, \ldots, X_p)$;
б) $Y = f(X) - \varepsilon$;
в) $Y = f(X_1, \ldots, X_p) + \varepsilon$;
г) жодної правильної відповіді.

19. У класичному курсі економетрії розглядають такі типи вибіркових даних:
а) просторові дані та часовий (динамічний) ряд;
б) cross-sectional data та time-series data;
в) репрезентативні та статичні дані;
г) жодної правильної відповіді.

20. Економіко-математичною моделлю називають:
а) формальний опис математичних об'єктів;
б) вербальний опис економічних об'єктів;
в) сукупність пов'язаних між собою математичними залежностями величин − факторів, всі чи частина яких мають економічний характер;
г) жодної правильної відповіді.

21. Моделювання – це:
а) вербальний опис економічних об'єктів;
б) графічний опис математичних об'єктів;
в) процесом побудови, реалізації та дослідження моделі;
г) жодної правильної відповіді.

22. Економіко-математична модель має відповідати певним вимогам:
а) адекватно відображати реальну економічну дійсність;
б) не відповідати встановленим критеріям;
в) враховувати найважливіші фактори;
г) не використовувати положення економічної теорії.

23. Моделі, в яких вхідні фактори та результат прямо залежать від часу, називають:
а) статичними;
б) динамічними;
в) кореляційними;
г) регресійними.

24. Економіко-математична модель складається з:

а) цільової функції $f(x_1, ..., x_n) \to extreme$;

б) обмежень на область використання шуканих величин: $g_k\ (x_1, ..., x_n) < b_k\ (k = 1,..,m)$;

в) y та x;

г) жодної правильної відповіді.

25. Цільова функція $f(x_1, ..., x_n)$ виражає значення:

а) техніко-економічних умов досліджуваного процесу;

б) критерію оптимальності, який зумовлений значеннями шуканих величин;

в) апроксимованої функції;

г) жодної правильної відповіді.

26. Обмеження на область використання шуканих величин визначають:

а) техніко-економічні умови досліджуваного процесу;

б) критерій оптимальності, який зумовлений значеннями шуканих величин;

в) апроксимовану функцію;

г) жодної правильної відповіді.

27. Серед економіко-математичних моделей виділяють такі:

а) ймовірнісно-статистичні моделі;

б) матричні та оптимального планування;

в) математичні та статистичні;

г) статистичні та статичні.

28. Назвіть основні етапи економіко-математичного моделювання:

а) постановочний, апріорний, параметричний;

б) інформаційний, ідентифікаційний, верифікаційний;

в) початковий, кінцевий;

г) жодної правильної відповіді.

29. До суттєвих показників якості інформації належать:

а) репрезентативність та адекватність;

б) змістовність та повнота;

в) доступність, стійкість та точність;

г) жодної правильної відповіді.

30. Синтаксичною мірою інформації є:

а) ентропія системи, що визначається за формулою Шеннона;

б) байт;

в) біт;

г) Гц.

ТЕМА 2 ЕЛЕМЕНТИ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

2.1 Випадкові величини та їх числові характеристики

Випадковою величиною називається змінна, що в результаті проведення випробування залежно від випадку набуває одного з можливих значень.

Прикладами випадкових величин є:

1) число новонароджених протягом доби у м. Вінниці;
2) кількість бракованих виробів в партії;
3) витрати електроенергії на підприємстві за місяць.

Випадкова величина називається *дискретною*, якщо множина її можливих значень є зчисленною (скінченною чи нескінченною).

Під *неперервною* випадковою величиною розуміють величину, множина можливих значень якої є деяким проміжком числової осі.

У прикладах 1 і 2 наявні дискретні випадкові величини, а у 3-му прикладі – неперервна випадкова величина.

Випадкові величини будемо позначати великими літерами латинської абетки $X, Y, Z, \ldots$, а їх можливі значення – відповідними маленькими літерами $x, y, z, \ldots$.

Законом розподілу випадкової величини називається будь-яке співвідношення, що встановлює зв'язок між можливими значеннями випадкової величини та відповідними їм ймовірностями.

Розглянемо дискретну випадкову величину X з можливими значеннями $x_1, x_2, \ldots, x_n$. Події $X = x_1$, $X = x_2$, … , $X = x_n$, (в результаті випробування випадкова величина набула значення $x_1, x_2, \ldots, x_n$, відповідно) є несумісними та єдино можливими, тобто утворюють повну групу. Позначивши ймовірності цих подій буквами p з відповідними індексами: $P(X = x_1) = p_1$, $P(X = x_2) = p_2$, … , $P(X = x_n) = p_n$, одержимо

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1. \tag{2.1}$$

Найпростішою формою подання закону розподілу дискретної випадкової величини є таблиця, в якій перераховані в порядку зростання усі можливі значення випадкової величини та відповідні їм ймовірності, тобто *X*:

x_1	x_2	…	x_i	…	x_n
p_1	p_2	…	p_i	…	p_n

Така таблиця називається *рядом розподілу* дискретної випадкової величини.

Ряд розподілу можна подати графічно, якщо вздовж осі абсцис відкладати значення випадкової величини, а вздовж осі ординат – відповідні ймовірності. З'єднавши отримані дискретні точки прямолінійними відрізками, одержуємо ламану, яка називається *багатокутником* або *полігоном розподілу ймовірностей* (рис. 2.1).

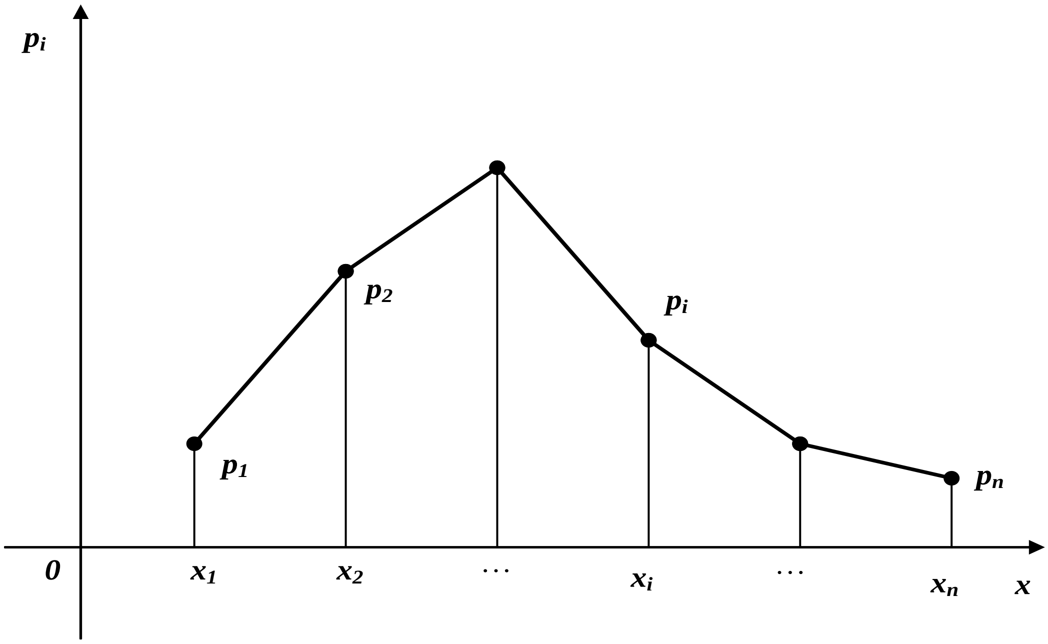


Рисунок 2.1 – Полігон розподілу ймовірностей

Приклад 2.1. Ймовірності того, що студент складе семестровий іспит з "Економічного аналізу" та "Фінансів" під час сесії, дорівнюють відповідно 0,9 та 0,7. Скласти закон розподілу кількості семестрових іспитів, які складе студент, та побудувати полігон цього розподілу.

Розв'язання. Можливі значення випадкової величини X – кількості складених іспитів – 0, 1, 2.

Нехай A_i = "Студент складе i-ий іспит" i = 1, 2. Тоді ймовірності того, що студент складе в сесію 0, 1, 2 іспити відповідно дорівнюють:

$$P(X=0)=P(\overline{A_1}\cdot\overline{A_2})=P(\overline{A_1})\cdot P(\overline{A_2})=(1-0{,}9)(1-0{,}7)=0{,}1\cdot 0{,}3=0{,}03;$$
$$P(X=1)=P(A_1\cdot\overline{A_2}+A_2\cdot\overline{A_1})=P(A_1\cdot\overline{A_2})+P(A_2\cdot\overline{A_1})=0{,}9\cdot 0{,}3+0{,}7\cdot 0{,}1=0{,}34;$$
$$P(X=2)=P(A_1\cdot A_2)=P(A_1)\cdot P(A_2)=0{,}9\cdot 0{,}7=0{,}63.$$

Таким чином, ряд розподілу випадкової величини X:

x_i	0	1	2
p_i	0,03	0,34	0,63

На рисунку 2.2 одержаний ряд розподілу подано графічно у вигляді полігону розподілу ймовірностей.

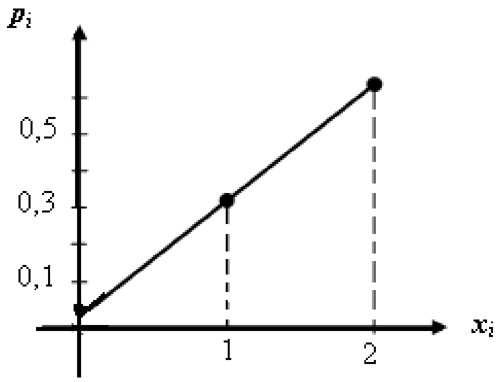


Рисунок 2.2 – Полігон розподілу ймовірностей для прикладу 2.1

Розглянемо найбільш уживані операції над випадковими величинами.

Нехай дано дві дискретні випадкові величини:

X:

x_1	x_2	…	x_n
p_1	p_2	…	p_n

та

Y:

y_1	y_2	…	y_m
p_1	p_2	…	p_m

Добутком kX випадкової величини X на сталу величину k називається випадкова величина, яка набуває значення kx_i з тими ж ймовірностями p_i, $i = 1, 2, \ldots, n$.

m*-им степенем* випадкової величини X, тобто X^m, називається випадкова величина, яка набуває значення x_i^m з тими ж ймовірностями p_i, $i = 1, 2, \ldots, n$.

Сумою (різницею чи *добутком)* випадкових величин X та Y називається випадкова величина, яка набуває усіх можливих значень виду $x_i + y_j$ ($x_i - y_j$ чи $x_i \cdot y_j$), $i = \overline{1,n}$, $j = \overline{1,m}$ із ймовірностями p_{ij} того, що випадкова величина X набуде значення x_i, а Y – y_j:

$$p_{ij} = P[(X = x_i)(Y = y_j)].$$

Якщо випадкові величини X та Y незалежні (закон розподілу однієї величини не залежить від того, які можливі значення набула інша величина), то за теоремою множення ймовірностей для незалежних подій маємо:

$$p_{ij} = P(X = x_i) \cdot P(Y = y_j) = p_i \cdot p_j \,. \qquad (2.2)$$

Приклад 2.2. Дано закони розподілу двох незалежних випадкових величин:

X:

x_i	0	2	4
p_i	0,5	0,2	0,3

та Y:

y_j	-2	0	2
p_j	0,1	0,6	0,3

Знайти закон розподілу випадкової величини $Z = X – Y$.

Розв'язання. Для зручності пошуку всіх значень різниці Z та їх ймовірностей складемо допоміжну таблицю, в кожній клітинці якої розмістимо в лівому кутку значення різниці $Z = X - Y$, а в правому кутку – ймовірності цих значень, одержані в результаті множення ймовірностей відповідних значень випадкових величин X та Y.

x_i \ y_j		-2	0	2
	p_i \ p_j	0,1	0,6	0,3
0	0,5	2 / 0,05	0 / 0,3	-2 / 0,15
2	0,2	4 / 0,02	2 / 0,12	0 / 0,06
4	0,3	6 / 0,03	4 / 0,18	2 / 0,09

Наприклад, якщо X = 2 (передостанній рядок таблиці), а Y = 0 (четвертий стовпець таблиці), то випадкова величина $Z = X – Y$ набуває значення Z = 2 – 0 = 2 з ймовірністю $P(Z = 2) = P(X = 2) \cdot P(Y = 0) = 0{,}2 \cdot 0{,}6 = 0{,}12$.

Оскільки серед дев'яти значень випадкової величини Z є ті, що повторюються, то їх відповідні ймовірності додаємо за теоремою додавання ймовірностей. Наприклад, значення Z = 2 може бути одержане, коли X = 2,

Y = 0 (з ймовірністю 0,12); X = 0, Y = -2 (з ймовірністю 0,05); X = 4, Y = 2 (з ймовірністю 0,09), тому

$$P(Z=2)=0{,}12+0{,}05+0{,}09=0{,}26 \text{ і т. п.}$$

Таким чином отримаємо розподіл:

Z:

z_i	-2	0	2	4	6
p_i	0,15	0,36	0,26	0,20	0,03

Розглянемо таку задачу. Відомі закони розподілу випадкових величин X та Y – кількість очок, що набрали 1-ий та 2-ий стрілець, відповідно.

X:

x_i	0	1	2	3	4	5	6	7	8	9	10
p_i	0,15	0,11	0,04	0,05	0,04	0,1	0,1	0,04	0,05	0,12	0,2

Y:

y_i	0	1	2	3	4	5	6	7	8	9	10
p_i	0,01	0,03	0,05	0,09	0,11	0,24	0,21	0,1	0,1	0,04	0,02

Потрібно з'ясувати, який з двох стрільців стріляє краще. Розглянувши ряди розподілу випадкових величин X та Y, важко відповісти на це запитання через велику кількість значень. Зрозуміло, що з двох стрільців краще стріляє той, хто в *середньому* набирає більшу кількість очок. Таким середнім значенням випадкової величини є її математичне сподівання.

Математичним сподіванням $M(X)$ дискретної випадкової величини X називають суму добутків усіх її значень та відповідних їм ймовірностей:

$$M(X)=\sum_{i=1}^{n} x_i p_i\,. \qquad (2.3)$$

Приклад 2.3. Обчислити $M(X)$ та $M(Y)$ у задачі про стрільців.

Розв'язання. За формулою (2.3) маємо:

$$M(X)=0\cdot 0{,}15+1\cdot 0{,}11+2\cdot 0{,}04+3\cdot 0{,}05+4\cdot 0{,}04+5\cdot 0{,}1+6\cdot 0{,}1+7\cdot 0{,}04+$$
$$+8\cdot 0{,}05+9\cdot 0{,}12+10\cdot 0{,}2=5{,}36;$$

$$M(Y)=0\cdot 0{,}01+1\cdot 0{,}03+2\cdot 0{,}05+3\cdot 0{,}09+4\cdot 0{,}11+5\cdot 0{,}24+6\cdot 0{,}21+7\cdot 0{,}1+$$
$$+8\cdot 0{,}1+9\cdot 0{,}04+10\cdot 0{,}02=5{,}36.$$

Розглянемо основні ***властивості*** математичного сподівання.

1. Математичне сподівання сталої величини є величина стала:

$$M(C) = C, \text{ де } C = const. \tag{2.4}$$

2. Сталий множник можна виносити за знак математичного сподівання, тобто:

$$M(kX) = kM(X), \text{ де } k = const. \tag{2.5}$$

3. Математичне сподівання алгебраїчної суми скінченної кількості випадкових величин дорівнює алгебраїчній сумі математичних сподівань, тобто:

$$M(X \pm Y) = M(X) \pm M(Y). \tag{2.6}$$

4. Математичне сподівання добутку скінченного числа випадкових величин дорівнює добутку їх математичних сподівань:

$$M(XY) = M(X)M(Y). \tag{2.7}$$

5. Якщо всі значення випадкової величини збільшити (зменшити) на сталу C, то на цю ж сталу збільшиться (зменшиться) математичне сподівання цієї випадкової величини:

$$M(X \pm C) = M(X) \pm C. \tag{2.8}$$

6. Математичне сподівання відхилення випадкової величини від її математичного сподівання дорівнює нулю:

$$M\left[X - M(X)\right] = 0. \tag{2.9}$$

Приклад 2.4. Знайти математичне сподівання випадкової величини $Z = 8X - 5Y + 6$, якщо відомо, що $M(X) = 3$, $M(Y) = 4$.

Розв'язання. Використовуючи властивості (1), (2), (3) математичного сподівання знаходимо:

$$M(Z) = 8M(X) - 5M(Y) + 6 = 8 \cdot 3 - 5 \cdot 4 + 6 = 10.$$

Одне математичне сподівання не може достатньою мірою охарактеризувати випадкову величину. В задачі про стрільців ми переконалися, що $M(X) = M(Y) = 5{,}36$, тобто середня кількість очок у обох стрільців однакова. Зрозуміло, що краще стріляє той стрілець, в якого менше відхилення кількості очок від середнього значення.

Дисперсією $D(X)$ випадкової величини X називається математичне сподівання квадрата її відхилення від математичного сподівання:

$$D(X) = M\left[X - M(X)\right]^2. \qquad (2.10)$$

Якщо випадкова величина X – дискретна зі скінченною кількістю значень, то

$$D(X) = \sum_{i=1}^{n} (x_i - M(X))^2 p_i . \qquad (2.11)$$

Із формули (2.11) випливає, що дисперсія має розмірність квадрата, що не завжди зручно. Тому як показник ступеня розсіювання використовують також величину $\sqrt{D(X)}$.

Середнім квадратичним відхиленням σ_x випадкової величини X називається арифметичне значення кореня квадратного з її дисперсії:

$$\sigma_x = \sqrt{D(X)}. \qquad (2.12)$$

Відмітимо основні ***властивості*** дисперсії випадкової величини.

1. Дисперсія сталої величини дорівнює нулю:

$$D(C) = 0, \text{ де } C = const. \qquad (2.13)$$

2. Сталий множник можна виносити за знак дисперсії, підносячи його при цьому до квадрата:

$$D(kX) = k^2 D(X). \qquad (2.14)$$

3. Дисперсія випадкової величини дорівнює різниці між математичним сподіванням квадрата випадкової величини та квадратом її математичного сподівання:

$$D(X) = M(X^2) - [M(X)]^2. \qquad (2.15)$$

Зауваження. Цю властивість досить часто використовують для обчислення дисперсії, оскільки вона дає спрощення розрахунків порівняно з основною формулою (2.10), якщо значення випадкової величини – ціле число, а математичне сподівання – неціле число.

4. Дисперсія алгебраїчної суми скінченного числа незалежних випадкових величин дорівнює сумі їх дисперсій:

$$D(X + Y) = D(X) + D(Y). \qquad (2.16)$$

Приклад 2.5. Знайти дисперсію випадкової величини $Z = 8X + 5Y + 7$, якщо відомо, що випадкові величини X та Y незалежні і $D(X) = 1,5$; $D(Y) = 1$.

Розв'язання. Використовуючи властивості (2.14) та (2.16), знайдемо

$$D(Z) = 8^2 \cdot D(X) + 5^2 \cdot D(Y) + 0 = 64 \cdot 1{,}5 + 25 \cdot 1 = 121.$$

Математичне сподівання, дисперсія, середнє квадратичне відхилення та інші числа, що в стислій формі описують найбільш істотні риси розподілу, називають *числовими характеристиками* випадкової величини.

Досить часто в практичних розрахунках використовують коефіцієнт варіації *CV* для обчислення величини ризику:

$$CV = \frac{\sigma_x}{M(X)}.$$

Приклад 2.6. На фінансовому ринку присутні акції трьох видів (*A, B, C*). Норма прибутку акцій залежить від ринкової кон'юнктури (%). Проаналізувати ситуацію і обрати тип акції, що найбільш приваблива для інвестора за критерієм мінімізації ризику.

Види проектів	Оцінка можливого результату					
	Песимістична		Стримана		Оптимістична	
	Прибуток X_{1i}	Ймовірність p_{1i}	Прибуток X_{2i}	Ймовірність p_{2i}	Прибуток X_{3i}	Ймовірність p_{3i}
A	59	0,25	29	0,53	19	0,22
B	49	0,3	39	0,45	29	0,25
C	39	0,27	29	0,5	19	0,23

Розв'язання. Визначимо сподівану норму прибутку для кожного виду акцій:

$$M(A) = 59 \cdot 0{,}25 + 29 \cdot 0{,}53 + 19 \cdot 0{,}22 = 34{,}3 \ (\%);$$

$$M(B) = 49 \cdot 0{,}3 + 39 \cdot 0{,}45 + 29 \cdot 0{,}25 = 39{,}5 \ (\%);$$

$$M(C) = 39 \cdot 0{,}27 + 29 \cdot 0{,}5 + 19 \cdot 0{,}23 = 29{,}4 \ (\%).$$

Визначимо дисперсію норм прибутку кожного виду акцій за формулою (2.15):

$$D(A) = 59^2 \cdot 0{,}25 + 29^2 \cdot 0{,}53 + 19^2 \cdot 0{,}22 - (34{,}3)^2 = 218{,}91 \quad (\%)^2;$$

$$D(B) = 49^2 \cdot 0{,}3 + 39^2 \cdot 0{,}45 + 29^2 \cdot 0{,}25 - (39{,}5)^2 = 54{,}75\ \ (\%)^2;$$

$$D(C) = 39^2 \cdot 0{,}27 + 29^2 \cdot 0{,}5 + 19^2 \cdot 0{,}23 - (29{,}4)^2 = 49{,}84\ \ (\%)^2.$$

Обчислимо середні квадратичні відхилення від сподіваних норм прибутків кожної акції:

$$\sigma_A = \sqrt{D(A)} = \sqrt{218{,}91} = 14{,}8 \quad (\%);$$

$$\sigma_B = \sqrt{D(B)} = \sqrt{54{,}75} = 7{,}4 \quad (\%);$$

$$\sigma_C = \sqrt{D(C)} = \sqrt{49{,}84} = 7{,}06 \quad (\%).$$

Обчислимо величину ризику для кожного виду акцій:

$$CV_A = \frac{14{,}8}{34{,}3} = 0{,}432; \qquad CV_B = \frac{7{,}4}{39{,}5} = 0{,}187; \qquad CV_C = \frac{7{,}06}{29{,}4} = 0{,}24.$$

Із одержаних результатів зрозуміло, що потрібно вибрати акцію виду B, оскільки в неї ризик є найменшим.

2.2 Функція розподілу та щільність випадкової величини. Неперервні випадкові величини

До цих пір ми розглядали закон розподілу випадкової величини як ряд розподілу або формулу, що дозволяє знаходити ймовірності довільних значень випадкової величини X. Однак такий опис не є універсальним, оскільки його неможливо застосувати до неперервної випадкової величини, яка має нескінченну незчисленну множину можливих значень.

Для опису закону розподілу випадкової величини можна розглядати не ймовірності подій $X = x$ для різних x, а ймовірності події $X < x$, де x – поточна змінна. Зрозуміло, що ймовірність $P(X < x)$ буде деякою функцією від змінної x.

Функцією розподілу випадкової величини X називається ймовірність того, що випадкова величина X набуде значення, меншого за x. Позначають функцію розподілу $F(x)$, тобто

$$F(x) = P(X < x). \tag{2.17}$$

Функцію $F(x)$ іноді називають *інтегральною функцією розподілу* або *інтегральним законом* розподілу.

Приклад 2.7. Дано ряд розподілу випадкової величини *X*:

x_i	1	4	5	7
p_i	0,4	0,1	0,3	0,2

Знайти та графічно зобразити її функцію розподілу.

Розв'язання. Будемо задавати різні значення x та знаходити для них $F(x) = P(X < x)$.

1. Якщо $x \le 1$, то зрозуміло, що $F(x) = P(X < x) = 0$.
2. Якщо $1 < x \le 4$, то $F(x) = P(X = 1) = 0{,}4$. Зрозуміло, що і $F(4) = P(X < 4) = 0{,}4$.
3. Якщо $4 < x \le 5$, то $F(x) = P(X = 1) + P(X = 4) = 0{,}4 + 0{,}1 = 0{,}5$.
4. Якщо $5 < x \le 7$, то $F(x) = [P(X = 1) + P(X = 4)] + P(X = 5) = 0{,}5 + 0{,}3 = 0{,}8$.
5. Якщо $x > 7$, то $F(x) = [P(X = 1) + P(X = 4) + P(X = 5)] + P(X = 7) =$
   $= 0{,}8 + 0{,}2 = 1$.

Маємо:
$$F(x) = \begin{cases} 0; & \textit{якщо} \quad x \le 1; \\ 0{,}4; & \textit{якщо} \quad 1 < x \le 4; \\ 0{,}5; & \textit{якщо} \quad 4 < x \le 5; \\ 0{,}8; & \textit{якщо} \quad 5 < x \le 7; \\ 1; & \textit{якщо} \quad x > 7. \end{cases}$$

Графічно зобразимо функцію $F(x)$ (рис. 2.3).

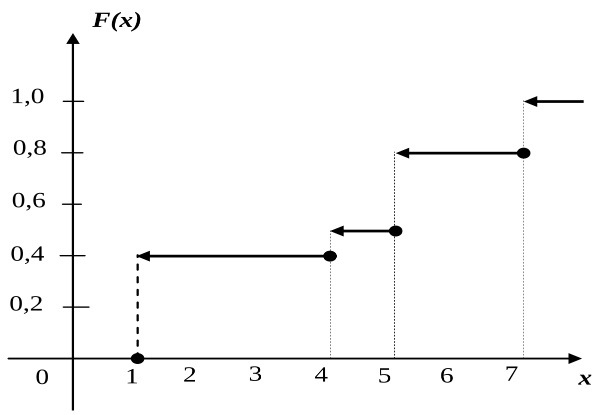


Рисунок 2.3 – Вигляд функції *F*(*x*) із прикладу 2.7

Зауваження. З попереднього прикладу зрозуміло, що функція розподілу *довільної дискретної випадкової величини* є східчастою функцією, стрибки якої відбуваються в точках, що відповідають можливим значенням випадкової величини і дорівнюють ймовірностям цих значень. Сума усіх стрибків функції розподілу дискретної випадкової величини дорівнює 1.

Розглянемо загальні ***властивості*** функції розподілу.

1. Значення функції розподілу належать відрізку [0, 1]. Дане твердження випливає з того, що функція розподілу – це ймовірність.

2. Функція розподілу є неспадною на всій числовій осі.

3. На мінус нескінченності функція розподілу дорівнює нулю, на плюс нескінченності дорівнює одиниці, тобто

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,\ F(+\infty)=\lim_{x\to+\infty}F(x)=1.$$

4. Ймовірність потрапляння випадкової величини на проміжок $[x_1, x_2)$ дорівнює приросту її функції розподілу на цьому інтервалі, тобто

$$P(x_1\le X<x_2)=F(x_2)-F(x_1). \qquad (2.18)$$

Враховуючи розглянуте поняття функції розподілу, *неперервною випадковою величиною* називають випадкову величину, функція розподілу якої неперервна та диференційовна в усіх точках. Можна довести, що ймовірність будь-якого окремо взятого значення неперервної випадкової величини дорівнює нулю. Такий результат означає, що нульову ймовірність можуть мати і можливі події.

Наслідок. Якщо X – неперервна випадкова величина, то ймовірність потрапляння випадкової величини у проміжок (x_1, x_2) не залежить від того, є цей проміжок відкритим чи закритим, тобто

$$P(x_1<X<x_2)=P(x_1\le X<x_2)=P(x_1<X\le x_2)=P(x_1\le X\le x_2).$$

Подання неперервної випадкової величини за допомогою функції розподілу не є єдиним. Введемо поняття *щільності ймовірності* неперервної випадкової величини.

Щільністю ймовірності (або просто *щільністю*) $f(x)$ неперервної випадкової величини X називається похідна її функції розподілу

$$F'(x)=f(x). \qquad (2.19)$$

Про випадкову величину кажуть, що вона розподілена зі щільністю $f(x)$ на певному проміжку осі абсцис. Функція $f(x)$ є однією з форм закону розподілу, але існує вона лише для ***неперервних*** випадкових величин. Щільність ймовірності іноді називають *диференціальним законом розподілу.* Графік щільності ймовірності $f(x)$ називають *кривою розподілу.*

Приклад 2.8. Знайти щільність розподілу ймовірності випадкової величини X, заданої функцією розподілу

$$F(x)=\begin{cases}0, & x\le 0\\ 2\sin x, & 0<x\le\dfrac{\pi}{6}.\\ 1, & x>\dfrac{\pi}{6}\end{cases}$$

Розв'язання. За формулою (2.19) маємо

$$f(x)=F'(x)=\begin{cases}0, & x\le 0\ \ i\ \ x>\dfrac{\pi}{6},\\ 2\cos x, & 0<x\le\dfrac{\pi}{6}.\end{cases}$$

Розглянемо загальні ***властивості*** щільності ймовірності неперервної випадкової величини.

1. Щільність ймовірності – невід'ємна функція, тобто $f(x)\ge 0$.

2. Ймовірність потрапляння неперервної випадкової величини у проміжок $[a,b]$ дорівнює визначеному інтегралу від її щільності ймовірності в межах від a до b, тобто

$$P(a\le X\le b)=\int_a^b f(x)dx. \tag{2.20}$$

Із геометричної точки зору одержана ймовірність дорівнює площі криволінійної трапеції, обмеженої кривою розподілу, віссю *Ox* та прямими $x=a$ та $x=b$.

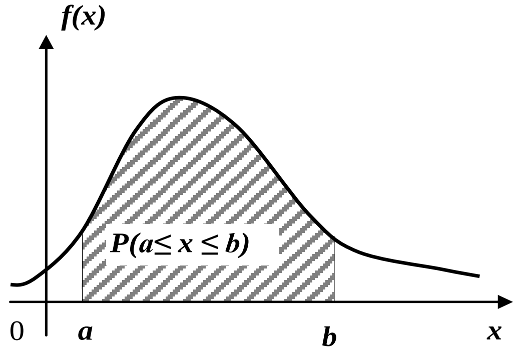


Рисунок 2.4 – Графічна інтерпретація ймовірності потрапляння неперервної випадкової величини на проміжок $[a,b]$

3. Функція розподілу неперервної випадкової величини знаходиться через щільність ймовірності за формулою:

$$F(x) = \int_{-\infty}^{x} f(x)dx. \qquad (2.21)$$

Із геометричної точки зору функція розподілу дорівнює площі фігури, обмеженої зверху кривою розподілу та розташованої лівіше точки x (рис. 2.5).

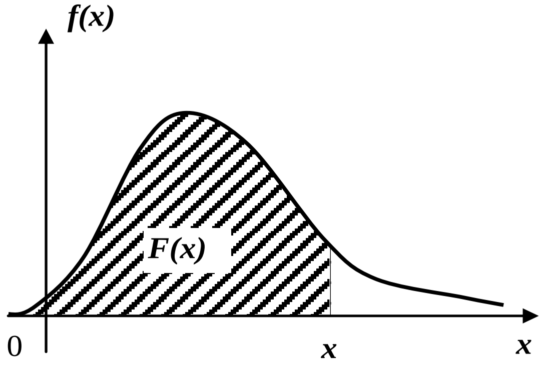


Рисунок 2.5 – Геометрична інтерпретація функції розподілу

4. Невласний інтеграл у нескінченних межах від щільності ймовірності неперервної випадкової величини дорівнює одиниці:

$$\int_{-\infty}^{+\infty} f(x)dx = 1. \qquad (2.22)$$

Поняття математичного сподівання та дисперсії, розглянуті для дискретної випадкової величини, можна поширити на неперервні випадкові величини. Для одержання відповідних формул достатньо у формулах (2.3) та (2.11) для дискретної випадкової величини замінити знак підсумовування $\sum_{i=1}^{n}$ знаком інтеграла з нескінченними межами $\int_{-\infty}^{+\infty}$, можливі значення x_i – неперервною змінною x, а ймовірність p_i – елементом ймовірності $f(x)dx$.

Зауваження. Під *елементом ймовірності* розуміють ймовірність потрапляння випадкової величини X у проміжок $[x, \quad x + dx]$.

У результаті отримаємо такі формули для математичного сподівання та дисперсії неперервної випадкової величини X:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx, \qquad (2.23)$$

(якщо інтеграл абсолютно збіжний) та

$$D(X)=\int_{-\infty}^{+\infty}(x-M(X))^2 f(x)dx, \tag{2.24}$$

(якщо інтеграл збіжний).

Усі властивості математичного сподівання та дисперсії, розглянуті для дискретних величин, справедливі і для неперервних. Зокрема, на практиці при обчисленні дисперсії використовують формулу:

$$D(X)=\int_{-\infty}^{+\infty}x^2 f(x)dx-\left(M(X)\right)^2. \tag{2.25}$$

Приклад 2.9. Дано функцію:

$$f(x)=\begin{cases}0, & якщо\ \ x\le 1\\ \dfrac{a}{x^4}, & якщо\ \ x>1\end{cases}.$$

Знайти: а) значення сталої a, при якому дана функція буде щільністю ймовірності деякої випадкової величини X; б) вираз для функції розподілу $F(x)$; в) обчислити ймовірність того, що випадкова величина X набуде значення з відрізку $[5,6]$; г) математичне сподівання та дисперсію випадкової величини X.

Розв’язання.

а) Для того, щоб дана функція була щільністю ймовірності неперервної випадкової величини, вона повинна бути невід’ємною, тобто $\frac{a}{x^4}\ge 0$ $(a\ge 0)$ і $\int_{-\infty}^{+\infty}f(x)dx=1$. Таким чином,

$$\int_{-\infty}^{+\infty}f(x)dx=\int_{-\infty}^{1}0\cdot dx+\int_{1}^{+\infty}\frac{a}{x^4}dx=\lim_{B\to+\infty}\int_{1}^{B}\frac{a}{x^4}dx=\frac{a}{3}\lim_{B\to+\infty}\left(-\frac{1}{x^3}\right)\Bigg|_1^B=$$

$$=\frac{a}{3}\lim_{B\to+\infty}\left(1-\frac{1}{B^3}\right)=\frac{a}{3}=1,\ \text{звідки}\ a=3;$$

б) за формулою (2.21) знайдемо $F(x)$.

Якщо $x\le 1$, то $F(x)=\int_{-\infty}^{x}f(x)dx=\int_{-\infty}^{x}0\cdot dx=0$.

Якщо $x > 1$, то $F(x) = 0 + \int\limits_1^x f(x)dx = \int\limits_1^x \frac{3}{x^4} \cdot dx = -\frac{1}{x^3}\Big|_1^x = 1 - \frac{1}{x^3}$.

Таким чином,

$$F(x) = \begin{cases} 0, & \textit{якщо} \quad x \le 1; \\ 1 - \frac{1}{x^3}, & \textit{якщо} \quad x > 1; \end{cases}$$

в) за формулою (2.20) маємо

$$P(5 \le X \le 6) = \int\limits_5^6 \frac{3}{x^4} dx = -\frac{1}{x^3}\Big|_5^6 = \frac{1}{5^3} - \frac{1}{6^3} = \frac{91}{27000} \approx 0{,}0034.$$

Ймовірність $P(5 \le X \le 6)$ можна знайти за формулою (2.18):

$$P(5 \le X \le 6) = F(6) - F(5) = \left(1 - \frac{1}{6^3}\right) - \left(1 - \frac{1}{5^3}\right) = \frac{91}{27000} \approx 0{,}0034;$$

г) за формулою (2.23) маємо

$$M(X) = \int\limits_{-\infty}^{+\infty} xf(x)dx = \int\limits_1^{+\infty} x \cdot \frac{3}{x^4} dx = 3 \lim_{B \to +\infty} \int\limits_1^B \frac{1}{x^3} dx = \frac{3}{2} \lim_{B \to +\infty} -\frac{1}{x^2}\Big|_1^B =$$

$$= \frac{3}{2} \lim_{B \to +\infty} \left(1 - \frac{1}{B^2}\right) = \frac{3}{2}.$$

Дисперсію $D(x)$ обчислимо за формулою (2.25). Для цього спочатку знайдемо

$$M(X^2) = \int\limits_{-\infty}^{+\infty} x^2 f(x)dx = \int\limits_1^{+\infty} x^2 \cdot \frac{3}{x^4} dx = 3 \lim_{B \to +\infty} \int\limits_1^B \frac{1}{x^2} dx = 3 \lim_{B \to +\infty} \left(-\frac{1}{x}\right)\Big|_1^B = 3.$$

Тоді $D(X) = \int\limits_{-\infty}^{+\infty} x^2 f(x)dx - \left(M(X)\right)^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}.$

2.3 Деякі розподіли випадкових величин

Закон розподілу Пуассона. Дискретна випадкова величина X має закон розподілу Пуассона з параметром $\lambda > 0$, якщо вона набуває значень 0, 1, 2, … , m, … (нескінченна, але зчисленна множина) з ймовірностями

$$P(X = m) = \frac{\lambda^m \cdot e^{-\lambda}}{m!} = P_m(\lambda). \qquad (2.26)$$

Ряд розподілу закону Пуассона такий:

x_i	0	1	2	…	m	…
p_i	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!}e^{-\lambda}$	…	$\frac{\lambda^m}{m!}e^{-\lambda}$	…

Математичне сподівання та дисперсія випадкової величини, розподіленої за законом Пуассона, збігаються і дорівнюють параметру його закону, тобто

$$M(X) = D(X) = \lambda\,. \qquad (2.27)$$

Оскільки ймовірність події A в кожному випробуванні мала, то закон розподілу Пуассона часто називають *законом рідкісних явищ.*

Рівномірний закон розподілу. Неперервна випадкова величина X має рівномірний закон розподілу на відрізку $[a, b]$, якщо щільність її ймовірності така:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{якщо } a \le x \le b, \\ 0, & \text{якщо } x < a, \quad x > b, \end{cases} \qquad (2.28)$$

а функція розподілу визначається за формулою:

$$F(x) = \begin{cases} 0, & \text{якщо } x \le a, \\ {(x-a)}\big/{(b-a)}, & \text{якщо } a < x \le b, \\ 1, & \text{якщо } x > b. \end{cases} \qquad (2.29)$$

Криву розподілу та графік функції розподілу випадкової величини X наведено на рис. 2.6, а, б.

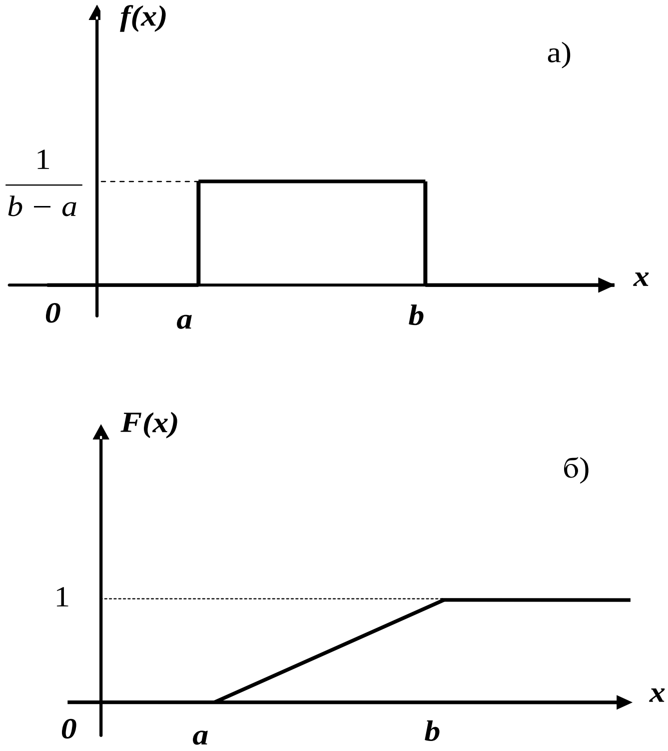


Рисунок 2.6 – Графік функції рівномірного розподілу

Якщо випадкова величина розподілена за рівномірним законом, то її математичне сподівання

$$M(X) = \frac{a+b}{2}, \tag{2.30}$$

а дисперсія

$$D(X) = \frac{(b-a)^2}{12}. \tag{2.31}$$

Рівномірний закон розподілу використовують при аналізі помилок округлення при проведенні числових обрахунків, в задачах масового обслуговування, при статистичному моделюванні спостережень.

Показниковий закон розподілу. Неперервна випадкова величина X має показниковий закон розподілу з параметром λ, якщо її щільність ймовірності така:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{якщо } x \geq 0, \\ 0, & \text{якщо } x < 0, \end{cases} \tag{2. 32}$$

а функція розподілу визначається за формулою:

$$F(x)=\begin{cases}0, & \text{якщо} \quad x<0,\\ 1-e^{-\lambda x}, & \text{якщо} \quad x\geq 0.\end{cases} \qquad (2.\ 33)$$

Криву розподілу та графік функції розподілу випадкової величини *X*, розподіленої за показниковим законом, наведено на рис. 2.7, а, б.

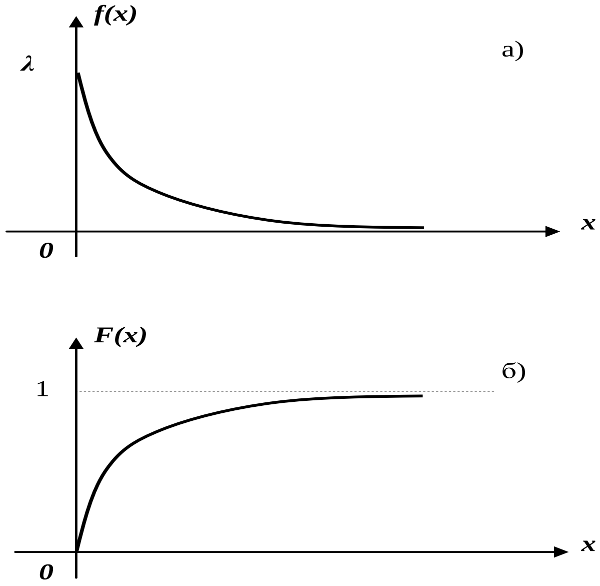


Рисунок 2.7 – Крива розподілу а) та графік функції розподілу випадкової величини X б)

Якщо випадкова величина розподілена за показниковим законом, то її математичне сподівання

$$M(X)=\frac{1}{\lambda}, \qquad (2.34)$$

а дисперсія

$$D(X)=\frac{1}{\lambda^2}. \qquad (2.35)$$

Показниковий закон відіграє важливу роль в теорії масового обслуговування та теорії надійності. Зокрема, інтервал часу T між двома сусідніми подіями в елементарному потоці має показниковий розподіл з параметром λ – інтенсивністю потоку.

Приклад 2.10. Довести, що якщо проміжок часу T, розподілений за показниковим законом, вже тривав деякий час τ, то це ніяк не впливає на закон розподілу проміжку $T_1=T-\tau$.

Розв'язання. Нехай функція розподілу проміжку часу T визначається за формулою (2.21), тобто $F(t)=1-e^{-\lambda t}$. Функція розподілу частини, що залишилася ($T_1=T-\tau$), за умови, що подія $T>\tau$ відбулась, є умовна ймовірність події $T_1<t$ відносно події $T>\tau$, тобто $F_1(t)=P_{T>\tau}(T_1>t)$.

Оскільки умовна ймовірність довільної події B відносно події A визначається за формулою $P_A(B)=\dfrac{P(AB)}{P(A)}$, то, приймаючи $A=(T>\tau)$, $B=(T_1<t)$, отримаємо

$$F_1(t)=P_{T>\tau}(T_1>t)=\frac{P\left[(T>\tau)(T_1<t)\right]}{P(T>\tau)}.$$

Добуток подій $T>\tau$ та $T_1=(T-\tau)<t$ рівносильний події $\tau<T<t+\tau$, ймовірність якої

$$P(\tau<T<t+\tau)=F(t+\tau)-F(\tau).$$

Оскільки $P(T>\tau)=1-P(T\le\tau)=1-F(\tau)$, то $F_1(t)=P_{T>\tau}(T_1>t)$ можна подати у вигляді:

$$F_1(t)=\frac{F(t+\tau)-F(\tau)}{1-F(\tau)}.$$

Враховуючи рівність (2.21), отримуємо

$$F_1(t)=\frac{e^{-\lambda\tau}-e^{-\lambda(t+\tau)}}{e^{-\lambda\tau}}=1-e^{-\lambda t}=F(t).$$

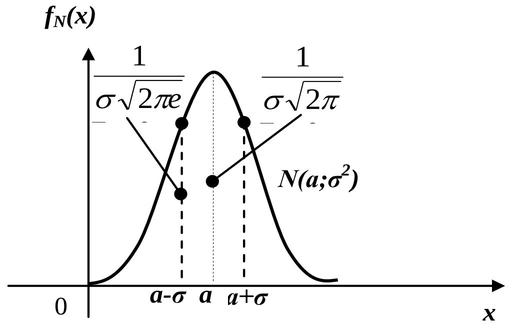
Зауваження. Доведена властивість широко використовується у марковських випадкових процесах.

Нормальний закон розподілу. Даний закон найчастіше застосовується на практиці, оскільки він є граничним законом, до якого наближаються інші закони.

Неперервна випадкова величина X має *нормальний закон розподілу* (*normal law of distribution*) (закон Гаусса) з параметрами a та σ^2, якщо її щільність ймовірності така:

$$f_N(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (2.36)$$

Криву нормального закону розподілу називають *нормальною* або *гаусовою* кривою. На рис. 2.8 наведено нормальну криву $f_N(x)$ з параметрами a та σ^2, тобто $N(a,\sigma^2)$, та графік функції розподілу випадкової величини *X*, що має нормальний закон.



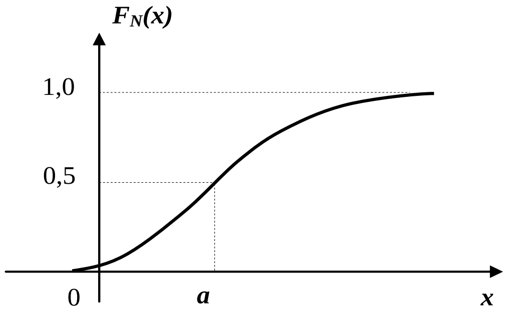


Рисунок 2.8 – Нормальна крива та графік функції розподілу випадкової величини X

Нормальна крива симетрична відносно прямої $x = a$. Значення $x = a$ є точкою максимуму, причому $f_{\max} = f(a) = \dfrac{1}{\sigma\sqrt{2\pi}}$. Існують дві точки перегину $x = a \pm \sigma$ з ординатою $f_{\text{пер}}(a \pm \sigma) = \dfrac{1}{\sigma\sqrt{2\pi e}} \approx \dfrac{0{,}242}{\sigma}$.

Математичне сподівання випадкової величини X, розподіленої за нормальним законом, дорівнює параметру a цього закону, тобто

$$M(X) = a, \tag{2.37}$$

а її дисперсія – параметру σ^2, тобто

$$D(X) = \sigma^2. \tag{2.38}$$

З'ясуємо, як буде змінюватися нормальна крива при зміні параметрів a та σ^2 (або σ). Якщо $\sigma = const$, а змінюється параметр, тобто центр симетрії, то нормальна крива буде зміщуватися вздовж осі абсцис, не змінюючи форми (рис. 2.9).

Якщо $a = const$, а змінюється параметр σ^2, то змінюється ордината максимуму кривої $f_{\max} = \dfrac{1}{\sigma\sqrt{2\pi}}$. При збільшенні σ ордината максимуму кривої зменшується, але оскільки площа під довільною кривою розподілу повинна залишатися рівною одиниці, то крива стає більш пласкою, розтягуючись вздовж осі абсцис; при зменшенні σ нормальна крива витягається вверх, одночасно стискаючись з боків. На рис. 2.10 зображено нормальні криві з параметрами $\sigma_1 < \sigma_2 < \sigma_3$.

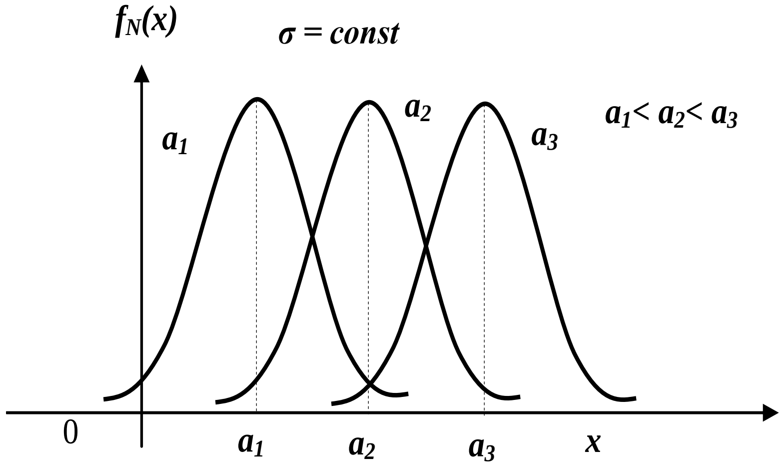


Рисунок 2.9 – Зсув нормальної кривої

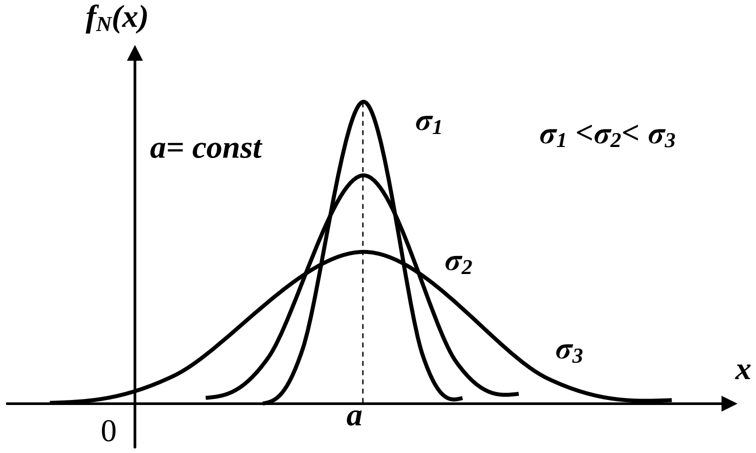


Рисунок 2.10 – Нормальні криві з параметрами $\sigma_1 < \sigma_2 < \sigma_3$

Таким чином, параметр a характеризує положення центра, а параметр σ^2 – форму нормальної кривої.

Нормальний закон з параметрами $a = 0$ та $\sigma^2 = 1$, тобто $N(0,1)$, називають стандартним або нормованим, а відповідну нормальну криву – стандартною або нормованою.

Функцію розподілу випадкової величини X, розподіленої за нормальним законом, можна виразити за допомогою функції Лапласа $\Phi(x)$ так:

$$F_N(x) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right). \tag{2.39}$$

Розглянемо основні ***властивості*** нормально розподіленої випадкової величини.

1. Ймовірність потрапляння випадкової величини X, розподіленої за нормальним законом, в інтервал $[x_1, x_2]$ дорівнює

$$P(x_1 \le X \le x_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)], \tag{2.40}$$

де

$$t_1 = \frac{x_1 - a}{\sigma}, \qquad t_2 = \frac{x_2 - a}{\sigma}.$$

2. Ймовірність того, що відхилення випадкової величини X, розподіленої за нормальним законом, від математичного сподівання не перевищить величину $\Delta > 0$ (за абсолютною величиною) дорівнює

$$P(|X - a| \le \Delta) = \Phi(t), \tag{2.41}$$

де

$$t = \frac{\Delta}{\sigma}. \tag{2.42}$$

На рис. 2.11, а) та 2.11, б) наведено геометричну інтерпретацію властивостей нормального закону.

Обчислимо за формулою (2.41) ймовірності $P(|X - a| \le \Delta)$ при різних значеннях Δ. Маємо при

$\Delta = \sigma \quad P(|X - a| \le \sigma) = \Phi(1) = 0{,}6827$ (додаток А);

$\Delta = 2\sigma \quad P(|X - a| \le 2\sigma) = \Phi(2) = 0{,}9545$;

$\Delta = 3\sigma \quad P(|X - a| \le 3\sigma) = \Phi(3) = 0{,}9973$, (рис. 2.12).

Обчислимо за формулою (2.41) ймовірності $P(|X - a| \le \Delta)$ при різних значеннях Δ. Маємо при

$\Delta = \sigma \quad P(|X - a| \le \sigma) = \Phi(1) = 0{,}6827$;

$\Delta = 2\sigma \quad P(|X - a| \le 2\sigma) = \Phi(2) = 0{,}9545$;

$\Delta = 3\sigma \quad P(|X - a| \le 3\sigma) = \Phi(3) = 0{,}9973$, (рис. 2.12).

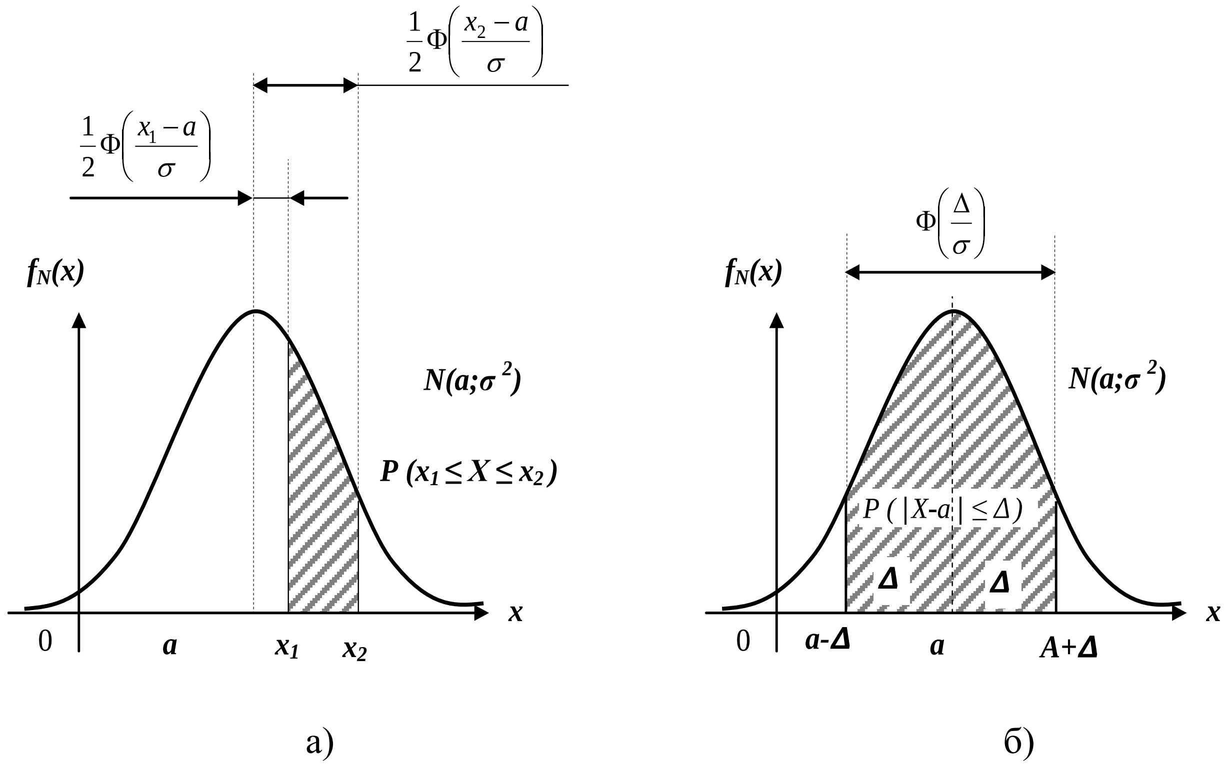


Рисунок 2.11– Геометрична інтерпретація властивостей нормального закону

Звідси випливає ***“правило трьох сигм”***: *якщо випадкова величина* X *має нормальний закон розподілу* з *параметрами* a *та* σ^2*, то практично достовірно, що її значення належать інтервалу* $(a-3\sigma, \quad a+3\sigma)$. Порушення "правила трьох сигм", тобто відхилення випадкової величини більше ніж на 3σ є подією практично неможливою, оскільки її ймовірність досить мала:

$$P(|X-a|>3\sigma)=1-P(|X-a|\le 3\sigma)=1-0{,}9973=0{,}0027.$$

Приклад 2.11. Припускаючи, що зріст чоловіків певної вікової групи є нормально розподілена випадкова величина з параметрами $a=173$ та $\sigma^2=36$, потрібно:

1) визначити щільність ймовірності та функцію розподілу;

2) знайти частки костюмів 4-го зросту (176 − 182 см) та 3-го зросту (170 − 176 см), котрі потрібно передбачити в загальному обсязі виробництва для даної вікової групи;

3) сформулювати “правило трьох сигм” для випадкової величини X.

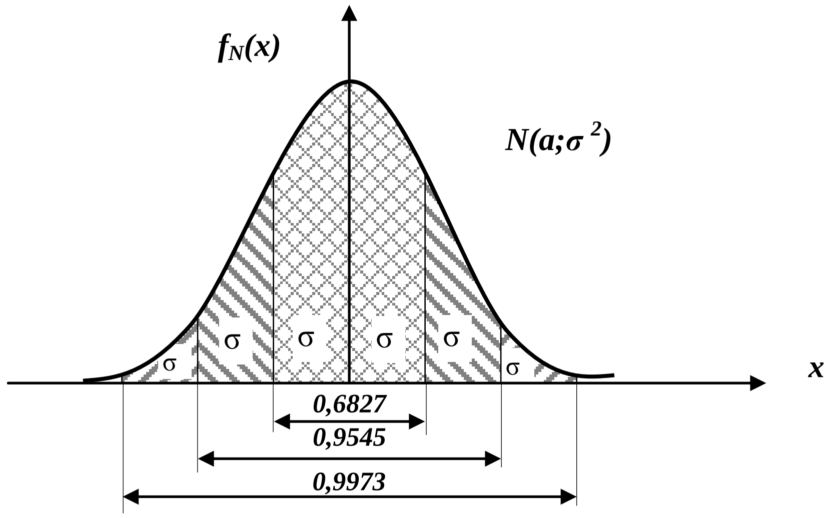


Рисунок 2.12 – Ймовірності при різних значеннях Δ

Розв'язання.

1. За формулами (2.36) та (2.39) запишемо:

$$f_N(x) = \frac{1}{6\sqrt{2\pi}} e^{-\frac{(x-173)^2}{2\cdot 36}}; \qquad F_N(x) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x-173}{6}\right).$$

2. Частку костюмів 4-го зросту в загальному обсязі виробництва обчислимо за формулою (2.40) так:

$$P(176 \le X \le 182) = \frac{1}{2}\left[\Phi\left(\frac{182-173}{6}\right) - \Phi\left(\frac{176-173}{6}\right)\right] = \frac{1}{2}[\Phi(1{,}5) - \Phi(0{,}5)] =$$

$$= \frac{1}{2}(0{,}8664 - 0{,}3829) = 0{,}2418;$$

аналогічно обчислюємо частку костюмів 3-го зросту:

$$P(170 \le X \le 176) = \frac{1}{2}\left[\Phi\left(\frac{176-173}{6}\right) - \Phi\left(\frac{170-173}{6}\right)\right] = \frac{1}{2}[\Phi(0{,}5) + \Phi(0{,}5)] = \Phi(0{,}5) = 0{,}3829.$$

3. Практично достовірно, що зріст чоловіків даної вікової групи знаходиться в межах від $a - 3\sigma = 173 - 18 = 155$ до $a + 3\sigma = 173 + 18 =$ $= 191$ (см), тобто $155 \le X \le 191$ (см).

Розподіл χ^2. Розподілом χ^2 із k *ступенями вільності* (додаток Б) називають розподіл суми квадратів k незалежних випадкових величин, розподілених за стандартним нормальним законом, тобто

$$\chi^2 = \sum_{i=1}^{k} Z_i^2, \tag{2.43}$$

де Z_i $\left(i = 1, 2, ..., k\right)$ має нормальний розподіл $N(0,1)$.

Щільність ймовірності χ^2-розподілу така:

$$\varphi(x) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}} \cdot \Gamma\left(\dfrac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, & \text{якщо} \quad x \geq 0, \\ 0, \quad \text{якщо} & x < 0, \end{cases}$$

де $\Gamma(y) = \int\limits_0^{+\infty} e^{-t} t^{y-1} dt$ – гамма-функція Ейлера (для цілих додатних значень $\Gamma(y) = (y-1)!$).

Криві χ^2-розподілу для різних значень ступенів вільності наведено на рис. 2.13.

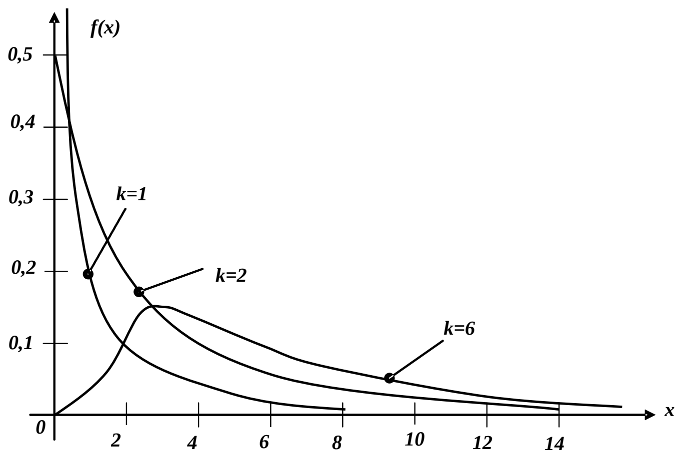


Рисунок 2.13 – Криві χ^2-розподілу

Очевидно, що χ^2-розподіл асиметричний з правосторонньою (додатною) асиметрією. При $k > 30$ розподіл випадкової величини $Z = \sqrt{2\chi^2} - \sqrt{2k-1}$ близький до стандартного нормального закону $N(0,1)$.

Розподіл Стьюдента. *Розподілом Стьюдента (distribution of Student),* або *t-розподілом,* (додаток В) називають розподіл випадкової величини

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}}, \tag{2.44}$$

де Z має нормальний розподіл $N(0,1)$,

χ^2 – незалежна від Z випадкова величина, що має χ^2-розподіл з k ступенями вільності.

Щільність ймовірності Стьюдента така:

$$\varphi(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{x^2}{n}\right)^{-\frac{k+1}{2}},$$

де $\Gamma(y) = \int\limits_0^{+\infty} e^{-t} t^{y-1}$ – гамма-функція Ейлера.

На рис. 2.14 наведено криву розподілу Стьюдента. Як і стандартна нормальна крива, крива t-розподілу симетрична відносно осі ординат, але, порівняно з нормальною, є більш пласкою.

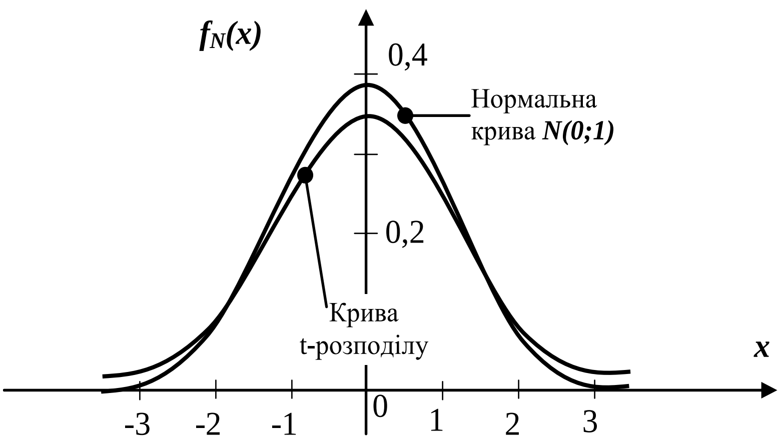


Рисунок 2.14 – Крива розподілу Стьюдента

Практично при $k > 30$ розподіл випадкової величини близький до стандартного нормального закону $N(0,1)$.

Математичне сподівання випадкової величини, що має розподіл Стьюдента, дорівнює нулю (оскільки крива розподілу симетрична), а дисперсія – $D(t) = \dfrac{k}{k-2}$.

Зауваження! Значення функції Стьюдента є табульованими $t_p = t_{\alpha,\nu}$ (додаток Г), причому $\alpha = 1 - p$, ймовірність p задається залежно від задачі, $\nu = n-2$, де n – число дослідів.

Розподіл Фішера-Снедекора. *Розподілом Фішера-Снедекора* або F-*розподілом* (додаток Д) називають розподіл випадкової величини

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}, \qquad (2.45)$$

де $\chi^2(k_1)$ та $\chi^2(k_2)$ – випадкові величини, що мають χ^2-розподіл відповідно з k_1 та k_2 ступенями вільності.

На рис. 2.15 зображено криві F-розподілу при деяких значеннях числа ступенів вільності k_1 та k_2.

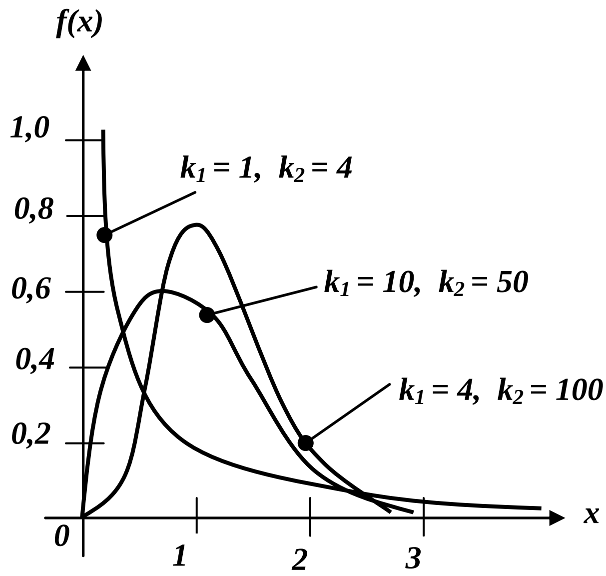


Рисунок 2.15 – Криві F-розподілу

Щільність ймовірності F-розподілу є такою:

$$\varphi(x) = \frac{\Gamma\left(\frac{k_1 + k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} k_1^{\frac{k_1}{2}} k_2^{\frac{k_2}{2}} x^{\frac{k_1}{2}-1} \left(k_1 x + k_2\right)^{-\frac{k_1+k_2}{2}}.$$

При $n \to \infty$ F-розподіл наближається до нормального закону.

2.4 Статистична гіпотеза та загальна схема її перевірки

Із теорією статистичного оцінювання параметрів тісно пов’язана перевірка статистичних гіпотез. Вона використовується щоразу, коли потрібно обґрунтувати висновок про переваги того чи іншого способу інвестування, вимірювання, стрільби, технологічного процесу, стосовно ефективності нового навчального методу, управління; довести значимість математичної моделі тощо.

Статистичною гіпотезою називають довільне припущення про вигляд чи параметри закону розподілу. Розрізняють *просту* та *складену* статистичні гіпотези. Проста гіпотеза повністю визначає теоретичну функцію розподілу випадкової величини. Ту гіпотезу, яку перевіряють, називають *нуль-гіпотезою* (або *основною гіпотезою*) та позначають H_0. Одночасно з нуль-гіпотезою розглядають *альтернативну*, або *конкуруючу*, гіпотезу H_1, яка є логічним запереченням гіпотези H_0.

Суть перевірки статистичної гіпотези полягає в тому, що використовується спеціально складена вибіркова характеристика (статистика) $\tilde{\theta}_n(x_1,....,x_n)$, отримана за вибіркою $X_1,..., X_n$, точний чи наближений розподіл якої відомий. Потім за цим вибірковим розподілом визначають критичне значення $\theta_{кр}$. Якщо ймовірність $P\left(\tilde{\theta}_n > \theta_{кр}\right) = \alpha$ мала, то гіпотеза H_0 є правильною. Правило, за яким гіпотеза H_0 приймається чи відкидається, називається статистичним критерієм чи статистичним тестом.

Таким чином, множину можливих значень статистики критерію $\tilde{\theta}_n$ розбивають на дві підмножини, які не перетинаються: критичну область (область відхилення гіпотези) W та область припустимих значень (область прийняття гіпотези) $\overline{W}$. Якщо спостережуване значення статистики $\tilde{\theta}_n$ потрапляє в критичну область W, то гіпотезу H_0 відкидають. При цьому можливі чотири випадки (табл. 2.1).

Ймовірність α припуститися помилки 1-го роду, тобто відкинути гіпотезу H_0, коли вона правильна, називається *рівнем значимості* чи *розміром критерію*.

Таблиця 2.1 – Прийняття рішення за різних рішень щодо гіпотези H_0

Гіпотеза H_0	*Приймається*	*Відкидається*
Правильна	Правильне рішення	Помилка 1-го роду
Неправильна	Помилка 2-го роду	Правильне рішення

Ймовірність припуститися помилки 2-го роду, тобто прийняти хибну гіпотезу, зазвичай позначають β.

Ймовірність (1 - β) не припуститися помилки 2-го роду, тобто відхилити гіпотезу H_0, коли вона хибна, називають *потужністю критерію*.

Користуючись термінологією статистичного контролю якості продукції можна сказати, що ймовірність α – “ризик постачальника”, пов’язаний із забракуванням усієї партії за результатами вибіркового контролю, а ймовірність β – "ризик споживача", пов'язаний із прийняттям за аналізом партії, що не задовольняє стандарт.

Застосовуючи юридичну термінологію, α – ймовірність винесення судом обвинувального вироку, коли насправді підсудний є невинним, β – ймовірність винесення судом виправдувального вироку у випадку реального скоєння злочину підсудним. У прикладних дослідженнях помилка 1-го роду означає ймовірність того, що сигнал не буде прийнято спостерігачем, а помилка 2-го роду – ймовірність того, що спостерігач прийме хибний сигнал.

Слід відмітити, що при практичних розрахунках зазвичай обирають таку критичну область, при якій потужність критерію є максимальною, а ймовірність потрапляння в неї статистики критерію $\tilde{\theta}_n$ буде мінімальною та рівною α, у випадку справедливості нуль-гіпотези, і максимальною в іншому випадку.

Іншими словами, критична область повинна бути такою, щоб при заданому рівні значимості α потужність критерію (1 - β) була максимальною.

Серед усіх критеріїв заданого рівня значимості α, що перевіряють гіпотезу H_0, критерій відношення правдоподібності є найбільш потужним (за Нейманом-Пірсоном). Основу методу максимальної правдоподібності складає функція правдоподібності, яка є щільністю ймовірності одночасної появи результатів вибірки $x_1, x_2, ..., x_n$:

$$L(x_1, x_2, ..., x_n; \theta) = \varphi(x_1, \theta) \cdot \varphi(x_2, \theta) \cdot ... \cdot \varphi(x_i, \theta) \cdot ... \cdot \varphi(x_n, \theta)$$

або

$$L(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} \varphi(x_i, \theta). \qquad (2.46)$$

Згідно з цим методом за оцінку параметра θ приймається таке значення $\tilde{\theta}_n$, яке максимізує функцію L.

Приклад 2.12. Випадкова величина має нормальний закон розподілу $N(a; \sigma^2)$, де $a = M(X)$ невідоме, а дисперсія – відома. Побудувати найбільш потужний критерій перевірки гіпотези H_0: $a = a_0$ на противагу альтернативній гіпотезі H_1: $a = a_1 > a_0$. Знайти: а) потужність критерію; б) мінімальний обсяг вибірки, що забезпечить задані рівень значимості α та потужність критерію (1 - β).

Розв'язання. Якщо правильною є гіпотеза H_0, тобто $X \approx N(a_0; \sigma^2)$, то функція правдоподібності, згідно з формулою (2.46), така:

$$L_0(x_1, x_2, ..., x_n) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^{n}(x_i - a_0)^2}{2\sigma^2}}.$$

Аналогічно, якщо правильною є гіпотеза H_1, тобто $X \approx N(a_1; \sigma^2)$, то

$$L_1(x_1, x_2, ..., x_n) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^{n}(x_i - a_1)^2}{2\sigma^2}}.$$

Найбільш потужний критерій базується на відношенні правдоподібності $\frac{L_1}{L_0}$. Знайдемо його натуральний логарифм:

$$\ln \frac{L_1}{L_0} = -\frac{\sum_{i=1}^{n}(x_i - a_1)^2}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i - a_0)^2}{2\sigma^2} =$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left[2x_i (a_1 - a_0) - \left(a_1^2 - a_0^2\right) \right] = \frac{1}{2\sigma^2} (a_1 - a_0) \sum_{i=1}^{n} (2x_i - a_1 - a_0) =$$

$$= \frac{1}{2\sigma^2} (a_1 - a_0) \left(2\bar{x} - a_1 - a_0\right) n.$$

Для побудови критерію знайдемо таку сталу C (або $\ln C = c$), що

$$P\left(\frac{L_1}{L_0} > C\right) = P\left(\ln \frac{L_1}{L_0} > c\right) = \alpha.$$

Отриманий вираз для рівня значимості α можна замінити рівносильним (враховуючи монотонність функції $\ln \frac{L_1}{L_0}$ відносно $\bar{x}$):

$$P(\bar{x} > c') = \alpha.$$

Для визначення c' потрібно врахувати, що якщо випадкова величина розподілена нормально, тобто $X \approx N(a_0;\sigma^2)$, то її середня $\bar{x}$ також розподілена нормально з параметрами a_0 та $\frac{\sigma^2}{\sqrt{n}}$, тобто $\bar{x} \approx N\left(a_0;\frac{\sigma^2}{\sqrt{n}}\right)$.

Отримуємо:

$$P(\bar{x} > c') = 1 - P(\bar{x} \le c') = 1 - \left[\frac{1}{2} + \frac{1}{2}\Phi\left(\frac{c'-a_0}{\sigma}\sqrt{n}\right)\right] = \frac{1}{2} - \frac{1}{2}\Phi\left(\frac{c'-a_0}{\sigma}\sqrt{n}\right) = \alpha,$$

звідки $\Phi\left(\frac{c'-a_0}{\sigma}\sqrt{n}\right) = 1 - 2\alpha$ або за таблицями $\frac{c'-a_0}{\sigma}\sqrt{n} = t_{1-2\alpha}$. Таким чином, межа критичної області W визначається значенням $c' = \frac{\sigma \cdot t_{1-2\alpha}}{\sqrt{n}} + a_0$.

Отже, найбільш потужним критерієм перевірки гіпотези H_0: $a = a_0$ на противагу альтернативній гіпотезі H_1: $a = a_1 > a_0$ є такий*: гіпотеза* H_0 *відхиляється, якщо* $\bar{x} > \frac{\sigma \cdot t_{1-2\alpha}}{\sqrt{n}} + a_0$; H_0 *приймається, якщо* $\bar{x} \le \frac{\sigma \cdot t_{1-2\alpha}}{\sqrt{n}} + a_0$.

1. Для пошуку потужності критерію визначимо спочатку ймовірність β допустити помилку 2-го роду – прийняти гіпотезу, коли вона хибна, тобто має місце альтернативна гіпотеза $X \approx N(a_1;\sigma^2)$ або $\bar{x} \approx N\left(a_1;\frac{\sigma^2}{\sqrt{n}}\right)$:

$$\beta = P\left(\bar{x} \le \frac{\sigma \cdot t_{1-2\alpha}}{\sqrt{n}} + a_0\right) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{a_0 + t_{1-2\alpha}\sigma/\sqrt{n} - a_1}{\sigma}\sqrt{n}\right) =$$

$$= \frac{1}{2} - \frac{1}{2}\Phi\left(\frac{\left(a_1 - a_0\sqrt{n}\right)}{\sigma} - t_{1-2\alpha}\right).$$

Таким чином, потужність критерію є такою:

$$1 - \beta = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{\left(a_1 - a_0\sqrt{n}\right)}{\sigma} - t_{1-2\alpha}\right).$$

Проаналізувавши отримане значення, можна переконатися, що зменшення рівня значимості α при сталому обсязі вибірки приводить до збільшення ймовірності β та, відповідно, до зменшення потужності критерію $(1-\beta)$. І лише при збільшенні обсягу вибірки можна, зменшуючи ймовірність α, одночасно зменшити ймовірність β (збільшувати потужність критерію $(1-\beta)$).

2. При заданих ймовірностях помилок 1-го та 2-го роду α та β з виразу для β не складно знайти відповідний обсяг вибірки за формулою:

$$n = \frac{\left(t_{1-2\alpha} + t_{1-2\beta}\right)^2 \sigma^2}{\left(a_1 - a_0\right)^2}.$$

Принцип тестування статистичної гіпотези не дає логічного доведення її істинності чи хибності. Більш того, прийняття гіпотези H_0 не потрібно розглядати як назавжди встановлений, абсолютно достовірний факт.

Розглянемо деякі методи тестування статистичних гіпотез, зауваживши попередньо, що в статистиці прийнято середнє квадратичне відхилення позначати s_x, s_y і т. п., а дисперсію – s_x^2, s_y^2 і т. п.

1. Перевірка гіпотез про рівність середніх. У промисловості задача порівняння середніх часто виникає при вибірковому контролі якості продукції, виготовленої на різних установках чи при різних технологічних режимах, у фінансовому аналізі – при порівнянні рівня дохідності різних активів і т. п.

Сформулюємо задачу. Нехай є дві сукупності, які характеризуються генеральними середніми $\overline{x_0}$ та $\overline{y_0}$ та відомими дисперсіями s_x^2 та s_y^2. Необхідно перевірити гіпотезу H_0 про рівність генеральних середніх. Для перевірки цієї гіпотези із сукупностей взято дві незалежні вибірки обсягів n_1 та n_2, за якими знайдено середні арифметичні $\overline{x}$ та $\overline{y}$ і вибіркові дисперсії D'_x та D'_y.

Якщо H_0 справедлива, то різниця $\overline{x} - \overline{y}$ має нормальний закон розподілу з математичним сподіванням $M(\overline{x} - \overline{y}) = \overline{x_0} - \overline{y_0} = 0$ та дисперсією $s^2_{\overline{x}-\overline{y}} = s^2_{\overline{x}} + s^2_{\overline{y}} = \dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}$. Тому при виконанні гіпотези H_0 статистика

$$t = \frac{\left(\overline{x} - \overline{y}\right) - M\left(\overline{x} - \overline{y}\right)}{s_{\overline{x}-\overline{y}}} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}} \tag{2.47}$$

має стандартний нормальний розподіл $N(0,1)$.

У випадку конкуруючої гіпотези H_1: $\overline{x_0} > \overline{y_0}$ (чи H_1: $\overline{x_0} < \overline{y_0}$) критичне значення статистики знаходять з умови

$$\Phi(t_{кр}) = \Phi(t_{1-2\alpha}) = 1 - 2\alpha, \qquad (2.48)$$

а при конкуруючій гіпотезі $H_1: \ \overline{x_0} \neq \overline{y_0}$ критичне значення статистики знаходять з умови

$$\Phi(t_{кр}) = \Phi(t_{1-\alpha}) = 1 - \alpha. \qquad (2.49)$$

Якщо спостережуване значення статистики t більше за критичне $t_{кр}$, визначене на рівні значимості α (за абсолютною величиною), то гіпотеза H_0 не приймається. У протилежному ви падку роблять висновок, що дана гіпотеза не суперечить результатам спостережень.

Приклад 2.13. Для перевірки ефективності нової технології відібрано дві групи робітників: в першій групі чисельністю n_1 = 50 чол., де застосовувалась нова технологія, вибірковий середній виробіток склав $\overline{x} = 85$ (виробів); в другій групі чисельністю n_2 = 70 чол. вибірковий середній виробіток склав $\overline{x} = 78$ (виробів). Попередньо встановлено, що дисперсії виробітку в обох групах відповідно $s_x^2 = 100$ та $s_y^2 = 74$. На рівні значимості α = 0,05 з'ясувати вплив нової технології на середню продуктивність.

Розв'язання. У нашому випадку тестуємо гіпотезу H_0: $\overline{x_0} = \overline{y_0}$, тобто середня продуктивність однакова за новою та старою технологіями. Як конкуруючу гіпотезу розглянемо гіпотезу $H_1: \ \overline{x_0} > \overline{y_0}$, оскільки справедливість цієї гіпотези означає ефективність застосування нової технології. За формулою (2.47) фактичне значення статистики критерію

$$t = \frac{85 - 78}{\sqrt{\dfrac{100}{50} + \dfrac{74}{70}}} = 4.$$

За формулою (2.49) знайдемо критичне значення статистики $\Phi(t_{кр}) = 1 - 2\alpha = 1 - 2 \cdot 0{,}05 = 0{,}9$; звідки за таблицею додатка В $t_{кр} = 1{,}64$. Оскільки фактичне значення статистики більше за критичне, то гіпотеза H_0 не приймається, тобто на п'ятивідсотковому рівні значимості можна стверджувати, що нова технологія дозволяє підвищити середню продуктивність робітників.

2. Перевірка гіпотез про рівність дисперсій двох сукупностей. Гіпотези про дисперсії виникають доволі часто, оскільки дисперсія характеризує такі важливі показники, як точність машин, приладів, технологічних процесів, ризик, пов'язаний з відхиленням дохідності активів та очікуваного рівня і т. д.

Сформулюємо задачу. Нехай є дві нормально розподілені сукупності, дисперсії яких дорівнюють s_1^2 та s_2^2. Необхідно перевірити нуль-гіпотезу про рівність дисперсій, тобто H_0: $s_1^2 = s_2^2 = s^2$ відносно конкуруючої гіпотези H_1: $s_1^2 > s_2^2$ чи H_1: $s_1^2 \neq s_2^2$.

Для перевірки гіпотези H_0 із цих сукупностей взято дві незалежні вибірки обсягом n_1 та n_2. Для оцінювання дисперсій використовуються "виправлені" вибіркові дисперсії $\hat{D}_1'$ та $\hat{D}_2'$. Тобто, задача перевірки гіпотези зводиться до порівняння цих дисперсій.

Вибіркові статистики $\frac{(n_1-1)\hat{D}_1'}{\sigma^2}$ та $\frac{(n_2-1)\hat{D}_2'}{\sigma^2}$ мають розподіл χ^2 відповідно з $k_1 = n_1 - 1$ та $k_2 = n_2 - 1$ ступенями вільності, а їх відношення $\frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}$ має розподіл Фішера-Снедекора з k_1 та k_2 ступенями вільності.

Отже, випадкова величина F визначається відношенням:

$$F = \frac{\frac{1}{n_1-1}\left[\frac{(n_1-1)\hat{D}_1'}{\sigma^2}\right]}{\frac{1}{n_2-1}\left[\frac{(n_2-1)\hat{D}_2'}{\sigma^2}\right]} = \frac{\hat{D}_1'}{\hat{D}_2'}. \tag{2.50}$$

Гіпотеза H_0 відкидається, якщо розрахована F-статистика більша табличного значення $F_{\alpha;k_1;k_2}$ (додаток Д).

3. Побудова теоретичного закону розподілу за експериментальними даними. Перевірка гіпотез про закон розподілу. Однією з найважливіших задач математичної статистики є встановлення теоретичного закону розподілу випадкової величини, яка характеризує досліджувану ознаку за емпіричним розподілом, що подається варіаційним рядом.

Припущення щодо вигляду закону розподілу може бути висунуте, виходячи з теоретичних передумов, досвіду аналогічних попередніх досліджень та на основі графічного зображення емпіричного розподілу.

На практиці найчастіше застосовують *χ^2-критерій Пірсона* для встановлення розбіжності між теоретичним та емпіричним законами розподілу. За міру розбіжності U обирають величину χ^2, рівну сумі квадратів відхилень статистичних ймовірностей ω_i від гіпотетичних ймовірностей p_i, розрахованих за припущеним розподілом і взятих з деякими вагами c_i:

$$U = \chi^2 = \sum_{i=1}^{m} c_i (\omega_i - p_i)^2 . \qquad (2.51)$$

Ваги c_i вводять таким чином, щоб за одних і тих самих відхилень $(\omega_i - p_i)^2$ більшу вагу мали ті відхилення, при яких ймовірність p_i мала, та меншу вагу – при яких p_i велика. Обравши $c_i = \dfrac{n}{p_i}$ можна показати, що при $n \to \infty$ статистика

$$U = \chi^2 = \sum_{i=1}^{m} \frac{n}{p_i} (\omega_i - p_i)^2 ,$$

чи

$$U = \chi^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i} \qquad (2.52)$$

має розподіл χ^2 із $k = m - r - 1$ ступенями вільності,
де m – число інтервалів емпіричного розподілу (варіаційного ряду);
r – число параметрів теоретичного розподілу, обчислених за експериментальними даними.

Числа $n_i = n\omega_i$ та np_i називають відповідно *емпіричними* та *теоретичними частотами*.

Схема застосування критерію χ^2 для перевірки нуль-гіпотези така.

1. Визначають міру розбіжності емпіричних та теоретичних частот χ^2 за (2.52).
2. Для обраного рівня значимості α за таблицею χ^2-розподілу знаходять критичне значення $\chi^2_{\alpha;k}$ при числі ступенів вільності $k = m - r - 1$ (додаток Б).
3. Якщо розраховане значення більше за критичне, тобто $\chi^2 > \chi^2_{\alpha;k}$, то гіпотеза H_0 не суперечить практичних даним.

Зауваження. При застосуванні критерію Пірсона необхідно, щоб в кожному інтервалі було не менше 5 спостережень.

Приклад 2.14. Для емпіричного розподілу кравчинь швейного цеху за виробітком дібрати відповідний теоретичний розподіл та на рівні значимості $\alpha = 0{,}05$ перевірити гіпотезу про згоду двох розподілів за допомогою критерію χ^2.

Виробіток у звітному році у відсотках до попереднього, x	Частота (кількість кравчинь), n_i
94,0 − 100,0	3
100,0 − 106,0	7
106,0 − 112,0	11
112,0 − 118,0	20
118,0 − 124,0	28
124,0 − 130,0	19
130,0 − 136,0	10
136,0 − 142,0	2
$\sum$	100

Розв'язання. За виглядом гістограми розподілу кравчинь за виробітком (рис. 2.16) можна припустити нормальний закон розподілу ознаки.

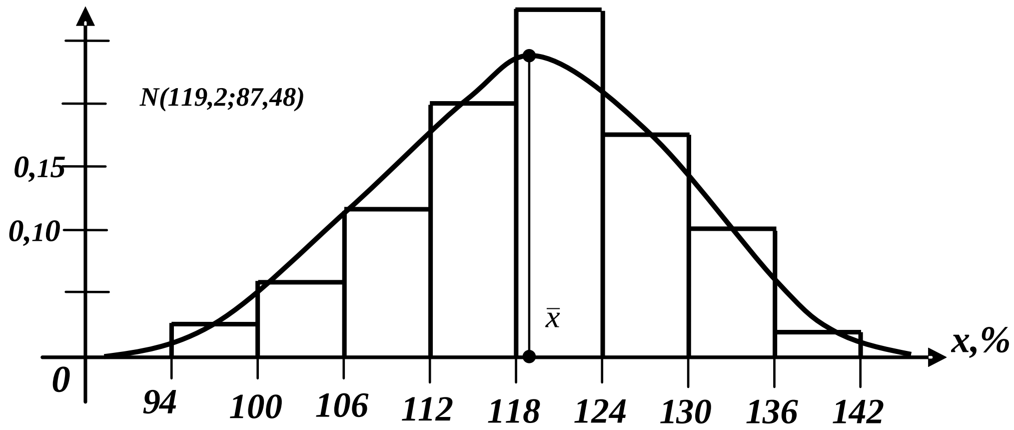


Рисунок 2.16 – Гістограма розподілу кравчинь за виробітком

Параметри нормального закону невідомі, тому замінюємо їх "найкращими" оцінками за вибіркою – вибірковою середньою $\overline{x}$ та "виправленою" вибірковою дисперсією $\hat{D}'$. Оскільки число спостережень $n = 100$ досить велике, то замість "виправленої" дисперсії $\hat{D}'$ можна взяти "звичайну" вибіркову дисперсію D.

Таким чином висуваємо гіпотезу H_0: випадкова величина X – виробіток кравчинь цеху – розподілена нормально з параметрами $a = 119{,}2$; $\sigma^2 = 87{,}48$, тобто $X \cong N(119{,}2;\quad 87{,}48)$.

Для розрахунку ймовірностей p_i потрапляння випадкової величини X в інтервал $[x_i, x_{i+1}]$ використовуємо функцію Лапласа згідно з властивістю нормального розподілу:

$$p_i(x_i \le X \le x_{i+1}) = \frac{1}{2}\left[\Phi\left(\frac{x_{i+1} - a}{\sigma}\right) - \Phi\left(\frac{x_i - a}{\sigma}\right)\right] \approx$$

$$\approx \frac{1}{2}\left[\Phi\left(\frac{x_{i+1}-119{,}2}{9{,}35}\right)-\Phi\left(\frac{x_i-119{,}2}{9{,}35}\right)\right].$$

Наприклад, $p_1(94 \le X \le 100) = \frac{1}{2}\left[\Phi\left(\frac{100-119{,}2}{9{,}35}\right)-\Phi\left(\frac{94-119{,}2}{9{,}35}\right)\right] =$

$= \frac{1}{2}[\Phi(-2{,}05)-\Phi(-2{,}69)] = \frac{1}{2}(-0{,}9596+0{,}9928) = 0{,}0166$ та теоретична частота, що відповідає першому інтервалу $np_1 = 100 \cdot 0{,}0166 \approx 1{,}7$ і т. п.

Для визначення статистики χ^2 зручно скласти таблицю 2.2.

Таблиця 2.2 – Проміжні розрахунки для визначення статистики χ^2

i	*Інтервал* $[x_i, x_{i+1}]$	*Емпіричні частоти* n_i	*Ймовірності* p_i	*Теоретичні частоти* np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
1	94 − 100	3 } 10	0,017	1,7 } 7,6	5,76	0,758
2	100 − 106	7	0,059	5,9		
3	106 − 112	11	0,141	14,1	9,61	0,682
4	112 − 118	20	0,228	22,8	7,84	0,344
5	118 − 124	28	0,247	24,7	10,89	0,441
6	124 − 130	19	0,182	18,2	0,64	0,035
7	130 − 136	10 } 12	0,087	8,7 } 11,6	0,16	0,014
8	136 − 142	2	0,029	2,9		
	$\sum$	100	0,990	99,0	–	$\chi^2 = 2{,}27$

Враховуючи, що в даному емпіричному розподілі частоти першого та останнього інтервалів менші 5, при використанні критерію χ^2-Пірсона доцільно об'єднати вказані інтервали із сусідніми (табл. 2.2).

Таким чином, спостережуване значення статистики $\chi^2 = 2{,}27$. Оскільки нова кількість інтервалів $m = 6$, а нормальний закон розподілу визначається $r = 2$ параметрами, то число ступенів вільності $k = m - r - 1 =$ $= 6 - 2 - 1 = 3$. Відповідне критичне значення статистики за таблицею додатка Б $\chi^2_{0,05;3} = 7{,}82$. Оскільки $\chi^2 < \chi^2_{0,05;3}$, то гіпотеза про обраний теоретичний нормальний закон $N(119{,}2;\ \ 87{,}48)$ узгоджується з практичними даними.

Зауваження. Для графічного зображення емпіричного закону та вирівнювального теоретичного нормального розподілу необхідно використовувати однаковий для двох розподілів масштаб.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Яку величину називають випадковою?
2. Яку випадкову величину називають дискретною, а яку – неперервною?
3. Що називають законом розподілу випадкової величини?
4. Дайте означення ряду розподілу. Охарактеризуйте полігон розподілу.
5. Вкажіть найбільш уживані операції над випадковими величинами.
6. Що називають математичним сподіванням дискретної випадкової величини? Сформулюйте властивості математичного сподівання, будь-які дві доведіть. Вкажіть математичний зміст математичного сподівання.
7. Що називають дисперсією та середнім квадратичним відхиленням випадкової величини? Сформулюйте основні властивості дисперсії, будь-які дві доведіть. Вкажіть механічний зміст дисперсії.
8. Дайте означення функції розподілу. Вкажіть загальний вид цієї функції для дискретної випадкової величини. Сформулюйте властивості функції розподілу.
9. Доведіть, що ймовірність будь-якого окремо взятого значення неперервної випадкової величини дорівнює нулю.
10. Дайте означення щільності ймовірності. Що називають кривою розподілу випадкової величини? Чи існує щільність ймовірностей для дискретної випадкової величини?
11. Запишіть формули для обчислення математичного сподівання та дисперсії неперервної випадкової величини.
12. Доведіть коректність означення закону Пуассона.
13. Чому дорівнюють математичне сподівання та дисперсія випадкової величини, розподіленої за законом Пуассона?
14. Доведіть, що сума двох незалежних випадкових величин, розподілених за законом Пуассона з параметрами λ_1 та λ_2, також розподілена за законом Пуассона з параметром $\lambda = \lambda_1 + \lambda_2$.
15. Охарактеризуйте рівномірний закон розподілу.
16. Охарактеризуйте показниковий закон та вкажіть область його застосування.
17. Довести, що якщо проміжок часу T, розподілений за показниковим законом, вже тривав деякий час τ, то це ніяк не впливає на закон розподілу проміжку $T_1 = T - \tau$.
18. Запишіть функцію щільності ймовірності нормального закону.
19. Як буде змінюватися нормальна крива при зміні параметрів a та σ^2?
20. Охарактеризуйте властивості нормально розподіленої випадкової величини.
21. Сформулюйте правило "трьох сигм".

22. Охарактеризуйте розподіли χ^2, Стьюдента та Фішера-Снедекора.
23. Що таке статистична гіпотеза? Яку гіпотезу називають нуль-гіпотезою, а яку – альтернативною?
24. Охарактеризуйте суть перевірки статистичної гіпотези.
25. У чому суть помилок 1-го та 2-го роду? Дайте означення рівня значимості та потужності критерію.
26. Сформулюйте критерій Неймана-Пірсона.
27. Охарактеризуйте перевірку гіпотез про рівність середніх.
28. Опишіть перевірку гіпотез про рівність дисперсій двох сукупностей.
29. Охарактеризуйте перевірку гіпотез про побудову теоретичного закону розподілу за експериментальними даними.
30. Проілюструйте перевірку гіпотез про закони розподілу.

ВПРАВИ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Завдання 2.1. Клієнти банку, не пов’язані між собою, не повертають кредити у вказаний термін з ймовірністю 0,4. Скласти закон розподілу числа повернених кредитів із 7 виданих. Знайти математичне сподівання, дисперсію та середнє квадратичне відхилення цієї випадкової величини.

Завдання 2.2. Знайти закон розподілу трьох пакетів акцій, якщо ймовірність отримання прибутку для кожного з них дорівнює відповідно 0,2; 0,4; 0,7. Знайти математичне сподівання та дисперсію даної випадкової величини, побудувати її функцію розподілу.

Завдання 2.3. Торговий агент має 7 телефонних номерів потенційних клієнтів та дзвонить їм до тих пір, доки не отримає замовлення. Ймовірність того, що потенційний покупець зробить замовлення, дорівнює 0,2. Скласти закон розподілу числа телефонних розмов, які здійснить агент. Знайти математичне сподівання та дисперсію даної випадкової величини.

Завдання 2.4. Із рекламною метою торгова фірма вкладає в кожну двадцяту одиницю товару грошовий приз розміром 700 грн. Скласти закон розподілу випадкової величини – розміру виграшу при чотирьох зроблених покупках. Знайти математичне сподівання та дисперсію цієї випадкової величини.

Завдання 2.5. Дискретна випадкова величина задана рядом розподілу:

$x=x_i$	0	2	3	6	7
p_i	0,1	0,3	0,2	0,1	0,3

Знайти:

а) функцію розподілу F(x) і побудувати її графік;

б) числові характеристики випадкової величини X.

Завдання 2.6. Дискретна випадкова величина X може приймати тільки два значення: x_1 і x_2, причому $x_1 < x_2$. Ймовірність того, що X прийме значення x_1, дорівнює 0,4. Знайти закон розподілу X, якщо математичне сподівання $\mathrm{M}(X) = 0{,}2$ і дисперсія $\mathrm{D}(X) = 0{,}96$.

Завдання 2.7. Щільність розподілу випадкової величини X задана виразом: $f(x) = \begin{cases} 0, & x < 0 \\ Ax^3, & 0 \le x \le 1 \\ 0, & x > 1 \end{cases}$. Знайти: коефіцієнт a, $F(x)$, математичне сподівання $M(X)$, дисперсію $D(X)$, $P(\frac{1}{2} < x < \frac{3}{4})$. Побудувати графіки $f(x)$ і $F(x)$.

Завдання 2.8. Випадкова величина X має рівномірний розподіл з математичним сподіванням $M(X)=3$ і дисперсією $D(X)=\frac{4}{3}$. Знайти щільність імовірності випадкової величини X, інтегральну функцію $F(x)$ і $P(2 < x < 3)$. Побудувати графіки $f(x)$ і $F(x)$.

Завдання 2.9. Математичне сподівання і середнє квадратичне відхилення нормально розподіленої випадкової величини X відповідно дорівнюють 10 і 2. Знайти ймовірність того, що в результаті випробування випадкова величина X прийме значення, що належить інтервалу (12;14).

Завдання 2.10. Хвилинна стрілка електричного годинника рухається стрибком в кінці кожної хвилини. Знайти ймовірність того, що в дану мить годинник покаже час, який відрізняється від справжнього не більше, ніж на 8 с.

Завдання 2.11. Є такі статистичні дані про число викликів спеціалізованих бригад швидкої допомоги в м. Вінниці протягом 300 год:

Число викликів за год., x_i	0	1	2	3	4	5	6	7	8	$\sum$
Частота n_i	15	71	75	68	39	17	10	4	1	300

Припускаючи, що число викликів швидкої розподілене за законом Пуассона, при рівні значимості $\alpha = 0{,}05$ перевірити гіпотезу про узгодження двох розподілів за допомогою критерію χ^2.

Завдання 2.12. Витрати сировини на одиницю продукції склали:

За старою технологією					За новою технологією					
x_i	30	30	30	Разом	y_j	30	30	30	30	Разом
n_i	1	4	4	9	n_j	2	6	4	1	13

Припускаючи, що витрати сировини за кожною технологією мають нормальний розподіл з однаковими дисперсіями, на рівні значимості

$\alpha = 0{,}05$ з'ясувати, чи дає нова технологія економію в середніх витратах сировини.

Завдання 2.13. Вступний іспит проводили на двох факультетах інституту. На першому факультеті серед $n_1 = 900$ абітурієнтів витримали іспит $m_1 = 500$; а на іншому факультеті серед $n_2 = 800$ абітурієнтів – $m_2 = 408$. На рівні значимості $\alpha = 0{,}05$ перевірити гіпотезу про відсутність суттєвих відмінностей в рівні підготовки абітурієнтів двох факультетів. Розглянути випадок, при якому конкуруючою гіпотезою є $H_1: \quad p_1 > p_2$.

Завдання 2.14. Встановлено, що середня вага пігулки ліків сильної дії (номінал) повинна дорівнювати 0,5 мг. Вибіркова перевірка $n = 100$ пігулок показала, що середня вага пігулки $\bar{x} = 0{,}53$ мг. На базі проведених досліджень можна вважати, що вага пігулки є нормально розподіленою випадковою величиною з середнім квадратичним відхиленням $\sigma = 0{,}11$ мг. На рівні значимості $\alpha = 0{,}05$: а) з'ясувати, чи можна вважати одержане у вибірці відхилення від номіналу випадковим; б) знайти потужність критеріїв, використаних в а).

Завдання 2.15. Є такі дані про число складених іспитів в сесію студентами-заочниками:

Число складених іспитів x_i	0	1	2	3	4	$\sum$
Число студентів n_i	1	1	1	3	3	41

На рівні значимості $\alpha = 0{,}05$ перевірити гіпотезу про те, що випадкова величина X – число складених студентами іспитів – розподілена за біноміальним законом, використовуючи критерій χ^2-Пірсона.

ТЕСТИ

1. Щільність розподілу випадкової величини – це функція $f(x)$, для якої (F – функція розподілу):

а) $F(x) = \int\limits_{x}^{+\infty} f(t)dt$;

б) $F(x) = \int f(x)dx + C$;

в) $F(x) = \int\limits_{0}^{x} f(t)dt$;

г) $F(x) = \int\limits_{-\infty}^{x} f(t)dt$.

2. Математичне сподівання неперервної випадкової величини з щільністю розподілу $f(x)$ дорівнює:

а) $\int\limits_{-\infty}^{+\infty} xf(x)dx$;

б) $\int\limits_{0}^{+\infty} xf(x)dx$;

в) $\int\limits_{0}^{+\infty} x^2 f(x)dx$;

г) $\int\limits_{-\infty}^{+\infty} x^2 f(x)dx$.

3. Середньоквадратичне відхилення випадкової величини є:

а) квадратним коренем з дисперсії цієї величини;

б) середнім значенням квадрата цієї величини;

в) відхиленням середнього значення квадрата випадкової величини від її середнього значення;

г) квадратом середнього значення цієї величини.

4. Випадкова величина ξ, задана функцією розподілу

$$F_\xi(x) = \begin{cases} 0, \; x \le 0, \\ 0,2x, \quad 0 < x \le 5, \\ 1, \; x > 5. \end{cases}$$

знайти щільність розподілу випадкової величини ξ .

а) $f_\xi(x) = \begin{cases} 0, \; x \le 0, \\ 0{,}1x^2, \quad 0 < x \le 5, \\ x, \; x > 5. \end{cases}$

б) $f_\xi(x) = \begin{cases} 0, \; x \le 0, \\ 0,2x \;, \quad 0 < x \le 5, \\ 0, \; x > 5. \end{cases}$

в) $f_\xi(x) = \begin{cases} 0, \; x \le 0, \\ 0,2, \quad 0 < x \le 5, \\ 0, \; x > 5. \end{cases}$

г) $f_\xi(x) = \begin{cases} 1, \; x \le 0, \\ 0{,}1x^2, \quad 0 < x \le 5, \\ 1, \; x > 5. \end{cases}$

5. Задано закон розподілу дискретної випадкової величини X . Знайти математичне сподівання *M(X)*, якщо

X	-2	-1	0	2
p	0,1	0,3	0,4	0,2

а) 0; б)1; в) –1; г) –0,1.

6. Задано закон розподілу дискретної випадкової величини X . Знайти дисперсію $D(X)$.

X	-2	-1	0	2
p	0,1	0,3	0,4	0,2

а) 1; б) 0,5; в)1,49; г)1,5.

7. Задана щільність $f_{\xi}(x)$ розподілу випадкової величини ξ. Знайти математичне сподівання $M(\xi)$, якщо

$$f_{\xi}(x) = \begin{cases} 2(1-x), \quad x \in (0;1), \\ 0, \; x \notin (0;1) \end{cases}$$

а) $\frac{2}{3}$; б) $\frac{1}{2}$; в) $\frac{1}{3}$; г) $\frac{1}{4}$.

8. Випадкова величина X нормально розподілена. $M(X) = 1$, $D(X) = 4$. Знайти $P(|X-1| < 2)$.

а) 0,68; б) 0,58; в) 0,78; г) 0,48.

9. Вантажі із залізничної станції вивозять автомобілями за кільцевими маршрутами. Визначити вантажопідйомність автомобіля на маршруті $M(X)$, якщо обсяг перевезень розподіляється за показниковим законом, коли $a = 0,25$.

а) 7; б) 4; в) 8; г) 5 .

10. Випадкова величина ξ має розподіл Пуассона з параметром λ. Які із рівностей є абсолютно правильними?

1) $P(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda}$, при k =0,1, 2,… ; 2) $M(\xi) = \lambda$; 3) $D(\xi) = \lambda^2$.

а) тільки 1 і 2;
б) тільки 1 і 3;
в) тільки 2 і 3;
г) інша відповідь.

11. Якщо ймовірність p появи події A в кожному випробуванні стала, то ймовірність $P_{m,n}$ того, що подія A з'явиться m раз в n незалежних випробуваннях, дорівнює:

а) $P_{m,n} = C_n^m \cdot p^m \cdot q^{n-m}$;

б) $P_{m,\,n} \approx \dfrac{\lambda^m \cdot e^{-\lambda}}{m!} = P_m(\lambda)$;

в) $P_{m,\,n} = p^m \cdot q^{n-m}$;

г) $P_{m,n} = C_n^m \cdot p^m \cdot q^{n+m}$.

12. Формула Пуассона застосовується у випадку, коли:

а) $\lambda = np \le 10$;

б) $\lambda = np > 10$;

в) $\lambda = np \le 100$;

г) $\lambda = np > 20$.

13. Дискретною ймовірнісною величиною називають величину *X*, яка має:

а) зчисленну множину можливих значень;

б) незчисленну множину можливих значень;

в) скінченну, але незчисленну множину можливих значень;

г) нескінченну, але зчисленну множину можливих значень.

14. Неперервною ймовірнісною величиною є величина *X*, що має:

а) зчисленну множину можливих значень;

б) нескінченну незчисленну множину можливих значень;

в) скінченну, але незчисленну множину можливих значень.

г) нескінченну, але зчисленну множину можливих значень.

15. Яка із перерахованих властивостей не відноситься до властивостей математичного сподівання:

а) $M(C)=C$;

б) $M(X_1 + X_2 + ... + X_n) = M(X_1) + M(X_2) + ... + M(X_n)$;

в) $M(CX)= C^2M(X),\ C = const$;

г) інша відповідь.

16. Дисперсія неперервної випадкової величини обчислюється за формулою:

а) $\int\limits_0^{+\infty} (x - M[X])^2 f(x)dx$, де $f(x)$ – щільність розподілу;

б) $\int\limits_{-\infty}^{+\infty} (x - M[X])^2 f(x)dx$, де $f(x)$ – щільність розподілу;

в) $\int\limits_{-\infty}^{+\infty} (x - M[X])^2 f(x)dx$, де $f(x)$ – довільна функція;

г) інша відповідь.

17. Яка з перерахованих властивостей не відноситься до властивостей дисперсії:

а) $D(X \pm C) = D(X) \pm C;$

б) $D(C)=0;$

в) $D(CX) = C^2 D(X), C = const;$

г) інша відповідь.

18. Функцією розподілу (інтегральною функцією розподілу) ВВ X називають функцію $F(x)$, значення якої дорівнюють:

а) $P(X < x)$;

б) $P(X > x)$;

в) $P(X = x)$;

г) інша відповідь.

19. Щільністю ймовірності неперервної випадкової величини називається:

а) інтеграл від її функції розподілу;

б) функція розподілу;

в) похідна її функції розподілу;

г) інша відповідь.

20. Яка з перерахованих властивостей не є властивістю щільності розподілу $f(x)$:

а) $\int\limits_{-\infty}^{\infty} f(x)dx = 1;$

б) $P(a \le X < b) = \int\limits_{a}^{b} f(x)dx;$

в) $F(x) = \int\limits_{-\infty}^{x} f(x)dx;$

г) $f(x) < 0.$

21. Яка з перерахованих властивостей не є властивістю функції розподілу $F(x)$:

а) $F(x) > 1;$

б) $P(a \le x < b) = F(b) - F(a);$

в) $\lim\limits_{x \to -\infty} F(x) = 0, \ \lim\limits_{x \to +\infty} F(x) = 1;$

г) $f(x) = F'(x).$

22. Математичне сподівання випадкової величини, розподіленої за законом Пуассона:

а) $M(X) = np$;
б) $M(X) = npq$;
в) $M(X) = \sqrt{np}$;
г) $M(X) = \lambda$.

23. Дисперсія випадкової величини, розподіленої за рівномірним законом, дорівнює:
а) $D(X) = \frac{(b-a)^2}{2}$;
б) $D(X) = \frac{(b-a)^2}{12}$;
в) $D(X) = \frac{(b+a)^2}{12}$;
г) $D(X) = \sqrt{np}$.

24. Неперервна випадкова величина має нормальний закон розподілу з параметрами a та σ^2, якщо її щільність ймовірності така:
а) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{(x-a)^2/(2\sigma^2)}$;
б) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x+a)^2/(2\sigma^2)}$;
в) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)}$;
г) $f(x) = -\frac{1}{\sigma\sqrt{2\pi}} e^{-(x+a)^2/(2\sigma^2)}$.

25. Для нормально розподіленої випадкової величини:
а) $M(X) = a$, $D(X) = \sigma^2$;
б) $M(X) = a$, $D(X) = \sigma^3$;
в) $M(X) = 0$, $D(X) = \sigma^2$;
г) $M(X) = a, D(X) = 0$.

26. Зміна параметра *a* нормально розподіленої неперервної випадкової величини характеризує:
а) форму кривої розподілу;
б) площу під кривою розподілу;
в) положення кривої розподілу;
г) інша відповідь.

27. Зміна параметра σ^2 нормально розподіленої неперервної випадкової величини характеризує:

а) форму кривої розподілу;

б) площу під кривою розподілу;

в) положення кривої розподілу;

г) інша відповідь.

28. Ймовірність потрапляння нормально розподіленої випадкової величини в інтервал $[x_1, x_2]$ обчислюється за формулою:

а) $P(x_1 \le X \le x_2) = \frac{1}{2}(\Phi(t_2) + \Phi(t_1))$, $t_1 = \frac{x_1 - a}{\sigma}$, $t_2 = \frac{x_2 - a}{\sigma}$, $\Phi(t)$ – функція Лапласа;

б) $P(x_1 \le X \le x_2) = \frac{1}{2}(\Phi(t_2) - \Phi(t_1))$, $t_1 = \frac{x_1 - a}{\sigma}$, $t_2 = \frac{x_2 - a}{\sigma}$, $\Phi(t)$ – функція Лапласа;

в) $P(x_1 \le X \le x_2) = \frac{1}{2}(\Phi(t_2) - \Phi(t_1))$, $t_1 = \frac{x_1 - a}{\sigma}$, $t_2 = \frac{x_2 - a}{\sigma}$, $\Phi(t)$ – функція Гаусса;

г) $P(x_1 \le X \le x_2) = \frac{1}{2}(\Phi(t_2) - \Phi(t_1))$, $t_1 = \frac{x_1 + a}{\sigma}$, $t_2 = \frac{x_2 + a}{\sigma}$, $\Phi(t)$ – функція Лапласа.

29. Випадкова величина X має показниковий закон розподілу з параметром λ, якщо його щільність ймовірності така:

а) $f(x) = \begin{cases} -\lambda e^{-\lambda x}, & \text{якщо } x \ge 0; \\ 0, & \text{якщо } x < 0; \end{cases}$

б) $f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{якщо } x \ge 0; \\ 0, & \text{якщо } x < 0; \end{cases}$

в) $f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{якщо } x < 0; \\ 0, & \text{якщо } x \ge 0; \end{cases}$

г) $f(x) = \begin{cases} e^{-\lambda x}, & \text{якщо } x \ge 0; \\ 0, & \text{якщо } x < 0. \end{cases}$

30. Якщо випадкова величина розподілена за показниковим законом, то її дисперсія дорівнює:

а) $D(X) = \frac{1}{\lambda^3}$;

б) $D(X) = \frac{1}{\lambda}$;

в) $D(X) = -\frac{1}{\lambda^2}$;

г) $D(X) = \frac{1}{\lambda^2}$.

31. Рівнем значущості критерію перевірки статистичної гіпотези є:

а) ймовірність того, що результат перевірки буде правильним;

б) ймовірність помилки першого роду;

в) ймовірність помилки другого роду;

г) максимальне відхилення вибіркового значення статистики критерію від критичного.

32. Основна гіпотеза підтверджується, якщо вибіркове значення статистики критерію:

а) менше критичного значення;

б) більше критичного значення;

в) потрапляє в критичну область;

г) не потрапляє в критичну область.

33. Які з оцінок є оцінками математичного сподівання?

1) $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$; 2) медіана; 3) $\frac{1}{2}(x_{\min} + x_{\max})$; 4) мода.

а) тільки 1, 3 і 4;

б) тільки 2, 3 і 4;

в) тільки 1, 2 і 3;

г) тільки 1.

34. Рівнем значущості критерію перевірки статистичної гіпотези є:

а) ймовірність того, що результат перевірки буде правильним;

б) ймовірність помилки першого роду;

в) ймовірність помилки другого роду;

г) максимальне відхилення вибіркового значення статистики критерію від критичного.

35. Яке з тверджень, щодо перевірки статистичних гіпотез, є помилковим?

а) помилкою першого типу є відхилення правильної гіпотези;

б) помилкою другого типу є підтвердження неправильної гіпотези;

в) перевірка статистичної гіпотези є логічним доведенням її правильності чи хибності;

г) для кожної статистичної гіпотези існує альтернативна гіпотеза.

ТЕМА 3 ОДНОФАКТОРНИЙ КОРЕЛЯЦІЙНИЙ АНАЛІЗ

Досить часто в практиці економічних досліджень намагаються визначити криву (поверхню), яка дає найкраще (в сенсі методу найменших квадратів) наближення вихідних даних. Відповідні методи наближення одержали назву регресійного аналізу.

Методи і моделі регресійного аналізу займають центральне місце в математичному апараті економетрії. Задачами регресійного аналізу є з'ясування форми залежності між змінними, оцінювання функції регресії, оцінювання невідомих значень (прогноз значень) залежної змінної.

3.1 Функціональна, статистична та кореляційна залежності

У природничих науках часто йде мова про функціональну залежність, коли кожному значенню однієї змінної відповідає певне значення іншої (наприклад, швидкість вільного падіння у вакуумі залежно від часу і т. п.).

В економіці у більшості випадків відповідні показники комерційної діяльності можуть знаходитись у нижченаведених видах зв'язку.

1. *Балансовий зв'язок* показників характеризує залежності між джерелами формування ресурсів та їх використанням. Наприклад, формула товарного балансу:

$$З_п + Н = В + З_к,$$

де $З_п$ – залишок товару на початок досліджуваного періоду,

Н – надходження товару,

В – вибуття товару,

$З_к$ – залишок товару на кінець досліджуваного періоду.

2. При *компонентному зв'язку* зміна статистичного показника відбувається під впливом зміни компонент, які входять до даного показника як множники. Наприклад, виручка дорівнює добутку кількості реалізованої продукції на ціну одиниці даної продукції.

3. *Статистичний зв'язок* показників, при якому кожному значенню однієї змінної відповідає множина можливих значень іншої змінної. Такий зв'язок стає можливим через те, що на досліджуваний (залежний) показник впливає велика кількість неврахованих факторів і, до того ж, при оцінюванні значень факторів, що впливають на даний показник, можливі помилки. Прикладом статистичного зв'язку може бути залежність врожайності зернових від кількості внесених добрив.

Важливе місце в економетричних дослідженнях займає *кореляційна залежність*, при якій кожному значенню однієї змінної відповідає умовне середнє (умовне математичне сподівання) значення іншої.

Умовною середньою величиною $\overline{y_x}$ називається середнє арифметичне значення ознаки *Y*, обчислене для конкретного значення показника *X*.

Рівняння виду:

$$\overline{y_x} = f(x) \tag{3.1}$$

називається рівнянням регресії. При цьому $f(x)$ називається регресією *Y* на *X*, а графік цієї функції – лінією регресії. *Залежну* змінну *Y* називають також *результативною ознакою, результативною функцією, ендогенною змінною,* а незалежну змінну *X* – *факторною ознакою, фактором, пояснювальною, екзогенною змінною та ін.*

Іноді кореляційну залежність можна подати у вигляді модельного рівняння регресії:

$$M_x(Y) = \varphi(x) \text{ або } M_y(X) = \psi(y), \tag{3.2}$$

де $\varphi(x) \neq const$, $\psi(y) \neq const$.

Для точного опису рівняння регресії необхідно знати умовний закон розподілу залежної змінної *Y* за умови, що змінна *X* набуде значення x, тобто $X = x$. У статистичній практиці таку інформацію отримати досить складно, оскільки, зазвичай, дослідник використовує лише вибірку пар значень (x_i, y_i) обмеженого обсягу n. У цьому випадку мова може йти лише про оцінку (апроксимацію) за вибіркою функції регресії. Такою оцінкою є вибіркова лінія (крива) регресії:

$$\overline{y_x} = \overline{f}(x, b_0, ..., b_p), \tag{3.3}$$

де $\overline{y_x}$ – умовна середня змінної *Y* при фіксованому значенні змінної $X = x$, $b_0, ..., b_p$ – параметри кривої. Рівняння (3.3) називають вибірковим рівнянням регресії.

За правильно визначеної апроксимованої функції $\overline{f}(x, b_0, ..., b_p)$ зі збільшенням обсягу вибірки вона буде збігатися за ймовірністю до функції регресії $f(x)$.

3.2 Однофакторна лінійна регресія: побудова та оцінювання параметрів

Прості лінійні регресійні моделі встановлюють залежність між двома змінними: факторною і результативною ознаками. Наприклад, між витратами на відпустку і складом сім'ї, між витратами на рекламу і обсягом реалізованої продукції.

У загальному вигляді **проста вибіркова лінійна** регресія записується так:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,, \tag{3.4}$$

де y_i – множина спостережень за змінною $y_i = \{y_{i1},..., y_{in}\}$;

x_i – множина спостережень за факторною ознакою $x_i = \{x_{i1},..., x_{in}\}$;

ε_i – множина помилок $\varepsilon_i = \{\varepsilon_{i1},..., \varepsilon_{in}\}$;

β_0 та β_1 – істинні параметри зв'язку, значення яких нам невідомі.

Регресійна модель називається **лінійною**, якщо вона лінійна за своїми параметрами. Модель (3.4) називають узагальненою парною лінійною кореляційно-регресійною моделлю.

Параметр β_1 називають коефіцієнтом регресії, параметр β_0 називають вільним членом кореляційно-регресійної моделі. Випадкова величина ε містить різноманітні стохастичні збурення, помилки спостереження та вимірювання, елементи випадковості людської реакції тощо.

Наше завдання – на основі заданих статистичних значень змінних x та y отримати оцінки b_0 та b_1 параметрів β_0 та β_1 (в сенсі певного критерію), тобто побудувати рівняння регресії

$$\overline{y_x} = b_0 + b_1 x, \tag{3.5}$$

де $\overline{y_x}$ – теоретичне (нормативне, прогнозне) значення результативної змінної y.

Приклад 3.1. Бюро економічного аналізу фабрики "ROSHEN" оцінює ефективність відділу маркетингу з продажу цукерок і тістечок. Для такого оцінювання є досвід праці у 5 географічних зонах з майже однаковими умовами (потенційні клієнти, ставлення до товарного знака і т. п.). У цих зонах зафіксовано протягом однакового періоду обсяги продажу (млн. коробок), витрати (млн. грн) фірми на просування товару на ринку. Вихідні дані наведено в табл. 3.1.

Реальні спостереження y_i зобразимо точками у системі координат (x_i, y_i) на рис. 3.1. Візуально можна припустити, що між даними є лінійна залежність, тобто їх можна апроксимувати прямою лінією.

Таблиця 3.1 – Вихідні дані для прикладу 3.1

№	y_i	x_i
1	25	5
2	30	6
3	35	9
4	45	12
5	65	18

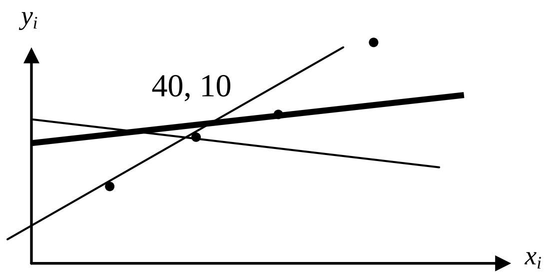


Рисунок 3.1 – Відхилення реальних даних від оцінюваних

Взагалі існує безліч прямих, які можна провести через множину спостережуваних точок. Для цього треба скористатися критерієм, який дасть змогу обрати "найкращу" з них, з точки зору цього критерію. Найпоширенішим є *критерій мінімізації суми квадратів відхилень.*

На рисунку будь-яка з прямих, яку можна провести через дані точки, має точки, які знаходяться над прямою і під нею. Встановимо відхилення від безпомилкової прямої так:

$$e_i = y_i - \overline{y}_i = y_i - b_0 - b_1 x_i, \tag{3.6}$$

де $\overline{y}_i$ – i-та точка на прямій, яка відповідає значенню x_i.

Відхилення або помилки називають ще залишками. Зрозуміло, що логічно обирати таку пряму, щоб сума квадратів помилок була мінімальною.

У цьому полягає критерій найменших квадратів: невідомі параметри b_0 та b_1 визначаються таким чином, щоб мінімізувати суму квадратів помилок:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = Q(b_0, b_1) \to min \tag{3.7}$$

Дослідимо функцію $Q(b_0, b_1)$ на мінімум. Необхідною умовою екстремуму є

$$\frac{\partial Q(b_0, b_1)}{\partial b_0} = 0; \quad \frac{\partial Q(b_0, b_1)}{\partial b_1} = 0. \tag{3.8}$$

У нашому випадку маємо:

$$\frac{\partial Q(b_0, b_1)}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0;$$

(3.9)

$$\frac{\partial Q(b_0,b_1)}{\partial b_1} = -2\sum_{i=1}^{n} x_i \cdot \left(y_i - b_0 - b_1 x_i\right) = 0,$$

звідки отримаємо систему лінійних рівнянь:

$$\begin{cases} \sum_{i=1}^{n} y_i = n \cdot b_0 + b_1 \cdot \sum_{i=1}^{n} x_i\ ; \\ \sum_{i=1}^{n} y_i x_i = b_0 \cdot \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^{\ 2}. \end{cases} \tag{3.10}$$

Систему (3.10) називають *системою нормальних рівнянь* для знаходження оцінених коефіцієнтів рівняння регресії.

Виразимо b_0 з першого рівняння і підставимо у друге, отримаємо b_1:

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}. \tag{3.11}$$

З метою спрощення помножимо чисельник і знаменник на $1/n$:

$$b_1 = \frac{\frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right) - \overline{x}\,\overline{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2}\ , \quad \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Вираз (3.11) можна записати ще так:

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\cdot\left(y_i - \overline{y}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} = \frac{cov(x,y)}{var(x)}. \tag{3.12}$$

Доведемо це:

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\cdot\left(y_i - \overline{y}\right) = \sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\sum_{i=1}^{n} y_i - \overline{y}\cdot\sum_{i=1}^{n} x_i + n\overline{x}\,\overline{y} = \sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}\,;$$

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\sum_{i=1}^{n}\left(x_i{}^2-2x_i\bar{x}+\bar{x}^2\right)=\sum_{i=1}^{n}x_i{}^2-2n\bar{x}^2+n\bar{x}^2=\sum_{i=1}^{n}x_i{}^2-n\bar{x}^2.$$

Чисельник (3.12) є коефіцієнтом коваріації між x та y. За означенням, **коефіцієнт коваріації** визначається за формулою:

$$\operatorname{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\cdot\left(y_i-\bar{y}\right)=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\bar{x}\cdot\bar{y}. \qquad (3.13)$$

Коваріація (cov) – це абсолютна міра зв’язку між двома величинами.

Знаменник (1.12) є **дисперсією** величини x, тобто:

$$D(X)=\operatorname{var}(x)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\frac{1}{n}\sum_{i=1}^{n}x_i{}^2-\bar{x}^2. \qquad (3.14)$$

Отже, кут нахилу прямої регресії можна встановити за формулами (3.11) чи (3.12).

Із першого рівняння системи (3.10) знайдемо параметр b_o:

$$\frac{1}{n}\sum_{i=1}^{n}y_i-b_0-b_1\frac{1}{n}\sum_{i=1}^{n}x_i=0 \quad \Rightarrow \quad b_0=\bar{y}-b_1\bar{x}.$$

Таким чином, точка (b_0, b_1) є точкою, “підозрілою” на екстремум.

Перевіримо, що дана точка є точкою мінімуму. Достатньою умовою мінімуму функції Q (b_0, b_1) є:

$$\Delta=\begin{vmatrix}\frac{\partial^2Q(b_0,b_1)}{\partial b_1^2} & \frac{\partial^2Q(b_0,b_1)}{\partial b_1\,\partial b_0}\\ \frac{\partial^2Q(b_0,b_1)}{\partial b_0\partial b_1} & \frac{\partial^2Q(b_0,b_1)}{\partial b_0^2}\end{vmatrix}>0 \quad \text{та} \quad \frac{\partial^2Q(b_0,b_1)}{\partial b_1^2}>0.$$

У нашому випадку маємо:

$$\frac{\partial^2Q(b_0,b_1)}{\partial b_1^2}=2\sum_{i=1}^{n}x_i^2>0;$$

для неперервних функцій $\frac{\partial^2Q(b_0,b_1)}{\partial b_1\,\partial b_0}=\frac{\partial^2Q(b_0,b_1)}{\partial b_0\partial b_1}$, тому

$$\frac{\partial^2 Q(b_0, b_1)}{\partial b_0 \partial b_1} = 2\sum_{i=1}^{n} x_i \ ;$$

$$\frac{\partial^2 Q(b_0, b_1)}{\partial b_0^2} = 2n \,.$$

$$\Delta = \begin{vmatrix} 2\sum\limits_{i=1}^{n} x_i^2 & 2\sum\limits_{i=1}^{n} x_i \\ 2\sum\limits_{i=1}^{n} x_i & 2n \end{vmatrix} = 4\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right) = \left\{ \begin{matrix} \textit{домножимо} \quad \textit{і} \quad \textit{поділимо} \\ \textit{даний} \quad \textit{вираз} \quad \textit{на} \quad n^2 \end{matrix} \right\} =$$

$$= 4n^2 \left(\frac{\sum\limits_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum\limits_{i=1}^{n} x_i}{n} \right)^2 \right) = 4n^2 \left(\frac{\sum\limits_{i=1}^{n} x_i^2}{n} - \overline{x}^2 \right) = 4n^2 \frac{\sum\limits_{i=1}^{n} \left(x_i - \overline{x} \right)^2}{n} = 4n \sum_{i=1}^{n} \left(x_i - \overline{x} \right)^2 > 0,$$

якщо існує хоча б одне значення $x_i \neq \overline{x}$.

Приклад 3.2. Припустимо, що ви збираєте дані про витрати родини на відпустку та її доходи і оцінюєте це таким рівнянням:

$$y = 5 + 0{,}86 \cdot x,$$

де y – споживання;

x – доход.

Визначити, що є залежною змінною, а що − незалежною? Поясніть взаємозв’язок між витратами на відпустку родини і її доходом. Наскільки зростуть витрати, якщо дохід зросте на одиницю?

Розв’язання. Оскільки витрати на відпустку залежать від доходу родини, то витрати є залежною, а дохід – незалежною змінною.

Дане рівняння свідчить про те, що витрати складаються з двох частин. Перша частина подана перетином b_0 = 5, що означає розмір витрат $\overline{y}$ = 5 при відсутності доходу. Друга частина складається з $0{,}86 \cdot x$, тобто зростання доходу на одну одиницю зумовить зростання витрат на відпустку на 0,86.

Приклад 3.3. На підставі даних про валовий регіональний продукт та величину основних засобів по регіонах центру України за 2011 рік (табл. 3.2) побудувати лінійну кореляційно-регресійну модель, яка описує залежність величини валового регіонального продукту від величини основних засобів.

Таблиця 3.2 – Дані для прикладу 3.3

№	Область	Валовий регіональний продукт у фактичних цінах, млн. грн	Основні засоби у фактичних цінах, млн. грн
1	Вінницька	8123	25993
2	Дніпропетровська	30040	128686
3	Кіровоградська	5594	19729
4	Полтавська	13983	52493
5	Черкаська	6623	23238

Розв'язання. Для побудови системи нормальних рівнянь (3.10) доцільно скласти допоміжну таблицю (табл. 3.3).

Таблиця 3.3 – Допоміжна таблиця для системи нормальних рівнянь

№	x_i	y_i	x_i^2	$x_i y_i$
1	25993	8123	675636049	211141139
2	128686	30040	16560086596	3865727440
3	19729	5594	389233441	110364026
4	52493	13983	2755515049	734009619
5	23238	6623	540004644	153905274
Разом	250139	64363	2092045779	5075147498

Таким чином маємо:

$$\begin{cases} 64363 = 5 \cdot b_0 + b_1 \cdot 250139; \\ 5075147498 = b_0 \cdot 250139 + b_1 \cdot 2092045779. \end{cases}$$

Розв'язавши одержану систему методом Крамера, знаходимо:

$$b_0 = 21778{,}112276 \ ; \quad b_1 = -0{,}178011 \ .$$

Отже, лінійна кореляційно-регресійна модель, яка описує залежність величини валового регіонального продукту від величини основних засобів, така:

$$\overline{y_x} = 21778{,}112276 - 0{,}178011 x_i \, .$$

3.3 Основні припущення класичного кореляційного аналізу

Розглянемо припущення, які є основою класичного кореляційно-регресійного аналізу (згідно з методом найменших квадратів).

1. Вектор помилок ε_i будемо називати збуренням.

2. Математичне сподівання збурення дорівнює нулю:

$$M(\varepsilon_i) = 0\,. \tag{3.15}$$

(або математичне сподівання залежної змінної y_i дорівнює лінійній функції регресії: $M(y_i) = \beta_0 + \beta_1 x_i$).

3. Дисперсія збурення ε_i (або залежної змінної y_i) є сталою для довільного i:

$$D(\varepsilon_i) = \sigma^2 \tag{3.16}$$

або $D(y_i) = \sigma^2$ – умова гомоскедастичності чи рівнозмінності збурень (залежної змінної).

4. Збурення ε_i та ε_j некорельовані:

$$M(\varepsilon_i \varepsilon_j) = 0, (i \neq j) \tag{3.17}$$

або

$$cov(\varepsilon_i, \varepsilon_j) = 0\,. \tag{3.18}$$

5. Незалежність між значеннями випадкової величини ε і значеннями незалежної змінної x:

$$cov(\varepsilon_i, x_i) = 0,\; i = \overline{1,n}\,. \tag{3.19}$$

6. Випадкова величина ε_i розподілена за нормальним законом з математичним сподіванням, рівним нулю, та дисперсією σ^2.

7. Якщо параметри вибіркової лінійної кореляційно-регресійної моделі розраховані методом найменших квадратів за зроблених вище припущень, то

$$M(b_0) = \beta_0,$$

$$M(b_1) = \beta_1.$$

8. Регресійну модель специфіковано правильно (обрано правильну функціональну форму моделі).

Економетричне дослідження містить етап специфікації моделі, яка має бути адекватною економічному об’єкту, процесу, явищу, що вивчається. При специфікації моделі потрібно з’ясувати такі запитання:

1) які змінні потрібно вносити в модель;

2) якою повинна бути функціональна форма моделі: лінійною чи нелінійною, якщо нелінійною, то якою: степеневою, показниковою тощо;

3) які можливі припущення щодо змінних *x*, *y*, *ε* можна зробити в моделі?

На етапі специфікації економетричної моделі потрібно залучати експертів або проводити послідовні економетричні дослідження для вдосконалення моделі.

Лема 3.1. Якщо під час кореляційно-регресійного аналізу виконуються перераховані вище припущення, то залежна змінна y має нормальний розподіл з математичним сподіванням

$$M(y_i) = \beta_0 + \beta_1 x_i$$

та дисперсією

$$D(y_i) = \sigma^2.$$

Доведення

$$M(y_i) = M(\beta_0 + \beta_1 x_i + \varepsilon_i) = M(\beta_0) + M(\beta_1 x_i) + M(\varepsilon_i) = \beta_0 + \beta_1 x_i,$$

оскільки β_0, β_1 – константи, а за припущенням $M(\varepsilon_i) = 0$.

$$D(y_i) = M\left(y_i - M(y_i)\right)^2 = M(\beta_0 + \beta_1 x_i + \varepsilon_i - \beta_0 - \beta_1 x_i)^2 = M(\varepsilon_i)^2 = \sigma^2.$$

Лема 3.2. Якщо параметри лінійної кореляційно-регресійної моделі (1.5) розраховано з використанням методу найменших квадратів за припущень 1–8, то

$$D(b_0) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2};$$

$$D(b_1) = \frac{\sigma^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}.$$

Доведення

Дисперсію параметра b_1 визначимо за формулою

$$D(b_1)=M\left(b_1-M(b_1)\right)^2=M(b_1-\beta_1)^2 .$$

Позначивши $\alpha_i=\dfrac{x_i-\overline{x}}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}$, можна записати $b_1=\sum\limits_{i=1}^{n}\alpha_i y_i$. Тоді

$$D(b_1)=D\left(\sum_{i=1}^{n}\alpha_i y_i\right)=\sum_{i=1}^{n}\alpha_i^2 D(y_i)=\sum_{i=1}^{n}\alpha_i^2\sigma^2 .$$

Знайдемо $\sum\limits_{i=1}^{n}\alpha_i^2$:

$$\sum_{i=1}^{n}\alpha_i^2=\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\right)^2=\frac{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^2}{\left(\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^2\right)^2}=\frac{1}{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^2} .$$

Остаточно формулу для обчислення дисперсії параметра b_1 подають так:

$$D(b_1)=\frac{\sigma^2}{\sum\limits_{i=1}^{n}\left(x_i-\overline{x}\right)^2} .$$

Дисперсію параметра b_0 визначимо за формулою

$$D(b_0)=M\left(b_0-M(b_0)\right)^2=M(b_0-\beta_0)^2 .$$

Оскільки $b_0=\overline{y_x}-b_1\overline{x}=\overline{y_x}-\overline{x}\sum\limits_{i=1}^{n}\alpha_i y_i=\dfrac{\sum\limits_{i=1}^{n}y_i}{n}-\overline{x}\sum\limits_{i=1}^{n}\alpha_i y_i=\sum\limits_{i=1}^{n}\left(\dfrac{1}{n}-\overline{x}\alpha_i\right)y_i$,

то

$$D(b_0)=D\left(\sum_{i=1}^{n}\left(\frac{1}{n}-\overline{x}\alpha_i\right)y_i\right)=\sum_{i=1}^{n}\left(\frac{1}{n}-\overline{x}\alpha_i\right)^2 D(y_i)=\sum_{i=1}^{n}\left(\frac{1}{n}-\overline{x}\alpha_i\right)^2\sigma^2 .$$

Обчислимо $\sum\limits_{i=1}^{n}\left(\dfrac{1}{n}-\overline{x}\alpha_i\right)^2$:

$$\sum_{i=1}^{n}\left(\frac{1}{n}-\bar{x}\alpha_i\right)^2=\sum_{i=1}^{n}\left(\frac{1}{n^2}-\frac{2\bar{x}\alpha_i}{n}+\bar{x}^2\alpha_i^2\right)=\frac{1}{n}-\frac{2\bar{x}}{n}\sum_{i=1}^{n}\alpha_i+\bar{x}^2\sum_{i=1}^{n}\alpha_i^2=$$

$$=\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+n\bar{x}^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{\sum_{i=1}^{n}\left(x_i^2-2\bar{x}x_i+\bar{x}^2\right)+n\bar{x}^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=$$

$$=\frac{\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2+n\bar{x}^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}=\frac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}.$$

Остаточна формула для обчислення дисперсії параметра b_0 така:

$$D(b_0)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}.$$

Виникає природне запитання, чи є оцінки b_0 та b_1 параметрів β_0 та β_1 найкращими? Відповідь на це запитання дає теорема Гаусса-Маркова: *якщо регресійна модель задовольняє припущення 1-5, то оцінки b_0 та b_1 мають найменшу дисперсію в класі усіх лінійних несуміщених оцінок.*

Таким чином, в певному сенсі, оцінки b_0 та b_1 є найбільш ефективними лінійними оцінками параметрів β_0 та β_1.

3.4 Елементи дисперсійного аналізу. Поняття про ступені вільності

У математичній статистиці дисперсійний аналіз розглядають як самостійний метод статистичного аналізу, а в економетрії він застосовується як допоміжний засіб для вивчення якості регресійної моделі. Згідно з основною ідеєю дисперсійного аналізу

$$\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2=\sum_{i=1}^{n}\left[\left(\overline{y_x}-\bar{y}\right)+\left(y_i-\overline{y_x}\right)\right]^2=\sum_{i=1}^{n}\left(\overline{y_x}-\bar{y}\right)^2+\sum_{i=1}^{n}\left(y_i-\overline{y_x}\right)^2+$$
$$+2\sum_{i=1}^{n}\left(\overline{y_x}-\bar{y}\right)\left(y_i-\overline{y_x}\right) \tag{3.20}$$

або

$$Q=Q_R+Q_e, \tag{3.21}$$

де $Q=\sum_{i=1}^{n}(y_i-\overline{y})^2$ – загальна сума квадратів відхилень залежної змінної від середнього значення;

$Q_R=\sum_{i=1}^{n}(\overline{y_x}-\overline{y})^2$ – сума квадратів відхилень, зумовлена регресією;

$Q_e=\sum_{i=1}^{n}(y_i-\overline{y_x})^2$ – залишкова сума квадратів, що характеризує невраховані фактори.

Переконаймося, що пропущений у (3.21) третій доданок $Q_3=2\sum_{i=1}^{n}\left(\overline{y_x}-\overline{y}\right)\left(y_i-\overline{y_x}\right)$ дорівнює нулю. Враховуючи, що $b_0=\overline{y}-b_1\overline{x}$ та $\overline{y_x}=\overline{y}+b_1(x-\overline{x})$, маємо:

$$\overline{y_x}-\overline{y}=b_1(x_i-\overline{x});$$

$$y_i-\overline{y_x}=y_i-b_0-b_1x_i=y_i-\left(\overline{y}-b_1\overline{x}\right)-b_1x_i=\left(y_i-\overline{y}\right)-b_1(x_i-\overline{x});$$

$$Q_3=2\sum_{i=1}^{n}\left(\overline{y_x}-\overline{y}\right)\left(y_i-\overline{y_x}\right)=2b_1\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})-2b_1^2\sum_{i=1}^{n}(x_i-\overline{x})^2=0$$

(враховуючи рівність (3.12)).

Результати дисперсійного аналізу можна подати у вигляді таблиці (табл. 3.4)

Із таблиці 3.4 видно, що кожна сума квадратів пов’язана з числом, яке називають ***ступенем вільності***. Це число показує, скільки незалежних елементів інформації, що утворилися з елементів y_1, y_2, …, y_n, потрібно для розрахунку даної суми квадратів.

У статистиці числом ступенів вільності певної величини часто називають різницю між кількістю різних дослідів і кількістю констант, встановлених в результаті цих дослідів, незалежно один від одного.

Таблиця 3.4 – Дані для дисперсійного аналізу

Сума квадратів	Число ступенів вільності	Середні квадрати
$Q_R=\sum_{i=1}^{n}(\overline{y_x}-\overline{y})^2$	$m-1$	$s_R^2=\dfrac{Q_R}{m-1}$
$Q_e=\sum_{i=1}^{n}(\overline{y_x}-y_i)^2$	$n-m$	$s^2=\dfrac{Q_e}{n-m}$
$Q=\sum_{i=1}^{n}(y_i-\overline{y})^2$	$n-1$	

де n – число дослідів;

m – число параметрів в рівнянні регресії.

За відсутності лінійної залежності між залежною та пояснювальними змінними випадкові величини $s_R^2 = \frac{Q_R}{m-1}$ та $s^2 = \frac{Q_e}{n-m}$ мають χ^2-розподіл відповідно з $(m-1)$ та $(n-m)$ ступенями вільності, а їх відношення – F-розподіл із тими ж ступенями вільності. Тому рівняння регресії є значимими на рівні α, якщо спостережуване значення статистики більше за табличне значення F-критерію Фішера-Снедекора

$$F = \frac{Q_R(n-m)}{Q_e(m-1)} = \frac{s_R^2}{s^2} > F_{\alpha;k_1;k_2}, \tag{3.22}$$

де $F_{\alpha;k_1;k_2}$ – табличне значення F-критерію Фішера-Снедекора (додаток Д), визначене на рівні значимості α при $k_1 = m-1$ та $k_2 = n-m$ ступенях вільності. Можна сказати, що значення F показує, якою мірою регресія краще оцінює значення залежної змінної порівняно з її середньою.

3.5 Коефіцієнти кореляції та детермінації. Критерій Фішера

Щільність зв'язку між ознаками оцінюють за допомогою коефіцієнта кореляції, який обчислюється за формулою:

$$r_{y/x} = \pm \sqrt{\frac{\sum_{i=1}^{n}(\overline{y_x} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{3.23}$$

де $\overline{y_x}$ – умовна середня ознаки Y;

$\bar{y}$ – загальне середнє арифметичне ознаки Y;

y_i – емпіричні (дослідні) значення ознаки Y (знак «–» в формулі ставиться при зворотному зв'язку).

На практиці досить часто для обчислення коефіцієнта кореляції використовують і такі підходи:

$$r_{y/x} = \frac{cov(x,y)}{s_x s_y} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} \qquad (3.24)$$

або

$$r_{y/x} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \qquad (3.25)$$

Властивості коефіцієнта кореляції

1. Якщо $r_{y/x} > 0$, то кореляційний зв'язок прямий, при якому збільшення однієї змінної приводить до збільшення іншої змінної.
2. Якщо $r_{y/x} < 0$, то кореляційний зв'язок зворотний, при якому збільшення однієї величини приводить до зменшення іншої.
3. Якщо $r_{y/x} = \pm 1$, то кореляційна залежність лінійна, тобто всі емпіричні точки лежать на прямій.
4. $r_{y/x} \in [-1,1]$.
5. Якщо $r_{y/x} = 0$, то лінійний зв'язок відсутній і лінія регресії паралельна осі *Ох*.

Кореляційний зв'язок між ознаками може бути слабким і сильним (щільним). Він оцінюється за шкалою Чеддока таким чином:

$0{,}1 < r_{y/x} < 0{,}3$: слабкий;

$0{,}3 < r_{y/x} < 0{,}5$: помірний;

$0{,}5 < r_{y/x} < 0{,}7$: помітний;

$0{,}7 < r_{y/x} < 0{,}9$: щільний;

$0{,}9 < r_{y/x} < 1$: дуже щільний.

Приклад 3.4. За даними прикладу 3.3 оцінити силу кореляційного зв'язку між величиною валового регіонального продукту та величиною основних засобів.

Розв'язання. Для оцінювання сили кореляційного зв'язку між величиною валового регіонального продукту та величиною основних засобів використаємо формулу (3.23), де $\overline{y_x} = 21778{,}112276 - 0{,}178011 x_i$, як було одержано в прикладі 3.3, $\overline{y} = \frac{64363}{5} = 12872{,}6$.

Для зручності розрахунку складемо допоміжну таблицю (табл. 3.5)

Таблиця 3.5 – Допоміжні дані для розрахунку коефіцієнта кореляції

№	x_i	y_i	$\overline{y_x}$	$\left(\overline{y_x}-\overline{y}\right)^2$	$\left(y_i-\overline{y}\right)^2$
1	25993	8123	17151,07	18305326	22558700
2	128686	30040	-1129,411	196056320	294719623
3	19729	5594	18266,13	29090201	52978018
4	52493	13983	12433,78	192562	1232988
5	23238	6623	17641,49	22742337	39057500
Разом	250139	64363	64363,07	266386746	410546829

Оскільки пряма $\overline{y_x}=21778{,}112276-0{,}178011x_i$ є спадною, то

$$r_{y/x}=-\sqrt{\frac{266386746}{410546829}}=-\sqrt{0{,}65}=-0{,}8.$$

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про тісний зворотний зв’язок між величиною валового регіонального продукту та величиною основних засобів.

Приклад 3.5. Знайти рівняння регресії залежності між видобутком вугілля на одного шахтаря за зміну Y та потужністю пласта X за такими даними, що характеризують процес видобутку вугілля на 10 шахтах (таблиця 3.6).

Таблиця 3.6 – Дані для прикладу 3.5

i	1	2	3	4	5	6	7	8	9	10
x_i	8	11	12	9	8	8	9	9	8	12
y_i	5	10	10	7	5	6	6	5	6	8

Обчислити коефіцієнт кореляції.

Розв’язання. Для знаходження рівняння регресії складемо систему нормальних рівнянь за допомогою допоміжної таблиці (табл. 3.7).

$$\begin{cases}68=10\cdot b_0+94b_1;\\664=94b_0+908b_1.\end{cases}$$

Розв’язавши цю системи, отримаємо

$$b_0=2{,}75,\ b_1=1{,}016\ .$$

Таким чином, рівняння регресії набуває вигляду:

$$\overline{y_x} = 2{,}75 + 1{,}016x\,.$$

Таблиця 3.7 – Допоміжна таблиця для прикладу 3.5

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	8	5	64	40	25
2	11	10	121	110	100
3	12	10	144	120	100
4	9	7	81	63	49
5	8	5	64	40	25
6	8	6	64	48	36
7	9	6	81	54	36
8	9	5	81	45	25
9	8	6	64	48	36
10	12	8	144	96	64
$\sum_{i=1}^{10}$	94	68	908	664	496

Для знаходження коефіцієнта кореляції використаємо формулу (3.25):

$$r_{y/x} = \frac{10 \cdot 664 - 94 \cdot 68}{\sqrt{10 \cdot 908 - 94^2} \cdot \sqrt{10 \cdot 496 - 68^2}} = 0{,}866,$$

тобто зв’язок між змінними досить тісний.

Одним із важливих питань кореляційного аналізу є перевірка значимості рівняння регресії, тобто встановлення адекватності математичної моделі експериментальним даним (реальному об’єкту). Однією з найбільш ефективних оцінок адекватності регресійної моделі є коефіцієнт детермінації, який обчислюється за формулою:

$$D = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q} \qquad (3.26)$$

або у відсотковому вигляді:

$$D = \frac{Q_R}{Q} 100\% \qquad (3.27)$$

і показує, яка частина варіації змінної Y залежить від змінної X, а яка − від неврахованих факторів. Рівняння регресії буде адекватне експериментальним даним, якщо $D \in [0{,}55;\ 1]$.

Відмітимо, що коефіцієнт детермінації є сенс розглядати лише за наявності вільного члена в рівнянні регресії, оскільки лише в цьому випадку рівність (3.21), а отже і рівність (3.26) є правильними. Якщо є відомим коефіцієнт детермінації, то критерій значимості (3.22) рівняння регресії або самого коефіцієнта детермінації можна записати так:

$$F = \frac{Q_R(n-m)}{Q_e(m-1)} = \frac{D(n-m)}{(1-D)(m-1)}. \tag{3.28}$$

У випадку парної лінійної регресії коефіцієнт детермінації дорівнює квадрату коефіцієнта кореляції. Дійсно,

$$D = \frac{Q_R}{Q} = \frac{\sum_{i=1}^{n}\left(\overline{y_x} - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} = \frac{\sum_{i=1}^{n} b_1^2\left(x_i - \overline{x}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} = \frac{b_1^2\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \Big/ n}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \Big/ n} = \frac{b_1^2 s_x^2}{s_y^2} = \left(\frac{b_1 s_x}{s_y}\right)^2 = r_{y/x}^2.$$

Приклад 3.6. За даними прикладу 3.5 знайти коефіцієнт детермінації та пояснити його зміст.

Розв'язання. Використовуючи таблицю 3.6, отримаємо

$$Q = \sum_{i=1}^{10}\left(y_i - \overline{y}\right)^2 = \sum_{i=1}^{10} y_i^2 - \frac{\left(\sum_{i=1}^{10} y_i\right)^2}{10} = 496 - \frac{68^2}{10} = 33{,}6;$$

$$Q_e = \sum_{i=1}^{n}(y_i - \overline{y_x})^2 = 0{,}14 + 2{,}48 + 0{,}31 + 0{,}37 + 0{,}14 + 0{,}39 + 0{,}15 + 1{,}94 + 0{,}39 + 2{,}09 = 8{,}39$$

$$Q_R = Q - Q_e = 33{,}6 - 8{,}39 = 25{,}21.$$

За формулою (3.26) маємо $D = \dfrac{25{,}21}{33{,}6} = 0{,}75$. Одержаний результат означає, що варіація залежної змінної Y – добовий видобуток вугілля на одного шахтаря – на 75% пояснюється зміною X – потужністю пласту.

Якщо $D \in [0{,}45;\ 0{,}55]$, то для встановлення адекватності моделі реальному об'єкту використовують F-критерій Фішера, який застосовується так.

1. Обчислюється F-відношення за формулою (3.28).

2. Задається рівень значимості $\alpha \cdot 100\%$ (зазвичай, $\alpha \cdot 100\%$ становить 5% або 1%). Наприклад, якщо можлива помилка $\alpha \cdot 100\%$ складає 5%, це означає, що ми можемо помилитись не більш як в 5% випадків, а в 95% випадків наші висновки правильні.

3. За статистичними таблицями F-розподілу Фішера обчислюється критичне значення $F_{кр} = F_{(\alpha, k_1, k_2)} = F_{(\alpha, m-1, n-m)}$ (додаток В).

4. Якщо $F > F_{кр}$, то з ризиком помилитися не більше ніж в $\alpha \cdot 100\%$ випадках можна стверджувати, що побудована модель адекватна реальному об'єкту.

Приклад 3.7. За даними прикладів 3.5 та 3.6 оцінити значущість рівняння регресії.

Розв'язання. За формулою (3.28) отримуємо значення F-відношення

$$F = \frac{25{,}21 \cdot (10-2)}{8{,}39} = 24{,}04 .$$

За таблицею F-розподілу (додаток В) $F_{0,05;1;8} = 4{,}20$. Оскільки розраховане значення більше за табличне, то з ймовірністю помилитися не більше ніж у 5% випадків можна стверджувати, що рівняння регресії є значущим.

3.6 Стандартна помилка оцінювання. Оцінювання коефіцієнта кореляції

Зобразимо у декартовій системі координат на площині пряму регресії $\overline{y_x} = b_0 + b_1 x$ і пряму середнього арифметичного значення результативної змінної $y = \overline{y}$ (рис. 3.2).

Із сукупності заданих точок спостережень візьмемо довільну точку $D(x_i, y_i)$ й опустимо з неї перпендикуляр DA на вісь абсцис. Очевидно, що

$$AD = AB + BC + CD$$

або

$$y_i = \overline{y} + \left(\overline{y_{x_i}} - \overline{y}\right) + \left(y_i - \overline{y_{x_i}}\right),$$

звідки

$$y_i - \overline{y} = \left(\overline{y_{x_i}} - \overline{y}\right) + \left(y_i - \overline{y_{x_i}}\right). \qquad (3.29)$$

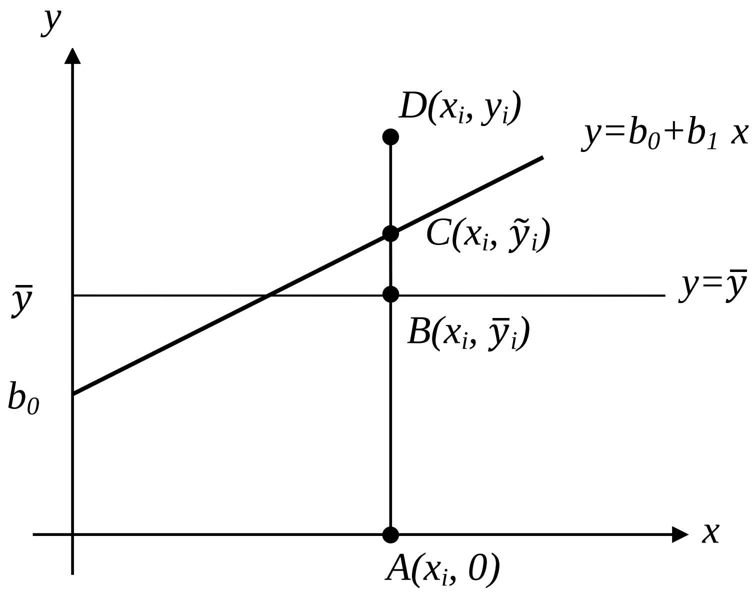


Рисунок 3.2 – Декомпозиція загального відхилення результативної змінної

У рівності (3.29) різницю $y_i - \overline{y}$ називають *загальним відхиленням* результативної змінної, різницю $\overline{y_{x_i}} - \overline{y}$ називають відхиленням, котре можна пояснити з точки зору кореляційно-регресійної моделі (*пояснене відхилення*). Справді, при будь-якому значенні факторної ознаки x завжди можна знайти величину цього відхилення, маючи тільки кореляційно-регресійну модель, бо $\overline{y}$ залишається незмінною величиною. Різницю $y_i - \overline{y_{x_i}}$ (*випадкове відхилення*) називають ще *непоясненим відхиленням,* тому що це відхилення не можна пояснити з точки зору кореляційно-регресійної моделі. Якщо факторна ознака x змінюється, то змінюються обидві величини y_i та $\overline{y_{x_i}}$, тому, маючи тільки кореляційно-регресійну модель, пояснити це відхилення неможливо.

Рівність (3.29) називають *формулою декомпозиції загального відхилення*: загальне відхилення результативної змінної можна розкласти на пояснене відхилення та непояснене відхилення.

Стандартною похибкою кореляційно-регресійної моделі (*стандартною похибкою оцінки за рівнянням регресії*) називають величину

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{y_{x_i}}\right)^2}{n}}. \qquad (3.30)$$

Стандартна похибка моделі характеризує розсіювання фактичних значень результативної змінної в околі теоретичних, знайдених за рівнянням регресії. З геометричної точки зору коефіцієнт кореляції характеризує ступінь розсіяності емпіричних точок навколо теоретичної лінії регресії. Чим ближче значення коефіцієнта кореляції до одиниці, тим щільніше розташовані емпіричні точки до лінії регресії.

Щоб обчислити стандартну похибку парної лінійної кореляційно-регресійної моделі, можна використати формули

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{n} e_i^{\,2}}{n}} \tag{3.31}$$

або

$$s_{yx} = \sqrt{\frac{\sum_{i=1}^{n} y_i^{\,2} - b_0 \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i y_i}{n}}\,. \tag{3.32}$$

Зауважимо, що s_{yx}^2 є зміщеною оцінкою дисперсії випадкових відхилень e_i. Для вибірок малих обсягів використовують незміщену оцінку $S_{yx}^2 = \frac{n}{n-2} s_{yx}^2$.

Якщо взаємозв’язок між змінними x та y функціональний, то всі випадкові відхилення $y_i - \overline{y_{x_i}}$ дорівнюють нулю і, отже, $s_{yx} = 0$.

Якщо кореляційний зв’язок між результативною та факторними змінними відсутній, то $\overline{y_{x_i}} = \overline{y}$ і, отже, $s_{yx} = s_y$, тобто стандартна похибка моделі збігається з середнім квадратичним значенням результативної змінної.

Отже,

$$0 \le s_{yx} \le s_y\,.$$

Усі випадкові величини, які ми оцінювали (випадкові відхилення, параметри b_0 та b_1, теоретичні значення результативної змінної $\overline{y_x}$), мали нормальний закон розподілу (або близький до нього), тому для їх оціню-вання можна було будувати симетричний довірчий інтервал, використо-вуючи таблиці нормального розподілу або розподілу Стьюдента.

Коефіцієнт кореляції загалом не є нормально розподіленою випадко-вою величиною, його областю допустимих значень є інтервал [-1, 1].

Найчастіше відмінність розподілу коефіцієнта кореляції від нормального виражена при тісному зв'язку між змінними, тобто коли коефіцієнт кореляції за абсолютним значенням близький до одиниці (рис. 3.3).

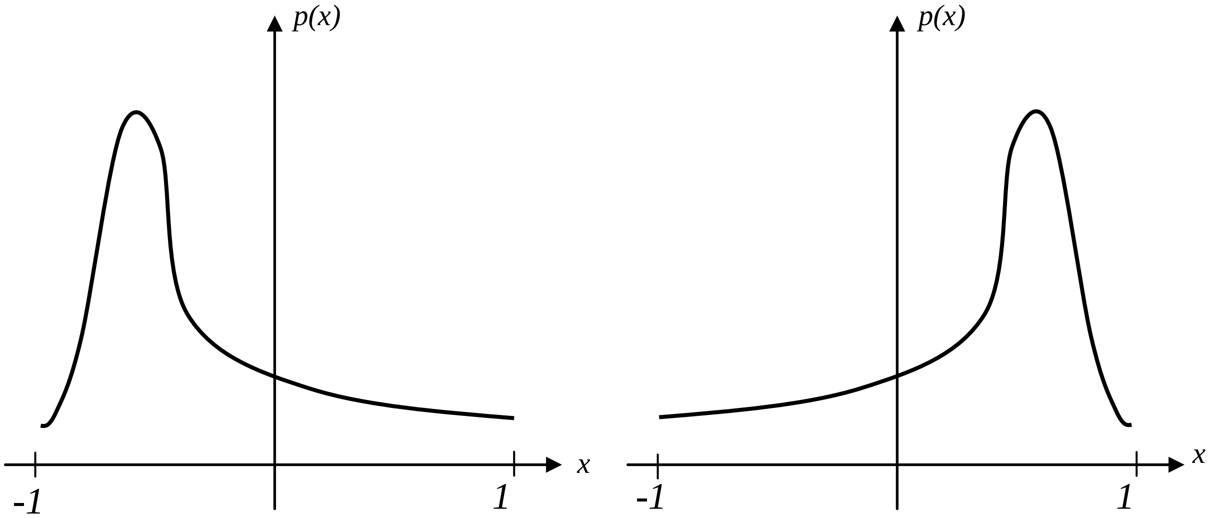


Рисунок 3.3 – Розподіл коефіцієнта кореляції при тісному зв'язку між змінними

Щоб мати змогу оцінити коефіцієнт кореляції у разі, коли його значення не наближене до нуля, Фішер у 1921 році запропонував такий метод.

Спочатку потрібно перейти від випадкової змінної r до випадкової змінної z:

$$z = \frac{1}{2} ln \frac{1+r}{1-r}. \tag{3.33}$$

Фішер довів, що випадкова змінна z розподілена за нормальним законом з математичним сподіванням:

$$E(z) = \frac{1}{2} ln \frac{1+\rho}{1-\rho}, \tag{3.34}$$

де ρ – істинне значення коефіцієнта кореляції для генеральної сукупності, та дисперсією

$$D_z = \frac{1}{n-3}. \tag{3.35}$$

Стандартна вибіркова похибка випадкової змінної z, згідно з (3.35), складає

$$s_z = \frac{1}{\sqrt{n-3}}, \qquad (3.36)$$

а гранична вибіркова похибка z при заданому значенні довірчої ймовірності p:

$$\Delta_z = t_p \cdot s_z.$$

Припустимо, що ξ – невідоме значення випадкової змінної z, яке відповідає істинному значенню коефіцієнта кореляції ρ. Тоді довірчий інтервал для змінної ξ має вигляд

$$z_r - t_p \cdot s_z \le \xi \le z_r + t_p \cdot s_z, \qquad (3.37)$$

де z_r – значення випадкової величини z, яке відповідає, згідно з перетворенням (3.33), вибірковому значенню коефіцієнта кореляції r.

Введемо позначення

$$z_1 = z_r - t_p \cdot s_z, \qquad z_2 = z_r + t_p \cdot s_z.$$

Тоді

$$z_1 \le \xi \le z_2. \qquad (3.38)$$

Здійснивши обернене перетворення від змінної z до змінної r за формулою

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}, \qquad (3.39)$$

отримаємо довірчий інтервал для істинного значення коефіцієнта кореляції ρ генеральної сукупності:

$$r_1 \le \rho \le r_2, \qquad (3.40)$$

де $r_1 = \dfrac{e^{2z_1} - 1}{e^{2z_1} + 1}$, $r_2 = \dfrac{e^{2z_2} - 1}{e^{2z_2} + 1}$.

Якщо значення коефіцієнта кореляції r наближене до нуля, тобто зв'язок між змінними слабкий, то його розподіл наближається до нормального (зазвичай, при великих обсягах вибірки) (рис. 3.4)

У такому разі стандартну похибку коефіцієнта кореляції визначають за формулою

$$s_r = \frac{1 - r^2}{\sqrt{n-1}}. \qquad (3.41)$$

Далі оцінювання істинного значення коефіцієнта кореляції ρ генеральної сукупності проводять за звичайною схемою.

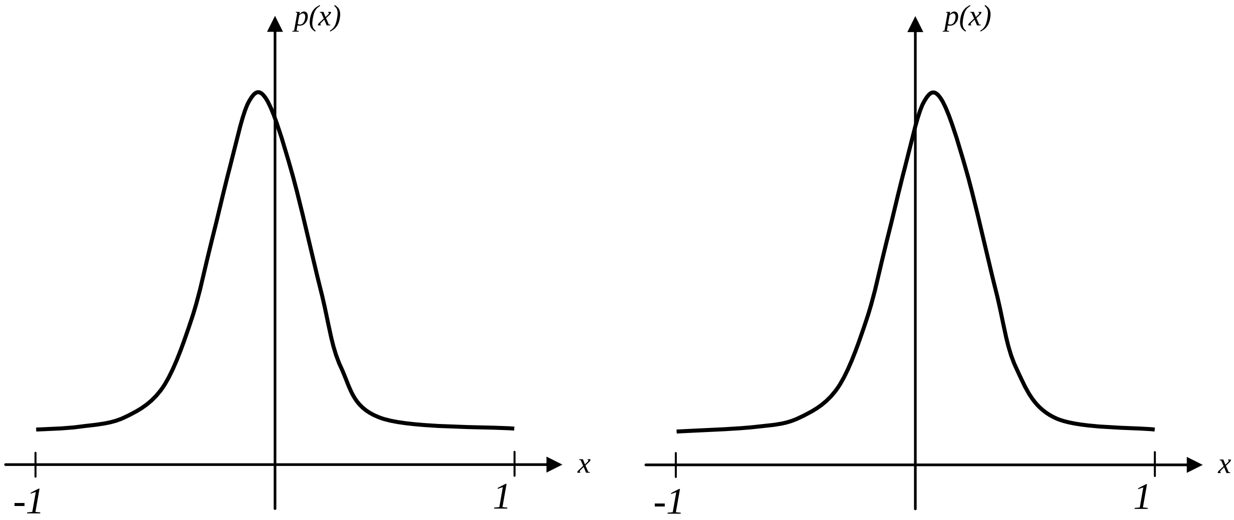


Рисунок 3.4 – Розподіл коефіцієнта кореляції при слабкому зв'язку між змінними

Приклад 3.8. За даними прикладу 3.4 оцінити істинне значення коефіцієнта кореляції між величиною валового регіонального продукту та величиною основних засобів для центральних регіонів України.

Розв'язання. Значення коефіцієнта кореляції, раніше обчислене в прикладі 3.4, дорівнює $r = -0{,}8$, $n = 5$. Отож, для оцінювання коефіцієнта кореляції насамперед перейдемо від випадкової змінної r до випадкової змінної z (3.33):

$$z = \frac{1}{2} ln \frac{1 - 0{,}8}{1 + 0{,}8} = \frac{1}{2} ln\, 0{,}11 \approx -1{,}0985.$$

За формулою (3.36) знайдемо стандартну вибіркову похибку $s_z = \frac{1}{\sqrt{5-3}} = \frac{1}{\sqrt{2}} = 0{,}709$; задаємо ймовірність $p = 0{,}95$ ($\alpha = 1 - p = 0{,}05$) та знаходимо за таблицею (додаток Ж) значення функції Стьюдента $t_p = 3{,}182$; враховуючи, що $\nu = 5 - 2 = 3$.

Гранична вибіркова похибка z при заданому значенні довірчої ймовірності $p = 0{,}95$ становить

$$\Delta_z = t_p \cdot s_z = 3{,}182 \cdot 0{,}709 = 2{,}256038.$$

За формулами (3.38) маємо:

$$z_1 = -1{,}0985 - 2{,}256038 = -3{,}354538\ ;$$

$$z_2 = -1{,}0985 + 2{,}256038 = 1{,}157538\ .$$

Тоді за формулами (3.40) будуємо довірчий інтервал:

$$r_1 = \frac{e^{2z_1} - 1}{e^{2z_1} + 1} = \frac{e^{-2\cdot 3{,}354} - 1}{e^{-2\cdot 3{,}354} + 1} \approx \frac{e^{-6{,}708} - 1}{e^{-6{,}708} + 1} \approx \frac{0{,}00122 - 1}{0{,}00122 + 1} \approx -\frac{0{,}99878}{1{,}00122} \approx -0{,}99\,;$$

$$r_2 = \frac{e^{2z_2} - 1}{e^{2z_2} + 1} = \frac{e^{2\cdot\cdot 1{,}157} - 1}{e^{-2\cdot\cdot 1{,}157} + 1} \approx \frac{e^{2{,}315} - 1}{e^{2{,}315} + 1} \approx \frac{10{,}1397 - 1}{10{,}1397 + 1} \approx \frac{9{,}1397}{11{,}1397} \approx 0{,}82\,;$$

$$-0{,}99 \le \rho \le 0{,}82\,.$$

Отже, з довірчою ймовірністю 0,95 можна стверджувати, що істинне значення коефіцієнта кореляції ρ генеральної сукупності має лежати в межах від -0,99 до 0,82.

Приклад 3.9. За даними прикладу 3.5 оцінити істинне значення коефіцієнта кореляції між видобутком вугілля на одного шахтаря за зміну та потужністю пласта.

Розв'язання. Значення коефіцієнта кореляції, раніше обчислене в прикладі 1.5, дорівнює r = 0,866; n = 10. Отож, для оцінювання коефіцієнта кореляції насамперед перейдемо від випадкової змінної r до випадкової змінної z (3.33):

$$z = \frac{1}{2} ln \frac{1 + 0{,}866}{1 - 0{,}866} = \frac{1}{2} ln 13{,}925 \approx 1{,}317\,.$$

За формулою (3.36) знайдемо стандартну вибіркову похибку $s_z = \frac{1}{\sqrt{10-3}} = \frac{1}{\sqrt{7}} = 0{,}378$; задамо ймовірність p = 0,95 (α = 1 – p = 0,05) та знайдемо за таблицею (додаток Ж) значення функції Стьюдента $t_p = 2{,}306$; враховуючи, що ν = 10 – 2 = 8.

Гранична вибіркова похибка z при заданому значенні довірчої ймовірності $p = 0{,}95$ становить

$$\Delta_z = t_p \cdot s_z = 2{,}306 \cdot 0{,}378 = 0{,}871668\,.$$

За формулами (3.38) маємо:

$$z_1 = 1{,}317 - 0{,}871668 = 0{,}890664\ ;$$

$$z_2 = 1{,}317 + 0{,}871668 = 2{,}188668\ .$$

Тоді за формулами (3.40) будуємо довірчий інтервал:

$$r_1 = \frac{e^{2z_1} - 1}{e^{2z_1} + 1} \approx \frac{e^{0{,}890664} - 1}{e^{0{,}890664} + 1} \approx \frac{2{,}4381 - 1}{2{,}4381 + 1} \approx \frac{1{,}4381}{3{,}4381} \approx 0{,}418;$$

$$r_2 = \frac{e^{2z_2} - 1}{e^{2z_2} + 1} \approx \frac{e^{4{,}377396} - 1}{e^{4{,}377396} + 1} \approx \frac{79{,}8509 - 1}{79{,}8509 + 1} \approx \frac{78{,}8509}{80{,}8509} \approx 0{,}975;$$

$$0{,}418 \le \rho \le 0{,}975.$$

Отже, з довірчою ймовірністю 0,95 можна стверджувати, що істинне значення коефіцієнта кореляції ρ генеральної сукупності має лежати в межах від 0,418 до 0,975.

У ряді прикладних задач потрібно оцінити значущість коефіцієнта кореляції. При цьому виходять з того, що за відсутності кореляційного зв'язку статистика $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ має t-розподіл Стьюдента із $n-2$ ступенями вільності.

Коефіцієнт кореляції є значущим на рівні α, якщо

$$|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > t_{1-\alpha;n-2}, \tag{3.42}$$

де $t_{1-\alpha;n-2}$ – табличне значення t-критерію Стьюдента (додаток Ж), обчислене на рівня значущості α при числі ступенів вільності $n-2$.

3.7 Приклади нелінійної кореляційної залежності

Пошук рівняння регресії – одна з найважливіших проблем кореляційного аналізу. На практиці в економічних дослідженнях іноді потрібно розглядати нелінійні рівняння регресії.

Як і у випадку лінійної регресії для знаходження оцінок параметрів рівнянь регресії потрібно розв'язати відповідні системи нормальних рівнянь. Наведемо деякі типи нелінійних кореляційних рівнянь та відповідних їм систем нормальних рівнянь, де для спрощення запису замість $\sum_{i=1}^{n}$ записано $\sum$.

Параболічна залежність: $\overline{y_x} = a_0 + a_1 x + a_2 x^2$,

$$\begin{cases} na_0 + a_1\sum x_i + a_2\sum x_i^2 = \sum y_i; \\ a_0\sum x_i + a_1\sum x_i^2 + a_2\sum x_i^3 = \sum x_i y_i; \\ a_0\sum x_i^2 + a_1\sum x_i^3 + a_2\sum x_i^4 = \sum x_i{}^2 y_i. \end{cases} \quad (3.43)$$

Гіперболічна залежність: $\overline{y_x} = a_0 + \dfrac{a_1}{x}$,

$$\begin{cases} na_0 + a_1\sum \dfrac{1}{x_i} = \sum y_i; \\ a_0\sum \dfrac{1}{x_i} + a_1\sum \dfrac{1}{x_i^2} = \sum \dfrac{y_i}{x_i}. \end{cases} \quad (3.44)$$

Показникова залежність: $\overline{y_x} = a_0 \cdot a_1{}^x$,

$$\begin{cases} n\ln a_0 + \ln a_1\sum x_i = \sum \ln y_i; \\ \ln a_0\sum x_i + \ln a_1\sum x_i^2 = \sum x_i \ln y_i. \end{cases} \quad (3.45)$$

Степенева залежність: $\overline{y_x} = a_0 \cdot x^{a_1}$,

$$\begin{cases} n\ln a_0 + a_1\sum \ln x_i = \sum \ln y_i; \\ \ln a_0\sum \ln x_i + a_1\sum (\ln x_i)^2 = \sum \ln x_i \ln y_i. \end{cases} \quad (3.46)$$

Зауваження! 1. Якщо позначити $a = ln\,a_0$, $b = ln\,a_1$, то показникову залежність можна звести до лінійної $\bar{z} = a + bx$.

2. Щільність показників також оцінюють за допомогою коефіцієнта кореляції, який можна обчислити за однією з формул, наведених у підрозділі 3.5.

3. Адекватність нелінійної моделі встановлюють за допомогою коефіцієнта детермінації, який обчислюють за формулою (3.27).

До нелінійних моделей приводять дослідження динаміки поведінки економічних систем. П. Самуельсон та В. Фріск визначають динамічну систему так: "…систему називають динамічною, якщо її поведінка в часі визначена функціональними рівняннями, в яких змінні в різні моменти часу присутні в явному вигляді". Такі дослідження дають можливість визначити перспективи розвитку цих систем, виявити можливі резерви, розробити комплекс адаптивних управлінських рішень, які забезпечать ефективне функціонування економічних об’єктів. Дослідженням детермінованої поведінки економічних систем в часі та визначенням оптимальної траєкторії їх розвитку займається економічна динаміка. Трендові, лагові та факторні моделі економічної динаміки є економетричними моделями. В основі динамічного аналізу економічних систем, явищ, процесів лежить поняття траєкторії – функції часу, яка описує стан об’єкта дослідження (значення деякого показника):

$$Q = Q(t), \quad t \in [0, T], \tag{3.45}$$

де $[0, T]$ – скінченний відрізок часу, на якому визначена траєкторія.

У теоретичних моделях досліджують також нескінченні траєкторії та траєкторії, для яких початковий момент часу може бути від’ємним. Під час дослідження траєкторії час *t* можна враховувати дискретно (моментами, інтервалами) або неперервно. Якщо час враховувати дискретно, то моделі економічної динаміки будуть описані скінченно-різницевими рівняннями, якщо ж неперервно – диференціальними рівняннями. *Динамічний (часовий) ряд* – це таблиця значень траєкторії, для якої час змінюється дискретно. За часовою ознакою економічні показники поділяють на моментні та інтервальні. Моментні показники дають кількісну характеристику об’єкта дослідження на деякий момент часу: кількість населення Вінницької області на кінець 2010 року, обсяг основних фондів деякого підприємства на початок 2011 року. Інтервальні показники характеризують об’єкт досліджень за деякий період часу: обсяг продукції, виготовленої деяким підприємством протягом кварталу, прибуток за місяць, виторг торговельного підприємства за один робочий день. Інтервальні показники мають властивість динамічної адитивності, тобто їх можна підсумовувати під час переходу від невеликих проміжків часу до триваліших. Моментні показники не є адитивними в часі. Залежно від показника, значення якого утворює траєкторію, динамічні ряди поділяють на моментні та інтервальні.

У математичній статистиці динамічний ряд розглядають як реалізацію випадкового процесу. У стаціонарних випадкових процесах, для яких характерна рівновага відносно деякого середнього рівня, основні характеристики обчислюють за однією реалізацією процесу (за результатами одного досліду). Проте динамічні ряди економічних показників здебільшого нестаціонарні, їм властиві тенденції, які відображають динамічність економіки. До цих тенденцій належать: нарощування виробничих ресурсів, підвищення науково-технічного рівня, вдосконалення управління системою тощо. Разом з динамічністю, економічним процесам властива інерційність, яка, насамперед, виявляється у вигляді розвитку (напрямок, темпи, коливання).

У складі динамічного ряду можна виділити такі компоненти.

1. Головна (вікова) тенденція або тренд.
2. Регулярні коливання відносно тренда (цикли).
3. Сезонні коливання відносно тренда.
4. Випадкова компонента, яка відображає вплив факторів стохастичного характеру.

Одним із найважливіших завдань дослідження економічної динаміки є встановлення загальної закономірності (тенденції) розвитку. Для виконання цього завдання використовують різноманітні методи зменшення коливальності ряду, серед яких можна виділити дві групи:

- згладжування ряду за допомогою середніх (методи згладжування);
- аналітичне вирівнювання ряду (моделі тренда).

Детальніше розглянемо аналітичне вирівнювання динамічного ряду – метод вираження головної тенденції розвитку у вигляді функції показника від часу. Таку функцію називають *моделлю тренда*. Динамічний ряд в межах періоду зі стабільними умовами розвитку має деяку закономірність динаміки – головну тенденцію. Різні економічні процеси або один і той самий процес, але у різні періоди свого розвитку, можуть суттєво відрізнятись за характеристиками розвитку. За основу типізації економічного розвитку використовують динаміку абсолютних приростів.

Базовий абсолютний приріст розраховують як різницю поточного та базового (початкового) рівнів динамічного ряду:

$$\delta_{t/0} = Q_t - Q_0 .$$

Ланцюговий абсолютний приріст розраховують як різницю поточного та базового (попереднього) рівнів динамічного ряду:

$$\delta_{t/t-1} = Q_t - Q_{t-1} .$$

Абсолютний приріст за одиницю часу характеризує швидкість динаміки. Знак абсолютного приросту засвідчує напрямок динаміки: якщо абсолютний приріст додатний, то значення рівнів ряду зростають, а якщо від’ємний, то спадають.

Базовий темп зростання розраховують як відношення поточного до базового (початкового) рівня динамічного ряду:

$$\eta_{t/0} = \frac{Q_t}{Q_0} .$$

Ланцюговий темп зростання розраховують як відношення поточного до базового (попереднього) рівня динамічного ряду:

$$\eta_{t/t-1} = \frac{Q_t}{Q_{t-1}}.$$

Темп зростання характеризує інтенсивність динаміки. Темп зростання може бути виражений у відсотках або числом, тоді його називають ще коефіцієнтом зростання.

Базовий темп приросту розраховують як відношення базового абсолютного приросту для поточного часового періоду до базового (початкового) рівня динамічного ряду:

$$\rho_{t/0} = \frac{Q_t - Q_0}{Q_0} = \frac{\delta_{t/0}}{Q_0}.$$

Ланцюговий темп приросту розраховують як відношення ланцюгового абсолютного приросту для поточного часового періоду до базового (попереднього) рівня динамічного ряду:

$$\rho_{t/t-1} = \frac{Q_t - Q_{t-1}}{Q_{t-1}} = \frac{\delta_{t/t-1}}{Q_{t-1}}.$$

Темп приросту характеризує відносну швидкість, тобто прискорення динаміки. Темп приросту в прикладних застосуваннях виражають у відсотках.

Можна виділити щонайменше чотири типи економічного розвитку (рис. 3.5 − 3.8):

1. *Рівномірний (постійний, стабільний) розвиток* – характеризується постійним або близьким до нього абсолютним приростом.

2. *Прискорений розвиток* – характеризується абсолютним приростом, який з плином часу збільшується (для спадання – за абсолютною величиною).

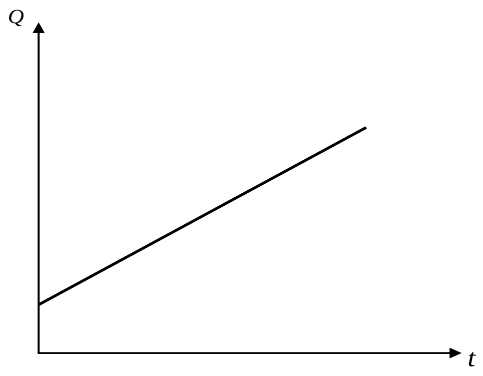
3. *Уповільнений розвиток* – характеризується абсолютним приростом, який з плином часу зменшується (для спадання – за абсолютною величиною).

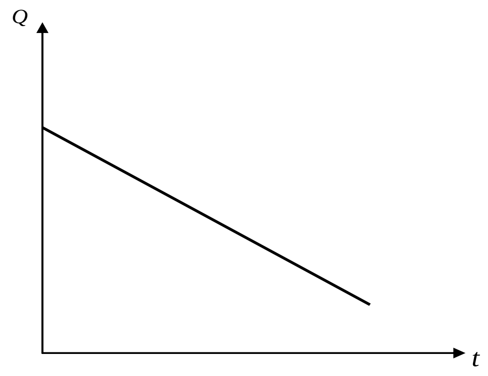
4. *Розвиток з якісною зміною характеристик динаміки протягом розглядуваного періоду часу.*

Зауваження! 1. Для спрощення викладення матеріалу будемо вважати, що часовим інтервалом у динамічному ряді є один рік.

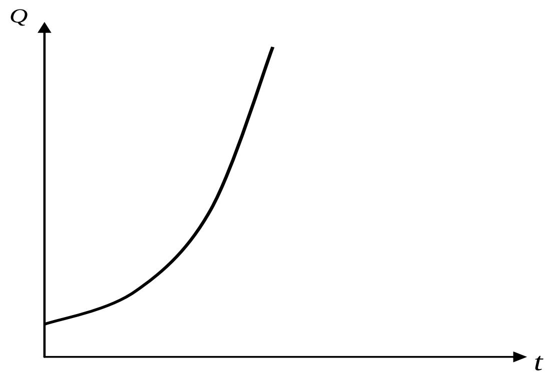
2. Для всіх трендових моделей вважатимемо, що $t \geq 0$.

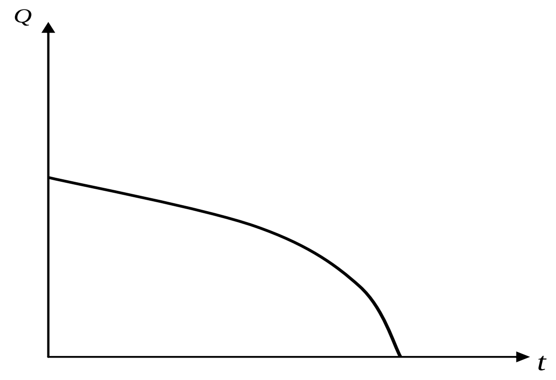
3. Для теоретичних досліджень важливою є гладкість моделі тренда. З огляду на це, обираючи трендові моделі, перевагу надають диференційованим функціям.



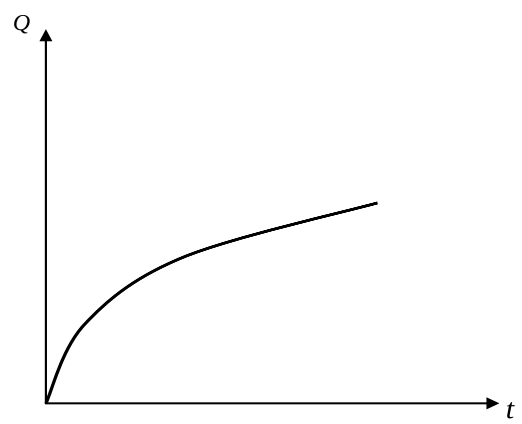


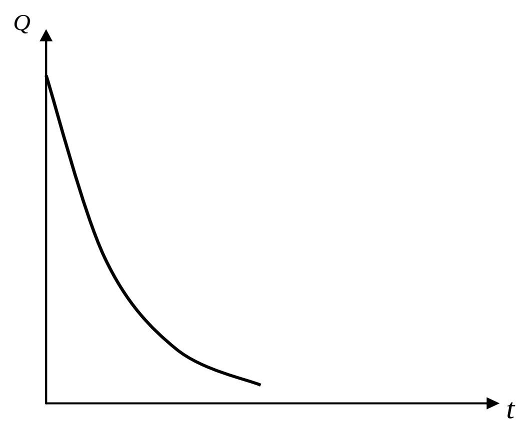
а) Рівномірне зростання б) Рівномірне спадання





а) Прискорене зростання б) Прискорене спадання





а) Уповільнене зростання б) Уповільнене спадання

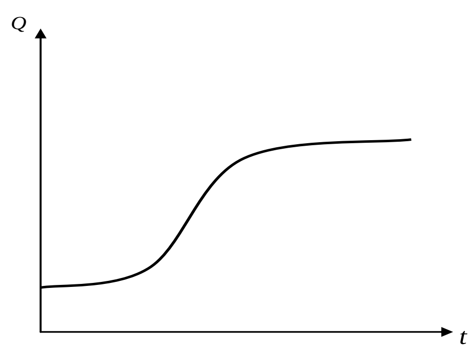


Рисунок 3.8 − Розвиток зі зміною характеристик динаміки

4. На відміну від фактичної траєкторії динаміки $Q(t)$ модель тренда будемо позначати $x(t)$, а замість фактичного значення показника динаміки ряду Q_t будемо використовувати позначення x_t.

Розглянемо деякі приклади нелінійних моделей.

Найчастіше прискорений розвиток описують показниковою та експоненціальною моделями тренда. Зокрема, показникова модель тренда

$$x(t) = a(1+b)^t, \quad a > 0, \tag{3.47}$$

де a – теоретичний початковий рівень,

$b = \rho$ – дискретний темп росту, використовують для опису дискретних процесів.

Експоненціальна модель тренда

$$x(t) = ae^{bt}, \quad a > 0, \tag{3.48}$$

де a – теоретичний початковий рівень,

$b = \tilde{\rho}$ – неперервний темп приросту:

$$\tilde{\rho}(t) = \frac{d\,ln\,x(t)}{dt} = \frac{d}{dt}(a + bt) = b,$$

використовують для опису неперервних процесів.

Якщо $b > 0$, то трендові моделі (3.47) та (3.48) описують прискорене зростання, а якщо $-1 < b < 0$ – прискорене спадання.

У межах прискореного розвитку можна виділити динаміку з постійним абсолютним прискоренням $\tilde{\varphi}$. Такий розвиток описують трендовою моделлю параболічного типу:

$$x(t) = a + bt + ct^2, a > 0, b > 0, c > 0. \tag{3.49}$$

Абсолютний приріст динаміки, яку описують трендовою моделлю (3.49),

$$\tilde{\delta}(t) = \frac{dx(t)}{dt} = b + 2ct.$$

Абсолютне прискорення динаміки

$$\tilde{\varphi}(t) = \frac{d\tilde{\delta}(t)}{dt} = \frac{d^2x(t)}{dt^2} = 2c = const,$$

темп приросту динаміки

$$\tilde{\rho}(t)=\frac{d\ln x(t)}{dt}=\frac{b+2ct}{a+bt+ct^2}.$$

Темп приросту динаміки, яку описують трендовою моделлю (3.49), може змінюватись двома способами: або монотонно спадати; або на початковому інтервалі часу зростати, а потім спадати.

Дослідимо зміну темпу приросту динаміки, яку описують параболічною моделлю, тобто знайдемо проміжки монотонності функції $\tilde{\rho}(t)=\dfrac{b+2ct}{a+bt+ct^2}$. Для цього розв'яжемо сукупність нерівностей:

$$\left[\begin{array}{l}\dfrac{d\tilde{\rho}(t)}{dt}>0;\\[2mm] \dfrac{d\tilde{\rho}(t)}{dt}<0.\end{array}\right.$$

$$\frac{d\tilde{\rho}(t)}{dt}=\frac{2c(a+bt+ct^2)-(b+2ct)^2}{(a+bt+ct^2)^2}.$$

Зрозуміло, що знак похідної $\dfrac{d\tilde{\rho}(t)}{dt}$ збігається зі знаком чисельника

$$-2c^2t^2-2bct+2ac-b^2.$$

Знайдемо проміжки знакосталості параболи $y(t)=-2c^2t^2-2bct+2ac-b^2$.

Для цього дослідимо рівняння:

$$2c^2t^2-2bct+(2ac-b^2)=0;$$

$$2c^2t^2+2bct-(2ac-b^2)=0;$$

$$D=(2bc)^2+4\cdot 2c^2\cdot(2ac-b^2)=4b^2c^2+16ac^3-8b^2c^2=16ac^3-4b^2c^2;$$

$$D=4c^2(4ac-b^2).$$

Якщо $4ac-b^2<0$, то парабола $y(t)=-2c^2t^2-2bct+2ac-b^2$ не перетинається з віссю Ox і має завжди від'ємні значення. Якщо $t>0$, $y(t)=-2c^2t^2-2bct+2ac-b^2<0$, якщо $2ac-b^2<0$, тобто $2ac<b^2$.

Якщо $4ac-b^2\ge 0$, то парабола $y(t)$ має два корені:

$$t_{1,2}=\frac{-2bc\pm 2c\sqrt{4ac-b^2}}{4c^2}=\frac{-b\pm\sqrt{4ac-b^2}}{2c},$$

серед яких (при додатних параметрах b,c і $2ac>b^2$) лише один додатний:

$$t_0 = \frac{-b+\sqrt{4ac-b^2}}{2c}.$$

У цій точці $\tilde{\rho}(t)$ досягає свого максимуму і відбувається перехід зростання на спадання.

Отже:

1. Якщо $2ac < b^2$, то $\tilde{\rho}(t)$ монотонно спадає на всьому інтервалі $[0, +\infty)$;

2. Якщо $2ac > b^2$, то на проміжку $\left[0, \frac{-b+\sqrt{4ac-b^2}}{2c}\right]$ $\tilde{\rho}(t)$ зростає, а на проміжку $\left[\frac{-b+\sqrt{4ac-b^2}}{2c}, \infty\right)$ спадає.

Трендова модель (3.48) відображає основні тенденції розвитку багатьох економічних процесів на сучасному етапі, коли абсолютні прирости продовжують збільшуватись, а темпи приросту спадають.

Моделлю тренда, яка описує динаміку уповільненого розвитку з насиченням може бути гіперболічна трендова модель першого порядку (узагальнена обернена модель) вигляду

$$x(t) = a + \frac{b}{t}. \tag{3.50}$$

Очевидно, що $\lim\limits_{t\to\infty} x(t) = a$ – межа насичення. Абсолютний приріст $\tilde{\delta}(t) = -\frac{b}{t^2}$ поступово спадає за абсолютною величиною.

Якщо $b < 0$, то узагальнена обернена модель описує динаміку уповільненого зростання з верхньою межею (рис. 3.9), а при $b > 0$ − динаміку уповільненого спадання з нижньою межею.

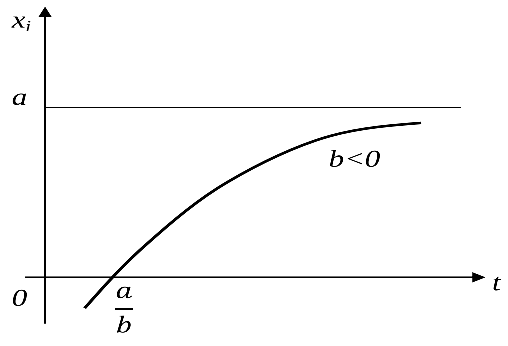


Рисунок 3.9 − Узагальнена обернена модель ($b < 0$)

Прикладами використання оберненої моделі в економічних дослідженнях є криві Енгеля та Філліпса. *Крива Енгеля:*

$$x(t) = a + \frac{b}{t}, \quad a > 0, \quad b < 0.$$

Наприкінці XIX ст. німецький статистик Е. Енгель сформулював емпіричні закони споживання і побудував криві, відповідно до яких із зростанням доходу частка витрат на харчування зменшується, а частка витрат на одяг і житло залишається стабільною. Криві, які пов'язують споживчі витрати на товари із загальними витратами або доходом, називають кривими Енгеля. Крива Енгеля (рис. 3.10) для деякого товару вказує на такі особливості:

- критичний рівень доходу – b/a, нижче якого товар не буде куплено;

- межу насичення (стелю) a, яку не можна збільшити, скільки б не зростав дохід.

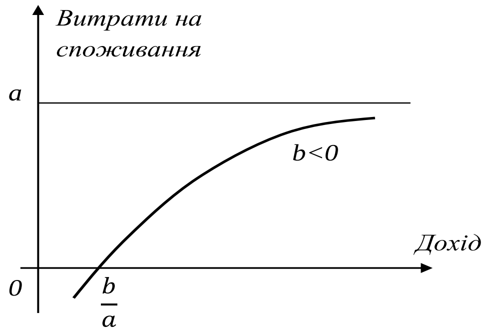


Рисунок 3.10 – Крива Енгеля

Крива Філліпса:

$$x(t) = a + \frac{b}{t}, \quad a < 0, \quad b > 0.$$

Економіст А. Філліпс, аналізуючи дані про норми зміни відсотка заробітної плати і відсотка безробіття для Англії за період від 1861 по 1957 рік, побудував криву, яка описує залежність норми зміни заробітної плати від норми безробіття і яку називають кривою Філліпса (рис. 3.11).

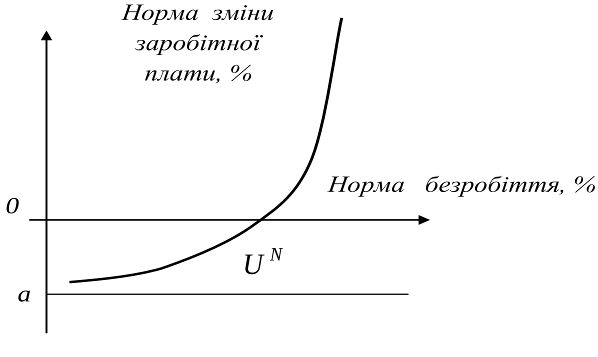


Рисунок 3.11 – Крива Філліпса

Крива Філліпса має такі особливості:

- межа насичення a є межею зміни заробітної плати;
- точка U^N є значенням природної норми безробіття.

Криву Філліпса в макроекономіці використовують, щоб розрахувати мінімальну заробітну плату, компенсацію за безробіття тощо.

Приклад 3.10. На підставі даних про валовий внутрішній продукт України за 2007 − 2011 рр. (табл. 3.8) Побудувати показникову трендову модель. Оцінити адекватність отриманої моделі.

Таблиця 3.8 – Дані для прикладу 3.10

Рік	*Квартал*	*Номер часового періоду*	*Валовий внутрішній продукт, млн. грн*
2007	I	1	52583
	II	2	60798
	III	3	75812
	IV	4	78151
2008	I	5	66981
	II	6	78607
	III	7	99405
	IV	8	100120
2009	I	9	88104
	II	10	101707
	III	11	122861
	IV	12	128780
2010	I	13	105423
	II	14	124116
	III	15	150434
	IV	16	157694
2011	I	17	133108

Розв'язання. Для знаходження параметрів показникової трендової моделі $x(t) = a_0 \cdot a_1^t$ складемо систему нормальних рівнянь (3.44), використавши допоміжну таблицю (табл. 3.9).

Маємо

$$\begin{cases} 17 ln\, a_0 + 153 ln\, a_1 = 195{,}1775 \\ 153 ln\, a_0 + 1785 ln\, a_1 = 1780{,}895 \end{cases}.$$

Розв'язавши побудовану систему нормальних рівнянь, знаходимо:
$ln\, a_1 \approx 0{,}059$, $a_1 \approx 1{,}0608$, $ln\, a_0 \approx 10{,}9459$, $a_0 \approx 56718{,}6105$.

Таблиця 3.9 – Таблиця для розрахунку системи нормальних рівнянь

i	t_i	x_i	t_i^2	$ln\,x_i$	$t_i\,ln\,x_i$
1	1	52583	1	10,8701	10,8701
2	2	60798	4	11,0153	22,0306
3	3	75812	9	11,236	33,708
4	4	78151	16	11,2664	45,0656
5	5	66981	25	11,1122	55,561
6	6	78607	36	11,2722	67,6332
7	7	99405	49	11,507	80,549
8	8	100120	64	11,5141	92,1128
9	9	88104	81	11,3863	102,4767
10	10	101707	100	11,5299	115,299
11	11	122861	121	11,7188	128,9068
12	12	128780	144	11,7659	141,1908
13	13	105423	169	11,5657	150,3541
14	14	124116	196	11,729	164,206
15	15	150434	225	11,9213	178,8195
16	16	157694	256	11,9684	191,4944
17	17	133108	289	11,7989	200,5813
$\sum$	**153**		**1785**	**195,1775**	**1780,859**

Таким чином, показникова трендова модель набуває вигляду

$$x(t) = 56718{,}6105 \cdot 1{,}0608^t.$$

Для оцінювання отриманої трендової моделі на адекватність використаємо коефіцієнт детермінації $D = \frac{Q_R}{Q} \cdot 100\%$. Для знаходження відповідних сум квадратів відхилень використаємо допоміжну таблицю (табл. 3.10).

Маємо

$$D = \frac{14146836905}{15434707332} \cdot 100\% \approx 92\%,$$

оскільки значення коефіцієнта детермінації близьке до 100%, то модель є адекватною.

Таблиця 3.10 – Допоміжна таблиця для розрахунку сум квадратів відхилень

i	t_i	x_i	$x(t_i)$	$Q_R = (x(t_i) - \bar{x})^2$	$Q = (x_i - \bar{x})^2$
1	1	52583	60167,1	1704442801	2388179161
2	2	60798	63825,26	1415771426	1652747716
3	3	75812	67705,84	1138803467	657409600
4	4	78151	71822,35	877916000,6	542936601
5	5	66981	76189,15	638211503,4	1188249841
6	6	78607	80821,45	425619504,9	521894025
7	7	99405	85735,4	247011627,7	4190209
8	8	100120	90948,11	110331735,9	1774224
9	9	88104	96477,75	24743127,91	178169104
10	10	101707	102343,6	794952,2512	65025
11	11	122861	108566,1	50610303,36	458345281
12	12	128780	115166,9	188098763,8	746819584
13	13	105423	122169,1	429196509,5	15768841
14	14	124116	129596,9	792137488,3	513656896
15	15	150434	137476,4	1297759625	2399236324
16	16	157694	145835	1969850507	3163162564
17	17	133108	154701,8	2835537561	1002102336
$\sum$	**153**		**1709548**	**14146836905**	**15434707332**

3.8 Моделювання сезонних коливань економічних явищ

При порівнянні квартальних та місячних даних багатьох соціально-економічних явищ можна виявити періодичні коливання, що виникають під впливом зміни пір року. Вони є результатом впливу природно-кліматичних умов, загальних економічних факторів. Прикладом таких явищ може бути попит на деякі види продукції, пасажирські перевезення, виробництво у сезонних галузях економіки (цукровій, консервній).

У широкому розумінні до сезонних відносять явища, які мають у своєму розвитку чітко виражену закономірність змін протягом року. Періодичні коливання, що мають визначений та сталий період, називають *“сезонними коливаннями”* або *“сезонними хвилями”*, а динамічний ряд в цьому випадку називають *сезонним рядом динаміки*.

Повсякденна життєдіяльність людей за умов періодичної зміни сезонів супроводжується специфічними змінами інтенсивності динаміки соціально-економічних процесів. У більшості галузей народного господарства це виявляється у вигляді річних чергувань підйомів і спадів випуску продукції, різного використання сировини та енергії, коливання рівнів продуктивності праці, собівартості, прибутку й інших показників.

Для деяких сфер людської діяльності річна динаміка характеризується призупиненням процесів у міжсезонні періоди (цукроваріння, рибальство, мисливство, навігація, туризм і т. п.). Яскраво виражений сезонний характер має сільськогосподарське виробництво, особливо рослинництво за умов відкритого ґрунту. Це викликає нерівномірність використання трудових ресурсів, напруженість у роботі транспорту, сховищ, баз. Сезонне коливання характерне також для грошового та товарного оборотів.

У практичній діяльності люди, впливаючи на природу, створюють більш сприятливі умови праці та побуту. Однак, на даному етапі свого розвитку людство не може керувати всіма силами природи. Наприклад, неможливо на власний розсуд змінювати час настання та тривалість несприятливих сезонів, наявність стихійних лих.

Знання сезонних особливостей попиту на окремі товари має важливе значення для торгівлі: розробка заходів підвищення ефективності торгівлі, покращення організації торгівлі, підвищення культури обслуговування покупців. З’ясування особливостей попиту населення на товари за сезонами важливе для розробки науково обґрунтованих нормативів, воно дозволяє уникнути нераціональних витрат та втрат.

Статистичні ряди внутрішньорічної динаміки зазвичай складають за матеріалами поточної звітності. Однією з необхідних умов вивчення сезонних коливань є зведення рядів динаміки до зіставного вигляду. При цьому необхідно мати на увазі, що різновеликі за тривалістю місяці та квартали річних періодів є однією з причин, які впливають на зміни рівнів рядів внутрішньорічної динаміки. Для усунення цієї причини об’ємні величини перераховують в середні, які характеризують інтенсивність розвитку досліджуваного явища за одиницю часу.

Для аналітичного вирівнювання, як методу виміру сезонних хвиль, використовують вирівнювання за рядом Фур’є, який подає сезонне явище у вигляді гармоніки.

Ряд Фур’є у загальному випадку можна записати у вигляді такої залежності:

$$\overline{y_t} = a_0 + \sum_{k=1}^{m}(a_k \cos kt + b_k \sin kt), \qquad (3.51)$$

де a_0, a_k, b_k – параметри досліджуваної моделі;

t – фактор часу;

k – порядковий номер гармоніки;

m – номер гармоніки, яка використовується з різним ступенем точності (зазвичай, від одного до чотирьох).

Звичайно кількість гармонік при побудові ряду Фур’є необхідно приймати не більше чотирьох. Після побудови усіх гармонік проводять оцінювання їх адекватності, тобто визначають, яка з гармонік найкраще описує сезонні коливання економічних явищ.

Як і в будь-якій моделі, необхідно оцінити її параметри. Оцінювання параметрів ряду Фур’є здійснюється на основі методу найменших квадратів за такими залежностями:

$$a_0 = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = \overline{y}, \tag{3.52}$$

$$a_k = \frac{2}{n} \sum_{i=1}^{n} y_i \cos kt_i , \tag{3.53}$$

$$b_k = \frac{2}{n} \sum_{i=1}^{n} y_i \sin kt_i , \tag{3.54}$$

де n − це кількість часових періодів, за які розглядається досліджуване явище (дні, місяці, квартали, роки), t_i – періоди, y_i – рівні динаміки.

Необхідно зауважити, що для побудови економетричної моделі потрібно для такого фактора, як час, зробити перерахунок від натурального масштабу до радіанного або ж градусного. Такий перехід пропонується здійснити за формулою:

$$t = \frac{2\pi}{n} \cdot t_i , \tag{3.55}$$

де t_i – фактор часу у радіанному або ж градусному вимірі;
n – кількість спостережень (кількість інтервалів часу).

Можна подати місяці року у радіанній формі (таблиця 3.11).

Таблиця 3.11 – Радіанна форма подання місяців року

№ місяця	Місяць	Радіани	№ місяця	Місяць	Радіани
1	Січень	0	7	Липень	π
2	Лютий	π/6	8	Серпень	7π/6
3	Березень	π/3	9	Вересень	4π/3
4	Квітень	π/2	10	Жовтень	3π/2
5	Травень	2π/3	11	Листопад	5π/3
6	Червень	5π/6	12	Грудень	11π/6

Перед тим, як визначати параметри економетричної моделі, необхідно скласти таблицю значень тригонометричних функцій (табл. 3.12).

Таблиця 3.12 – Таблиця значень тригонометричних функцій

№	t_i, рад	$\cos t_i$	$\cos 2t_i$	$\sin t_i$	$\sin 2t_i$
1	0	1	1	0	0
2	π/6	0,866	0,5	0,5	0,866
3	π/3	0,5	-0,5	0,866	0,866
4	π/2	0	-1	1	0
5	2π/3	-0,5	-0,5	0,866	-0,866
6	5π/6	-0,866	0,5	0,5	-0,866
7	π	-1	1	0	0
8	7π/6	-0,866	0,5	-0,5	0,866
9	4π/3	-0,5	-0,5	-0,866	0,866
10	3π/2	0	-1	-1	0
11	5π/3	0,5	-0,5	-0,866	-0,866
12	11π/6	0,866	0,5	-0,5	-0,866

Розглянемо застосування ряду Фур’є на конкретному прикладі.

Приклад 3.11. Дано статистичні щомісячні дані щодо реалізації зимового одягу (табл. 3.13). Необхідно описати сезонні коливання реалізації зимового одягу на основі рядів Фур’є (з використанням двох гармонік) і обрати гармоніку, яка найбільш адекватно описує ці сезонні коливання.

Таблиця 3.13 – Дані прикладу 3.11

№	Місяці	t_i, рад	y_i
1	Січень	0	37
2	Лютий	π/6	40
3	Березень	π/3	44
4	Квітень	π/2	52
5	Травень	2π/3	46
6	Червень	5π/6	70
7	Липень	π	60
8	Серпень	7π/6	48
9	Вересень	4π/3	46
10	Жовтень	3π/2	38
11	Листопад	5π/3	36
12	Грудень	11π/6	35
Сума			552

Розв’язання. Для подальших розрахунків необхідно скласти допоміжну таблицю (табл. 3.14) для першої гармоніки, що значно полегшить оцінювання параметрів моделі.

Таблиця 3.14 – Допоміжні розрахунки для першої гармоніки

№	$y_i\,cos\,t_i$	$y_i\,sin\,t_i$	$\overline{y_{1,t_i}}$	$\left(\overline{y_{1,t_i}}-\overline{y}\right)^2$	$\left(y_i-\overline{y}\right)^2$
1	37,000	0,000	34,96	121,8816	81
2	34,641	20,000	39,31	44,7561	36
3	22,000	38,105	45,45	0,3025	4
4	0,000	52,000	51,74	32,9476	36
5	-23,000	39,837	56,49	110,0401	0
6	-60,622	35,000	58,43	154,5049	576
7	-60,000	0,000	57,04	121,8816	196
8	-41,569	-24,000	52,69	44,7561	4
9	-23,000	-39,837	46,55	0,3025	0
10	0,000	-38,000	40,26	32,9476	64
11	18,000	-31,177	35,51	110,0401	100
12	30,311	-17,500	33,57	154,5049	121
Σ	**-66,239**	**34,428**	**552,0**	**928,8656**	**1218**

За формулами (3.51) – (3.53) розрахуємо параметри моделі для першої гармоніки:

$$a_{01}=\overline{y}=\frac{552}{12}=46;\ a_1\approx\frac{2\cdot(-66{,}24)}{12}\approx-11{,}04;\ b_1\approx\frac{2\cdot 34{,}43}{12}\approx 5{,}74.$$

Отримавши значення параметрів, можна записати ряд Фур’є для першої гармоніки:

$$\overline{y_{1t}}=46-11{,}04\cdot cos\,t+5{,}74\cdot sin\,t. \qquad (3.56)$$

Підставивши значення t_i в рівняння (3.56) одержуємо теоретичні значення змінної $\overline{y_{1,t_i}}$ (див. табл. 3.14). Замість значень косинусів та синусів у рівняння підставляємо відповідні їм значення з таблиці 3.14.

Розрахуємо абсолютне значення коефіцієнта кореляції для першої гармоніки за формулою (3.23), використовуючи попередньо підраховані проміжні дані з таблиці 3.14, таким чином

$$r_{y_1/t}=\sqrt{\frac{928{,}8656}{1218}}\approx\sqrt{0{,}76}\approx 0{,}87.$$

Оскільки значення коефіцієнта кореляції близьке до одиниці, то це свідчить про щільний зв’язок між коливаннями реалізації зимового одягу та часом.

Надалі необхідно за тим же принципом, що і для першої гармоніки, оцінити параметри для другої гармоніки (табл. 3.15).

Таблиця 3.15 – Допоміжні розрахунки для другої гармоніки

$y_i\,cos\,2t_i$	$y_i\,sin\,2t_i$	$\overline{y_{2,t_i}}$	$\left(\overline{y_{2,t_i}} - \overline{y}\right)^2$
37	0,000	37,88	65,9344
20	34,641	39,64	40,4496
-22	38,105	42,87	9,7969
-52	0,000	48,82	7,9524
-23	-39,837	56,16	103,2256
35	-60,622	61,01	225,3001
60	0,000	59,96	194,8816
24	41,569	53,03	49,4209
-23	39,837	43,97	4,1209
-38	0,000	37,35	74,8225
-18	-31,177	35,18	117,0724
17,5	-30,311	36,15	97,0225
17,5	**-7,794**	**552,0**	**989,9998**

Таким чином, маємо значення параметрів для другої гармоніки:

$$a_{02} = \overline{y} = \frac{552}{12} = 46\,;\ a_2 \approx \frac{2 \cdot 17{,}5}{12} \approx 2{,}92\,;\ b_1 \approx \frac{2 \cdot (-7{,}79)}{12} \approx -1{,}3\,.$$

Запишемо ряд Фур'є для другої гармоніки:

$$\overline{y_{2t}} = 46 - 11{,}04 \cdot cos\,t + 5{,}74 \cdot sin\,t + 2{,}92 \cdot cos\,2t - 1{,}3 \cdot sin\,2t\,. \qquad (3.57)$$

Розраховуємо абсолютне значення коефіцієнта кореляції для першої гармоніки за формулою (3.23), використовуючи попередньо підраховані проміжні дані з таблиці 3.14,

$$r_{y_2/t} = \sqrt{\frac{989{,}9998}{1218}} \approx \sqrt{0{,}813} \approx 0{,}9\,.$$

Оскільки значення коефіцієнта кореляції близьке до одиниці, то це свідчить про щільний зв'язок між коливаннями реалізації зимового одягу та часом.

Порівнюючи отримані значення щільності зв'язку для двох гармонік можна зробити висновок, що ряд Фур'є для другої гармоніки є більш адекватним та краще описує сезонні коливання. На основі обраного рівняння будуємо графік залежності продажів від місяців, використовуючи прикладний пакет *Mathcad* (рис. 3.12).

Приклад 3.12. Дано статистичні дані про середньоденний товарообіг торговельного підприємства по місяцях 2010 року (табл. 1.16). Опишіть сезонні коливання товарообігу з використанням однієї гармоніки та визначте розрахункові рівні товарообігу підприємства щомісячно.

Таблиця 3.16 – Дані прикладу 3.12

Місяць	Обсяг товарообігу, тис. грн
Січень	65,1
Лютий	66,5
Березень	74,4
Квітень	73,6
Травень	67,2
Червень	100,0
Липень	90,0
Серпень	72,6
Вересень	68,9
Жовтень	70,4
Листопад	66,3
Грудень	77,2

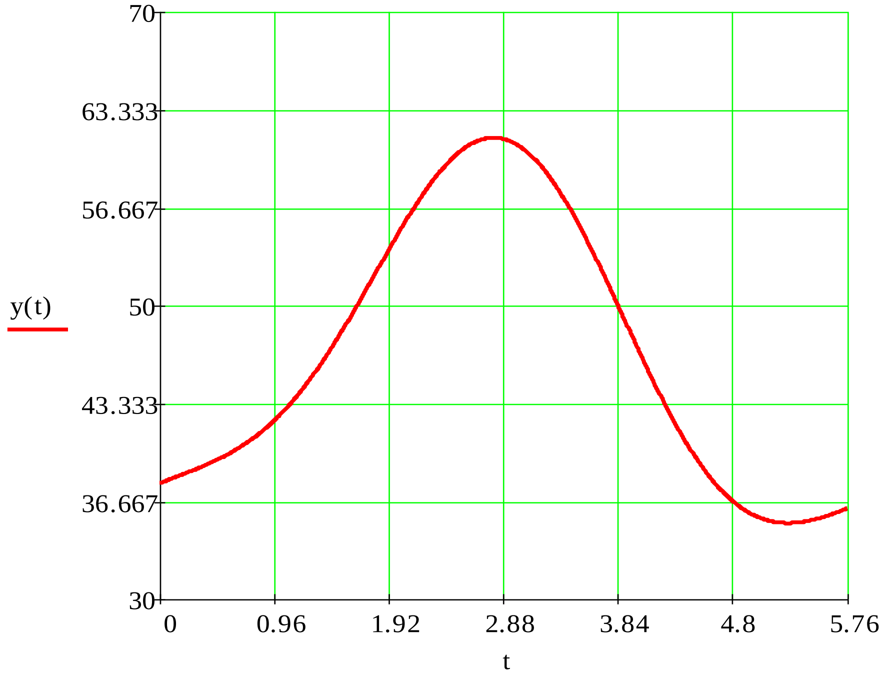


Рисунок 3.12 – Щомісячний графік продажів

Розв’язання. Для подальших розрахунків необхідно скласти допоміжну таблицю (табл. 3.17), використавши таблицю 3.13, що значно полегшує оцінювання параметрів моделі.

Таблиця 3.17 – Допоміжна таблиця для прикладу 3.12

t_i, рад.	Обсяг товарообігу, тис. грн (y_i)	$cos\,t_i$	$sin\,t_i$	$y_i\,cos\,t_i$	$y_i\,sin\,t_i$
0	65,1	1	0	65,1	0
π/6	66,5	0,866	0,5	57,6	33,3
π/3	74,4	0,5	0,866	37,2	64,4
π/2	73,6	0	1	0	73,6
2π/3	67,2	-0,5	0,866	-33,6	58,2
5π/6	100,0	-0,866	0,5	-86,6	50,0
π	90,0	-1	0	-90,0	0
7π/6	72,6	-0,866	-0,5	-62,9	-36,3
4π/3	68,9	-0,5	-0,866	-34,5	-59,7
3π/2	70,4	0	-1	0	-70,4
5π/3	66,3	0,5	-0,866	33,2	-57,4
11π/6	77,2	0,866	-0,5	66,9	-38,6
Σ	**893,0**	*	*	**-47,6**	**17,1**

Застосовуючи першу гармоніку ряду Фур’є, визначимо параметри рівняння (3.50):

- за формулою (3.51): $a_0 = \frac{1}{12} \cdot 893 \approx 74{,}4$;

- за формулою (3.52): $a_1 = \frac{2}{12}(-47{,}6) \approx -7{,}9$;

- за формулою (3.53): $b_1 = \frac{2}{12} \cdot 17{,}1 \approx 2{,}9$.

За одержаними параметрами математична модель така:

$$\overline{y_t} = 74{,}4 - 7{,}9 \cdot cos\,t + 2{,}9 \cdot sin\,t\,. \qquad (3.58)$$

За даними моделі (3.58) визначаємо для кожного місяця розрахункові теоретичні рівні $\overline{y_{t_i}}$, одержані значення занесемо до таблиці 3.18.

$$\overline{y_{t\,\text{січень}}} = 74{,}4 - 7{,}9 \cdot 1 + 2{,}9 \cdot 0 = 66{,}5 \text{ тис. грн;}$$

$$\overline{y_{t\,\text{лютий}}} = 74{,}4 - 7{,}9 \cdot 0{,}866 + 2{,}9 \cdot 0{,}5 = 69{,}0 \text{ тис. грн;}$$

………………………………………………..

$$\overline{y_{t\,\text{грудень}}} = 74{,}4 - 7{,}9 \cdot 0{,}866 + 2{,}9 \cdot (-0{,}5) = 66{,}1 \text{ тис. грн.}$$

Оскільки $\sum_{i=1}^{12} \overline{y_{t_i}} = 892{,}8$, то це свідчить про достатньо точний розподіл вирівняних даних. Відхилення у 0,2 пояснюється заокругленням у розрахунках.

Таблиця 3.18 – Щомісячні розрахункові теоретичні рівні $\overline{y_{t_i}}$

Місяць	$\overline{y_{t_i}}$	Місяць	$\overline{y_{t_i}}$
Січень	66,5	Липень	82,3
Лютий	69,0	Серпень	79,8
Березень	73,0	Вересень	75,8
Квітень	77,3	Жовтень	71,5
Травень	80,9	Листопад	67,9
Червень	82,7	Грудень	66,1

3.9 Алгоритм розв'язання практичної задачі однофакторного кореляційного аналізу

Нехай для виробничого підприємства відомі такі показники: кількість виробленого та реалізованого продукту ***К*** *(тис. одиниць),* ціна ***Ц*** *(тис. грошових од. за одиницю продукції),* витрати ***В*** виробництва за повною собівартістю *(млн. грошових од.).* Необхідно провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку та розробити стратегію підприємства на майбутні періоди.

Відомо, що прибуток підприємства обчислюється за формулою:

$$П = К \cdot Ц - В.$$

Зрозуміло, що в даному випадку ціна та витрати виробництва за повною собівартістю, а отже і прибуток, кореляційно залежать від кількості виробленого та реалізованого продукту. Тому в наших розрахунках ми будемо одержувати функцію прибутку, аналітичний вираз якої в спрощеному вигляді можна записати так:

$$\overline{П}_К = К \cdot \overline{Ц}_К - \overline{В}_К. \qquad (3.59)$$

При розв'язуванні даної практичної задачі потрібно дотримуватися певного ***алгоритму***.

1. *Знаходимо аналітичні рівняння регресії для ціни та витрат:* $\overline{Ц_К}$ *=Ц(К),* $\overline{В_К}$ *=В(К).*

Для виконання цього пункту потрібно спочатку нанести на координатні площини *КОЦ* та *КОВ* емпіричні дискретні точки та зробити припущення щодо форми відповідної лінії регресії. За допомогою нормальних систем (3.10) або (3.43) – (3.45) оцінюємо значення параметрів обраних рівнянь регресії.

2. Перевіряємо адекватність знайдених рівнянь регресії. Оцінюємо щільність зв'язку між кількістю виробленого та реалізованого продукту і ціною та кількістю і витратами.

Адекватність знайдених рівнянь регресії перевіряємо за допомогою коефіцієнта детермінації, який обчислюємо за формулою (3.26) або за (3.27). Варто нагадати, що рівняння будуть значущими, якщо коефіцієнт детермінації D належить відрізку [0,55; 1]. Якщо $D \in [0{,}45;\ 0{,}55]$, то для встановлення адекватності моделі реальному об'єкту використовують F-критерій Фішера.

За умови, що $0 < D < 0{,}45$, знайдене рівняння регресії неадекватне, наше припущення щодо форми лінії регресії недостовірне і потрібно робити інше припущення.

Оцінити щільність зв'язку між відповідними показниками та зробити необхідні висновки можна за допомогою коефіцієнта кореляції ($r_{Ц/К}$ та $r_{В/К}$), який обчислюється за однією з формул (3.23) − (3.27).

Зауваження! Неадекватність одержаних рівнянь регресії може бути пов'язана з використанням нераціональних методів розв'язування систем нормальних рівнянь або неточних округлень. Для усунення останнього недоліку пропонується в розрахунках використовувати числа не менше як із п'ятьма знаками після коми.

3. Будуємо функцію прибутку та знаходимо кількість виробленого і реалізованого продукту, при якій прибуток буде максимальним.

Для побудови функції прибутку підставляємо знайдені рівняння регресії для ціни та витрат у рівність (3.59). Одержану функцію прибутку досліджуємо на екстремум (знаходимо точку максимуму K_0).

4. Робимо висновки за даним критерієм про стан підприємства та розробляємо стратегію на майбутні періоди.

Для одержання висновків щодо стану підприємства складається допоміжна порівняльна таблиця (табл. 3.19). Значення $К_п$, $Ц_п$ та $В_п$ – статистичні дані динамічного ряду за останній період. Оптимальне значення кількості виробленого та реалізованого продукту – це точка максимуму функції прибутку $К_{0.}$ Підставивши це значення в рівняння регресії $\overline{Ц}_к = Ц(К)$, $\overline{B}_К = B(К)$, отримаємо оптимальні ціну та витрати.

Аналізуємо значення відхилення фактичного значення показника від оптимального, робимо відповідні висновки та розробляємо стратегію підприємства на майбутні періоди

Таблиця 3.19 – Допоміжна порівняльна таблиця

Показник	K	$Ц$	B	$ЦK$	$П$
Фактичний за останній період	K_n	$Ц_n$	B_n	$Ц_n K_n$	$П_n$
Оптимальний	K_0	$\overline{Ц_к}=Ц(K_0)$	$\overline{B_к}=B(K_0)$	$Ц(K_0)\,K_0$	$Ц(K_0)\,K_0-B(K_0)$
Відхилення фактичного від оптимального	K_n-K_0	$Ц_n-Ц(K_0)$	$B_n-B(K_0)$	$Ц_n K_n-$ $-Ц(K_0)\,K_0$	$П_n-$ $-(Ц(K_0)\,K_0-B(K_0))$

Зауваження! Якщо в п. 1 ми встановили, що крива прибутку має точку мінімуму, а не максимуму, то ми не зможемо розв’язати задачу за критерієм максимального прибутку. Ми можемо лише визначити мінімальний обсяг виробництва та мінімальні ціну й витрати.

Приклад 3.13. Для виробничого підприємства відомі такі показники за 6 періодів: кількість виробленого та реалізованого продукту *К* (тис. одиниць), ціна *Ц* (тис. грошових од. за одиницю продукції), витрати *В* виробництва за повною собівартістю (млн. грошових од.) (табл. 3.20).

Необхідно:

1) знайти кореляційну залежність ціни *Ц* та витрат *В* від кількості виробленого і реалізованого продукту *К*: $\overline{Ц}=Ц(K)$, $\overline{B}=B(K)$. Оцінити щільність зв’язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації;

2) провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку;

3) зробити висновки та розробити стратегію підприємства на майбутні періоди.

Таблиця 3.20 – Дані для прикладу 3.13

Період / Показник	1	2	3	4	5	6
К	48	37	25	23	22	21
Ц	1,653	4,794	7,288	8,165	9,218	10,000
В	55	120	137	153	173	182

Розв’язання. Якщо побудувати емпіричні точки, то можна переконатися в тому, що залежність $\overline{Ц}=Ц(K)$ доцільно шукати у вигляді лінійної функції

$$\overline{Ц}=a_1K+a_0. \tag{3.60}$$

Для пошуку параметрів функції (3.60) використаємо систему нормальних рівнянь (3.10):

$$\begin{cases} a_1 \sum K_i + n a_0 = \sum Ц_i; \\ a_1 \sum K_i^2 + a_0 \sum K_i = \sum K_i Ц_i. \end{cases}$$

Складаємо таку допоміжну розрахункову таблицю (табл. 3.21):

Таблиця 3.21 – Допоміжна розрахункова таблиця для розрахунку $\hat{Ц}$

Період	K_i	$Ц_i$	K_i^2	$K_i\,Ц_i$	$\overline{Ц}_i$	$(\overline{Ц}_i - \overline{Ц})^2$	$(Ц_i - \overline{Ц})^2$
1	48	1,653	2304	79,344	1,575786	27,851099	27,040000
2	37	4,794	1369	177,378	4,685484	4,698126	4,239481
3	25	7,288	625	182,200	8,078100	1,500870	0,189225
4	23	8,165	529	187,795	8,643536	3,206019	1,721344
5	22	9,218	484	202,796	8,926254	4,298382	5,593225
6	21	10,000	441	210,000	9,208972	5,550604	9,903609
Σ	**176**	**41,118**	**5752**	**1039,513**	–	**47,105100**	**48,686884**

Тобто, в нашому випадку система нормальних рівнянь має вигляд:

$$\begin{cases} 176a_1 + 6a_0 = 41{,}118; \\ 5752a_1 + 176a_0 = 1039{,}513. \end{cases}$$

Розв’яжемо отриману систему методом Гаусса (додаток Ж). Маємо:

$$\left(\begin{array}{cc|c} 176 & 6 & 41{,}118 \\ 5752 & 176 & 1039{,}513 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 3{,}492857\cdot 10^{-2} & 2{,}33625\cdot 10^{-1} \\ 1 & 3{,}0598052\cdot 10^{-2} & 1{,}80722\cdot 10^{-1} \end{array}\right) \sim$$

$$\sim \left(\begin{array}{cc|c} 0 & 3{,}492857\cdot 10^{-3} & 5{,}2903\cdot 10^{-2} \\ 1 & 3{,}0598052\cdot 10^{-2} & 1{,}8072\cdot 10^{-1} \end{array}\right).$$

Із першого рядка останньої матриці випливає, що $a_0 = \dfrac{5{,}2903\cdot 10^{-2}}{3{,}492857\cdot 10^{-3}} \approx 15{,}14605$. Використовуючи другий рядок останньої матриці знаходимо наступний параметр рівняння регресії:

$$a_1 = 1{,}8072\cdot 10^{-1} - 3{,}0598052\cdot 10^{-2} a_0 \approx -0{,}282718 .$$

Тому шукане рівняння регресії має вигляд:

$$\overline{Ц}_i = -0{,}282718K + 15{,}14605. \qquad (3.61)$$

Підставляючи по черзі значення K_i у рівняння (3.61), отримаємо відповідні значення $\overline{Ц}_i$, якими заповнюємо шосту графу таблиці 3.21, а за цими значеннями та значеннями $Ц_i$ і $\overline{Ц}_i$, обчислюємо елементи сьомого та восьмого стовпців цієї таблиці.

Щільність зв'язку між ознаками $Ц$ та $К$ обчислюємо за допомогою коефіцієнта кореляції:

$$r_{Ц/К} = -\sqrt{\frac{\sum(\overline{Ц}_i - \overline{Ц})^2}{\sum(Ц_i - \overline{Ц})^2}} = -\sqrt{\frac{47{,}1051}{48{,}686884}} = -\sqrt{0{,}967511} \approx -0{,}984.$$

(Знак "–" для попередніх розрахунків обираємо з тієї причини, що пряма (3.61) є спадною).

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про щільний зворотний зв'язок між ціною товару та його реалізованою кількістю. Коефіцієнт детермінації D є підкореневим виразом останньої формули. Цей коефіцієнт $D \approx 0{,}97$ показує, що варіація результативної ознаки (ціни товару) на 97% відбувається під впливом фактора кількості реалізованого продукту, а на 3% − під впливом не врахованих в моделі факторів.

Аналогічно, якщо побудувати емпіричні точки, то можна переконатися в тому, що залежність $\overline{B}_i = B(К)$ доцільно шукати у вигляді лінійної функції

$$\overline{B}_i = b_1 К + b_0. \quad (13.62)$$

Для пошуку параметрів функції (3.62) використаємо систему нормальних рівнянь (3.10):

$$\begin{cases} b_1 \sum К_i + n b_0 = \sum B_i; \\ b_1 \sum К_i^2 + b_0 \sum К_i = \sum К_i B_i. \end{cases}$$

Складаємо таку допоміжну розрахункову таблицю (табл. 3.22):

Таблиця 3.22 – Допоміжна розрахункова таблиця для розрахунку $\overline{B}_i$

Період	$К_i$	B_i	$К_i^2$	$К_i\,B_i$	$\overline{B}_i$	$(\overline{B}_i - \overline{B})^2$	$(B_з - \overline{B})^2$
1	48	55	2304	2640	60,60632	5784,16224	6668,35560
2	37	120	1369	4440	105,42758	975,46405	277,55560
3	25	137	625	3425	154,3235	311,99923	0,1156
4	23	153	529	3519	162,47282	666,30167	266,99560
5	22	173	484	3806	166,54748	893,26146	1320,5056
6	21	182	441	3822	170,62214	1153,42695	2055,7156
Σ	**176**	**820**	**5752**	**21652**	**820**	**9784,6156**	**10589,3336**

Тобто, в нашому випадку система нормальних рівнянь має вигляд:

$$\begin{cases} 176b_1 + 6b_0 = 820; \\ 5752b_1 + 176a_0 = 21652. \end{cases}$$

Розв'яжемо отриману систему методом Гаусса. Маємо:

$$\left(\begin{array}{cc|c} 176 & 6 & 820 \\ 5752 & 176 & 21652 \end{array}\right) \sim \left(\begin{array}{cc|c} 176 & 6 & 820 \\ 0 & -3536 & -905888 \end{array}\right).$$

Із другого рядка останньої матриці випливає, що $b_0 = \frac{905888}{3536} \approx 256{,}1900$. Використовуючи перший рядок останньої матриці, знаходимо такий параметр рівняння регресії:

$$b_1 = \frac{1}{176}(820 - 6b_0) \approx -4{,}07466.$$

Тому шукане рівняння регресії має вигляд:

$$\overline{B}_i = -4{,}07466\ K + 256{,}1900\ . \tag{3.63}$$

Підставляючи по черзі значення K_i в рівняння (3.63), отримаємо відповідні значення $\overline{B}_i$, якими заповнюємо шосту графу таблиці 3.22, а за цими значеннями та значеннями B_i і $\overline{B}$, обчислюємо елементи сьомої та восьмої граф цієї таблиці.

Оскільки пряма (3.63) є спадною (коефіцієнт біля K від'ємний), то щільність зв'язку між ознаками B та K обчислюємо за допомогою коефіцієнта кореляції так:

$$r_{B/K} = -\sqrt{\frac{\sum(\overline{B}_i - \overline{B})^2}{\sum(B_i - \overline{B})^2}} = -\sqrt{\frac{9784{,}6156}{10589{,}3336}} = -\sqrt{0{,}92400621} \approx -0{,}961.$$

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про щільний зворотний зв'язок між витратами на виробництво та кількістю реалізованої продукції. Коефіцієнт детермінації D є підкореневим виразом останньої формули. Цей коефіцієнт $D \approx 0{,}92$ показує, що варіація результативної ознаки (витрат) на 92% відбувається під впливом фактора кількості реалізованого продукту, а на 8% – під впливом не врахованих в моделі факторів.

Використовуючи рівняння (3.61) та (3.63), знаходимо формулу прибутку від реалізації продукту:

$$\overline{П} = \overline{Ц} \cdot K - \overline{B} = (-0{,}282718K + 15{,}14605)K + 4{,}07466K - 256{,}1900 =$$
$$= -0{,}282718K^2 + 19{,}22071K - 256{,}1900.$$

Прирівнюючи до нуля першу похідну отриманої функції, знаходимо оптимальне значення кількості реалізованого продукту за критерієм максимального прибутку:

$$(\overline{П})' = -0{,}565436 \cdot K + 19{,}22071 = 0\,.$$

Звідки $K \approx 34$ (оскільки $(\overline{П})'(30)>0$, $(\overline{П})'(40)<0$, то $K \approx 34$ є точкою максимуму).

Таким чином, оптимальний обсяг випуску та реалізації продукту складає 34 одиниці. За цією величиною можна обчислити оптимальну ціну продукту, за рівнянням (3.61) оптимальні витрати, за рівнянням (3.63) виручку від реалізації та оптимальний прибуток. Результати розрахунків наведено в таблиці 3.23. Для порівняння в даній таблиці наведено також фактичні дані за останній період.

Таблиця 3.23 – Порівняльна таблиця

Показник	K	$Ц$	B	$ЦК$	$П$
Фактичний за останній період	21	10,000	182	210	28
Оптимальний	34	5,534	117,652	188,156	70,504
Відхилення фактичного від оптимального	-13	4,466	64,348	21,844	-42,504

Висновки:

1. Прибуток підприємства на 42,504 млн. грош. од. менший оптимального за рахунок відповідного перевищення витрат на 64,348 млн. грош. од.;

2. Виручка від реалізації продукту перевищує оптимальну на 21,844 млн. грош. од. за рахунок збільшення ціни на 4,466 тис. грош. од.

Стратегія підприємства: зростання обсягу випуску та реалізації продукту, зменшення ціни та витрат.

Приклад 3.14. Для виробничого підприємства відомі такі показники за 6 періодів: кількість виробленого та реалізованого продукту K (тис. одиниць), ціна $Ц$ (тис. грошових од. за одиницю продукції), витрати B виробництва за повною собівартістю (млн. грошових од.) (табл. 3.24).

Необхідно:

1) знайти кореляційну залежність ціни $Ц$ та витрат B від кількості виробленого та реалізованого продукту К: $\overline{Ц} = Ц(K)$, $\overline{B} = B(K)$.

Оцінити щільність зв'язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації;

2) провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку;

3) зробити висновки та розробити стратегію підприємства на майбутні періоди.

Таблиця 3.24 – Дані для прикладу 3.14

Період	1	2	3	4	5	6
К	68	31	26	25	14	12
Ц	6	17	19	19	24	25
В	401	420	520	484	280	255

Розв'язання. Якщо побудувати емпіричні точки, то можна переконатися в тому, що залежність $\overline{Ц_к} = Ц(К)$ є лінійною функцією

$$\overline{Ц_К} = a_1 К + a_0 . \tag{3.64}$$

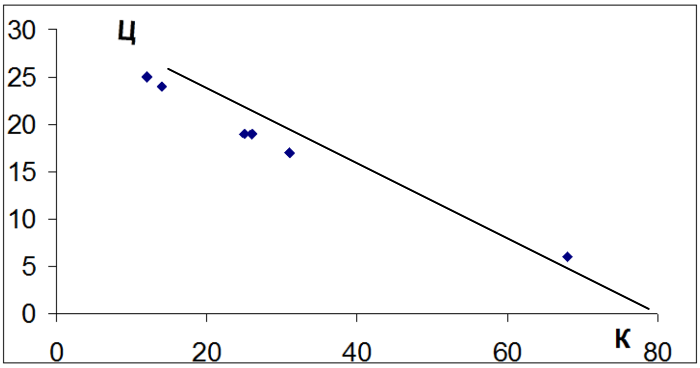


Рисунок 3.13 − Апроксимація даних *Ц*(*К*) лінійною залежністю

Для знаходження параметрів функції (3.64) використаємо систему нормальних рівнянь (3.10):

$$\begin{cases} a_1 \sum K_i + n a_0 = \sum Ц_i \\ a_1 \sum K_i^2 + a_0 \sum K_i = \sum K_i Ц_i \end{cases}.$$

Складаємо таку допоміжну розрахункову таблицю (табл. 3.25):

Тобто, в нашому випадку система нормальних рівнянь має вигляд:

$$\begin{cases} 176 a_1 + 6 a_0 = 110; \\ 7226 a_1 + 176 a_0 = 2540. \end{cases}$$

Таблиця 3.25 Допоміжна розрахункова таблиця

Період	K_i	$Ц_i$	K_i^2	$K_i Ц_i$	$\overline{Ц_K}$	$(\overline{Ц_K}-\overline{Ц})^2$	$(Ц_i-\overline{Ц})^2$
1	68	6	4626	408	5,465	165,587	152,111
2	31	17	961	527	17,779	0,308	1,778
3	26	19	676	494	19,443	1,231	0,444
4	25	19	625	475	19,775	2,08	0,444
5	14	24	196	336	23,436	26,039	32,111
6	12	25	144	300	24,102	33,275	44,444
$\sum$	**176**	**110**	**7226**	**2540**	**110**	**228,519**	**231,333**

Розв’язавши отриману систему методом Гаусса, знаходимо:

$$a_0 \approx 28{,}095315\ ,\ a_1 \approx -0{,}333\,.$$

Тому шукане рівняння регресії має вигляд:

$$\overline{Ц_K} = -0{,}333K + 28{,}095315\,. \qquad (3.65)$$

Підставляючи по черзі значення K_i в рівняння (3.65), отримаємо відповідні значення $\overline{Ц_K}$, якими заповнюємо шосту графу таблиці 3.25, а за цими значеннями та значеннями $Ц_i$ і $\overline{Ц}$ обчислюємо елементи сьомої та восьмої граф цієї таблиці.

Щільність зв’язку між ознаками $Ц$ та K обчислюємо за допомогою коефіцієнта кореляції:

$$r_{Ц/K} = \sqrt{\frac{\sum(\overline{Ц_K}-\overline{Ц})^2}{\sum(Ц_i-\overline{Ц})^2}} = \sqrt{\frac{228{,}519}{231{,}333}} = -\sqrt{0{,}987835} \approx -0{,}994.$$

(Знак “–” для попередніх розрахунків обираємо з тієї причини, що пряма (3.65) є спадною).

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про щільний зворотний зв’язок між ціною товару та його реалізованою кількістю. Коефіцієнт детермінації D є підкореневим виразом останньої формули. Цей коефіцієнт $D \approx 0{,}98$ показує, що варіація результативної ознаки (ціни товару) на 98% відбувається під впливом фактора кількості реалізованого продукту, а на 2% − під впливом не врахованих в моделі факторів.

Аналогічно, якщо побудувати емпіричні точки, то визначимо тип залежності $B(K)$ за допомогою графіка, що апроксимує дані кореляційної таблиці до функції, зображеної на рис. 3.13.

Із рисунка 3.13 можна припустити, що залежність між витратами та кількістю виробленої та реалізованої продукції має квадратичний характер:

$$\overline{B_к} = b_0 + b_1 \cdot K + b_2 \cdot K^2. \qquad (3.66)$$

Для пошуку параметрів параболічної залежності скористаємося системою нормальних рівнянь (3.42):

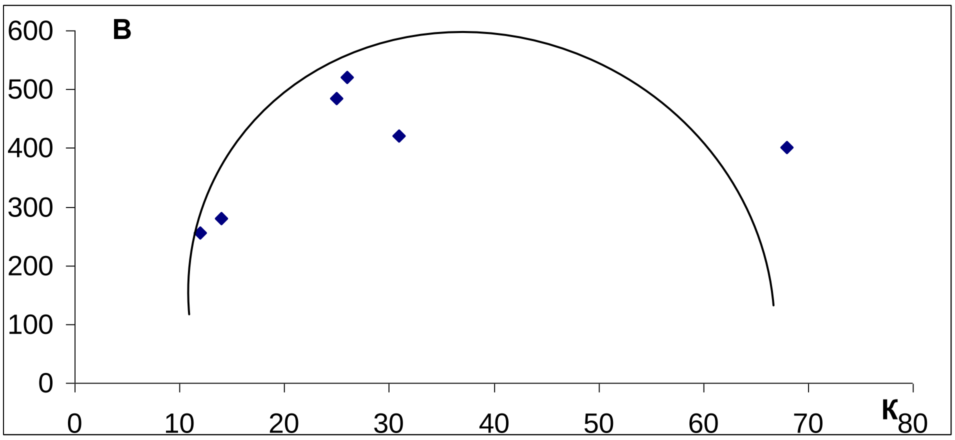


Рисунок 3.14 − Апроксимація даних $B(K)$ параболічною залежністю

$$\begin{cases} nb_0 + b_1 \sum K_i + b_2 \sum K_i^2 = \sum B_i; \\ b_0 \sum K_i + b_1 \sum K_i^2 + b_2 \sum K_i^3 = \sum K_i B_i; \\ b_0 \sum K_i^2 + b_1 \sum K_i^3 + b_2 \sum K_i^4 = \sum K_i{}^2 B_i. \end{cases}$$

Для розрахунку значень параметрів параболічної однофакторної моделі необхідно скласти відповідну допоміжну таблицю 3.26.

Таблиця 3.26 – Допоміжна розрахункова таблиця

№	K_i	B_i	K_i^2	K_i^3	K_i^4	$K_i B_i$	$K_i{}^2 B_i$
1	68	401	4626	314568	21390624	27268	1855026
2	31	420	961	29791	923521	13020	403620
3	26	520	676	17576	456976	13520	351520
4	25	484	625	15625	390625	12100	302500
5	14	280	196	2744	38416	3920	54880
6	12	255	144	1728	20736	3060	36720
$\sum$	**176**	**2360**	**7226**	**382032**	**23220898**	**72888**	**3004266**

Таким чином, система нормальних рівнянь набуває вигляду:

$$\begin{cases} 6b_0 + 176b_1 + 7226b_2 = 2360; \\ 176b_0 + 7226b_1 + 382032b_2 = 72888; \\ 7226b_0 + 382032b_1 + 23220898b_2 = 3004266. \end{cases}$$

Розв’язавши систему засобами *Mathcad* знаходимо:

$$b_0 = 2{,}471; \qquad b_1 = 24{,}788; \qquad b_2 = -0{,}279.$$

Тому шукане рівняння регресії таке:

$$\overline{B_к} = 2{,}471 + 24{,}788 \cdot K - 0{,}279 \cdot K^2. \qquad (3.67)$$

Підставляючи по черзі значення K_i в рівняння (3.67), знайдемо умовні середні значення витрат ($\overline{B_k}$) та відповідні суми квадратів відхилень (табл. 3.27), врахувавши, що $\overline{B} = \dfrac{2360}{6} \approx 393{,}333$.

Таблиця 3.27 – Допоміжна розрахункова таблиця

№	$\overline{B_k}$	$(\overline{B_K} - \overline{B})^2$	$(B_i - \overline{B})^2$
1	397,959	21,39988	58,78289
2	502,78	11978,65000	711,12890
3	458,355	4227,86000	16044,53000
4	447,796	2966,21800	7220,50500
5	294,819	9705,00800	12844,37000
6	259,751	17844,15000	19136,02000
$\sum$	**2361,46**	**46843,28**	**57015,33**

Щільність зв’язку між ознаками *B* та *K* обчислюємо за допомогою коефіцієнта кореляції так:

$$r_{B/K} = \sqrt{\frac{\sum(\overline{B_K} - \overline{B})^2}{\sum(B_i - \overline{B})^2}} = \sqrt{\frac{46843{,}28}{57015{,}33}} = \sqrt{0{,}819837} \approx 0{,}91.$$

Оскільки значення коефіцієнта кореляції близьке до одиниці, то це свідчить про щільний прямий зв’язок між витратами на виробництво та кількістю реалізованої продукції. Коефіцієнт $D \approx 0{,}91$ показує, що варіація результативної ознаки (витрат) на 91% відбувається під впливом фактора кількості реалізованого продукту, а на 9% − під впливом не врахованих в моделі факторів. Отже, побудоване рівняння регресії є адекватним реальному об'єкту.

Використовуючи рівняння (3.65) та (3.67), знаходимо формулу прибутку від реалізації продукту:

$$\overline{П_К} = \overline{Ц_К}К - \overline{В_К} = (-0{,}333К + 28{,}095315)К - 2{,}471 - 24{,}788К + 0{,}279К^2 =$$
$$= -0{,}054К^2 + 3{,}307315К - 2{,}471.$$

Прирівнюючи до нуля першу похідну отриманої функції, знаходимо оптимальне значення кількості реалізованого продукту за критерієм максимального прибутку: $(\overline{П_К})' = -0{,}108К + 3{,}307315 = 0$. Звідки $К \approx 31$ – точка максимуму, оскільки крива функції прибутку є параболою з від'ємним коефіцієнтом біля $К^2$.

Таким чином, оптимальний обсяг випуску та реалізації продукту складає 31 одиницю. За цією величиною можна обчислити оптимальну ціну продукту, за рівнянням (3.65), оптимальні витрати, за рівнянням (3.67), виручку від реалізації та оптимальний прибуток. Результати розрахунків подано в таблиці 3.28.

Таблиця 3.28 – Порівняльна таблиця

Показник	$К$	$Ц$	$В$	$ЦК$	$П$
Фактичний за останній період	12	25	255	300	45
Оптимальний	31	17,77	502	550,87	48,87
Відхилення фактичного від оптимального	-19	7,23	-247	-250,87	-3,87

Висновки

1. Лінійна залежність $\widehat{Ц} = Ц(К)$ та квадратична $\widehat{В} = В(К)$ є гарними наближеннями для вихідних даних, що засвідчується отриманими значеннями коефіцієнтів кореляції. Такий результат дозволяє прогнозувати значення кількості виробленого та реалізованого продукту при наявності значень щодо ціни одиниці продукції або витрат на цей обсяг продукції за повною собівартістю. І навпаки, маючи заплановану оптимальну кількість виготовленого продукту можна визначити, виходячи з оцінених рівнянь, оптимальну ціну, яку необхідно встановити, та витрати, що може собі дозволити виробник.

2. Прибуток підприємства на 3,87 млн. грош. од. менший оптимального за рахунок недовиробництва.

3. Кількість виробленої та реалізованої продукції менша за оптимальну на 19 тис. одиниць за рахунок перевищення ціни на 7,23 тис. горш. од.

Стратегія підприємства: зростання обсягу випуску та реалізації продукту, зменшення ціни.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Який зв’язок між економічними показниками називають балансовим?
2. Що називають умовною середньою величиною?
3. Яку залежність називають кореляційною? Наведіть приклади.
4. Що таке рівняння та лінія регресії?
5. Запишіть загальний вигляд простої лінійної вибіркової регресії.
6. За яким критерієм оцінюють параметри лінійного рівняння регресії?
7. Охарактеризуйте метод найменших квадратів для оцінювання параметрів лінійної регресії та виведіть відповідну систему нормальних рівнянь.
8. Запишіть формулу для обчислення коефіцієнта коваріації.
9. Сформулюйте основні припущення кореляційного аналізу.
10. Доведіть, що за умови виконання основних припущень кореляційного аналізу залежна змінна y має нормальний розподіл.
11. Доведіть, що розраховані за методом найменших квадратів та з використанням основних припущень кореляційного аналізу параметри лінійної регресійної моделі мають такі властивості: $$D(b_0)=\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2};\ D(b_1)=\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}.$$
12. Охарактеризуйте елементи дисперсійного аналізу.
13. Сформулюйте поняття ступеня вільності.
14. Опишіть F-критерій Фішера-Снедекора.
15. Що характеризує коефіцієнт кореляції $r_{y/x}$ і як його можна обчислити?
16. Охарактеризуйте геометричний зміст коефіцієнта кореляції та його властивості.
17. За допомогою якого показника встановлюють адекватність рівняння регресії реальному об’єкту і як його обчислюють?
18. Опишіть стандартну помилку оцінювання.
19. Опишіть метод Фішера оцінювання коефіцієнта кореляції.
20. Які дослідження приводять до нелінійних моделей? Наведіть приклади таких моделей. Спробуйте вивести одну з систем нормальних рівнянь для оцінювання параметрів нелінійної кореляційної залежності.
21. Що називають траєкторією та трендом? Що таке динамічний ряд? Які компоненти можна виділити у складі данамічного ряду?
22. Яку функцію називають моделлю тренда?
23. Охарактеризуйте динаміку абсолютних приростів.
24. Вкажіть типи економічного розвитку.

25. Якими моделями тренда описують прискорений розвиток? Опишіть ці моделі.
26. Якою трендовою моделлю описують динаміку з абсолютним прискоренням $\tilde{\varphi}$?
27. Дослідіть темпи приросту динаміки, яку описують параболічною моделлю.
28. Охарактеризуйте криві Енгеля і Філліпса та вкажіть сфери їх застосування.
29. Що називають сезонними хвилями? Для яких сфер життєдіяльності людини вони характерні?
30. Охарактеризуйте вирівнювання сезонних рядів динаміки рядом Фур'є.
31. Опишіть алгоритм розв'язання практичної задачі кореляційного аналізу.

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Завдання 3.1. Побудувати економетричну модель залежності *y* від фактора *x*. Початкові дані наведено у таблиці (відповідно до варіанта). Оцінити щільність зв'язку між змінними та адекватність побудованої моделі.

1.

i	y_i	x_i
1	6,25	1,25
2	7,5	1,5
3	8,75	2,25
4	11,25	3
5	16,25	4,5

2.

i	y_i	x_i
1	97,5	54
2	67,5	45
3	52,5	30
4	45	22,5
5	67,5	15

3.

i	y_i	x_i
1	8,75	1,75
2	10,5	2,1
3	12,25	3,15
4	15,75	4,2
5	22,75	6,3

4.

i	y_i	x_i
1	110,5	61,2
2	76,5	51
3	59,5	34
4	51	25,5
5	76,5	17

5.

i	y_i	x_i
1	11,25	2,25
2	13,5	2,7
3	15,75	4,05
4	20,25	5,4
5	29,25	8,1

6.

i	y_i	x_i
1	78	43,2
2	54	36
3	42	24
4	36	18
5	54	12

7.

i	y_i	x_i
1	13,75	2,75
2	16,5	3,3
3	19,25	4,95
4	24,75	6,6
5	35,75	9,9

8.

i	y_i	x_i
1	117	64,8
2	81	54
3	63	36
4	54	27
5	81	18

9.

i	y_i	x_i
1	16,25	3,25
2	19,5	3,9
3	22,75	5,85
4	29,25	7,8
5	42,25	11,7

10.

i	y_i	x_i
1	136,5	75,6
2	94,5	63
3	73,5	42
4	63	31,5
5	94,5	21

11.

i	y_i	x_i
1	18,75	3,75
2	22,5	4,5
3	26,25	6,75
4	33,75	9
5	48,75	13,5

12.

i	y_i	x_i
1	162,5	90
2	112,5	75
3	87,5	50
4	75	37,5
5	112,5	25

13.

i	y_i	x_i
1	21,25	4,25
2	25,5	5,1
3	29,75	7,65
4	38,25	10,2
5	55,25	15,3

14.

i	y_i	x_i
1	169	93,6
2	117	78
3	91	52
4	78	39
5	117	26

15.

i	y_i	x_i
1	23,75	4,75
2	28,5	5,7
3	33,25	8,55
4	42,75	11,4
5	61,75	17,1

16.

i	y_i	x_i
1	175,5	97,2
2	121,5	81
3	94,5	54
4	81	40,5
5	121,5	27

17.

i	y_i	x_i
1	50	10
2	60	12
3	70	18
4	90	24
5	130	36

18.

i	y_i	x_i
1	182	100,8
2	126	84
3	98	56
4	84	42
5	126	28

19.

i	y_i	x_i
1	52,5	10,5
2	63	12,6
3	73,5	18,9
4	94,5	25,2
5	136,5	37,8

20.

i	y_i	x_i
1	188,5	104,4
2	130,5	87
3	101,5	58
4	87	43,5
5	130,5	29

21.

i	y_i	x_i
1	57,5	11,5
2	69	13,8
3	80,5	20,7
4	103,5	27,6
5	149,5	41,4

22.

i	y_i	x_i
1	195	108
2	135	90
3	105	60
4	90	45
5	135	30

23.

i	y_i	x_i
1	67,5	13,5
2	81	16,2
3	94,5	24,3
4	121,5	32,4
5	175,5	48,6
∑	540	135

24.

i	y_i	x_i
1	201,5	111,6
2	139,5	93
3	108,5	62
4	93	46,5
5	139,5	31

25.

i	y_i	x_i
1	145	57,2
2	91	41
3	64	14
4	51	20,5
5	91	27

Завдання 3.2. Для виробничого підприємства відомі такі показники за 6 періодів: кількість виробленого та реалізованого продукту ***К*** (тис. одиниць), ціна ***Ц*** (тис. грошових од. за одиницю продукції), витрати ***В*** виробництва за повною собівартістю (млн. грошових од.)

Необхідно:

1) знайти кореляційну залежність ціни *Ц* та витрат *В* від кількості виробленого та реалізованого продукту *К*: $\overline{Ц_к}$ =*Ц*(*К*), $\overline{В_к}$ =*В*(*К*).

Оцінити щільність зв'язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації;

2) провести аналіз на оптимальність обсягу реалізації продукції за критерієм максимального прибутку;

3) зробити висновки та розробити стратегію підприємства на майбутні періоди.

1.

Період / Показник	1	2	3	4	5	6
К	55	44	39	34	33	27
Ц	75	266	266	280	280	300
В	3240	7011	8989	8989	8500	7776

2.

Період / Показник	1	2	3	4	5	6
К	8	6	5,6	4,3	4	3
Ц	47	135	140	140	190	200
В	340	709	730	709	690	570

3.

Період / Показник	1	2	3	4	5	6
К	25	19	5	11	8	6
Ц	26	62	62,5	62,5	80	82
В	404	640	778	729	458	377

4.

Період / Показник	1	2	3	4	5	6
К	68	31	26	25	14	12
Ц	6	17	19	19	24	25
В	401	420	520	484	280	255

5.

Період / Показник	1	2	3	4	5	6
К	19	9,3	7,6	7,4	6	5
Ц	23	78	85	85	107	108
В	400	453	485	620	620	529

6.

Період / Показник	1	2	3	4	5	6
К	1560	778	555	545	348	300
Ц	5	14	15	15	22	23
В	5538	7859	7950	8100	7413	6555

7.

Період / Показник	1	2	3	4	5	6
К	45	38,5	37,5	37	33	20
Ц	8	20	21	21	29	36
В	248	736	785	790	930	713

8.

Період / Показник	1	2	3	4	5	6
К	130	117	92	60	31	28
Ц	16	53	61	110	165	170
В	1301	4350	5432	5199	2802	2666

9.

Період / Показник	1	2	3	4	5	6
К	2700	2413	2200	2150	1066	900
Ц	0,57	1,66	1,75	1,75	2,14	2,22
В	1486	4005	3850	3763	2047	1793

10.

Період / Показник	1	2	3	4	5	6
К	48	37	25	23	22	231
Ц	1,653	4,794	7,288	8,165	9,218	10
В	55	120	137	153	173	182

11.

Період / Показник	1	2	3	4	5	6
К	589	528	367	342	327	134
Ц	2,484	7,203	10,95	12,286	13,85	15,025
В	1027	3214	3428	4380	4017	1877

12.

Період / Показник	1	2	3	4	5	6
К	90	69	48	45	43	18
Ц	3,968	11,506	17,491	19,597	22,123	24,000
В	228	715	762	980	898	403

13.

Період / Показник	1	2	3	4	5	6
К	97	73	50	47	45	19
Ц	3,306	9,588	14,476	16,33	18,436	20
В	196	599	630	812	746	353

14.

Показник \ Період	1	2	3	4	5	6
К	229	175	122	114	109	44
Ц	1,356	3,931	5,976	6,695	7,559	8,2
В	171	535	572	733	672	308

15.

Показник \ Період	1	2	3	4	5	6
К	275	211	147	137	125	51
Ц	1,306	3,787	5,757	6,45	7,282	7,9
В	198	623	666	852	781	345

16.

Показник \ Період	1	2	3	4	5	6
К	25	19	5	11	8	6
Ц	26	62	62,5	62,5	80	82
В	404	640	778	729	458	377

17.

Показник \ Період	1	2	3	4	5	6
К	19	9,3	7,6	7,4	6	5
Ц	23	78	85	85	107	108
В	400	453	485	620	620	529

18.

Показник \ Період	1	2	3	4	5	6
К	45	38,5	37,5	37	33	20
Ц	8	20	21	21	29	36
В	248	736	785	790	930	713

19.

Показник \ Період	1	2	3	4	5	6
К	2700	2413	2200	2150	1066	900
Ц	0,57	1,66	1,75	1,75	2,14	2,22
В	1486	4005	3850	3763	2047	1793

20.

Період / Показник	1	2	3	4	5	6
К	589	528	367	342	327	134
Ц	2,484	7,203	10,95	12,286	13,85	15,025
В	1027	3214	3428	4380	4017	1877

21.

Період / Показник	1	2	3	4	5	6
К	97	73	50	47	45	19
Ц	3,306	9,588	14,476	16,33	18,436	20
В	196	599	630	812	746	353

22.

Період / Показник	1	2	3	4	5	6
К	275	211	147	137	125	51
Ц	1,306	3,787	5,757	6,45	7,282	7,9
В	198	623	666	852	781	345

23.

Період / Показник	1	2	3	4	5	6
К	19	9,3	7,6	7,4	6	5
Ц	23	78	85	85	107	108
В	400	453	485	620	620	529

24.

Період / Показник	1	2	3	4	5	6
К	90	69	48	45	43	18
Ц	3,968	11,506	17,491	19,597	22,123	24,000
В	228	715	762	980	898	403

25.

Період / Показник	1	2	3	4	5	6
К	48	37	25	23	22	231
Ц	1,653	4,794	7,288	8,165	9,218	10
В	55	120	137	153	173	182

ТЕСТИ

1. Балансовий зв’язок характеризує залежності:

а) між джерелами формування ресурсів та їх використанням;

б) між показником і компонентами, що входять до нього як множники;

в) за яких кожному значенню однієї змінної відповідає множина можливих значень іншої змінної.

2. Компонентний зв’язок показників характеризує залежності:

а) між джерелами формування ресурсів та їх використанням;

б) між показником і компонентами, що входять до нього як множники;

в) при яких кожному значенню однієї змінної відповідає множина можливих значень іншої змінної.

3. Статистичний зв’язок показників характеризує залежність:

а) між джерелами формування ресурсів та їх використанням;

б) між показником і компонентами, що входять до нього як множники;

в) за якої кожному значенню однієї змінної відповідає множина можливих значень іншої змінної.

4. Умовною середньою величиною $\overline{y_x}$ називають:

а) середнє геометричне значення ознаки *Y*, обчислене для конкретного значення показника *X*;

б) середнє арифметичне значення ознаки *Y*, обчислене для конкретного значення показника *X*;

в) довільне значення ознаки *Y*.

5. Рівнянням регресії називають вираз виду:

а) $\overline{y_x} < f(x)$;

б) $\overline{y_x} > f(x)$;

в) $\overline{y_x} = f(x)$.

6. Коефіцієнтом коваріації *cov*(*x*,*y*) між змінними *x* та *y* визначається за такою залежністю:

а) $cov(x, y) = \frac{1}{n} + \bar{x} \cdot \bar{y}$;

б) $cov(x,y)=\frac{1}{n}\sum_{i=1}^{n}x_i y_i-\overline{x}\cdot\overline{y}$;

в) $cov(x,y)=\sum_{i=1}^{n}x_i y_i-\overline{x}\cdot\overline{y}$.

7. Дисперсію величини x визначають за формулою:

а) $D(X)=\frac{1}{n^2}\sum_{i=1}^{n}x_i^{\;2}-\overline{x}^{\,2}$;

б) $D(X)=\frac{1}{n}\sum_{i=1}^{n}x_i^{\;2}+\overline{x}^{\,2}$;

в) $D(X)=\frac{1}{n}\sum_{i=1}^{n}x_i^{\;2}-\overline{x}^{\,2}$.

8. При оцінці параметрів лінійної регресії використовують критерій:

а) мінімізації суми квадратів відхилень;

б) максимізації суми квадратів відхилень;

в) мінімізації суми квадратів значень змінної.

9. Яке з перерахованих тверджень не є припущенням класичного кореляційного аналізу?

а) математичне сподівання збурення дорівнює одиниці;

б) дисперсія збурення ε_i є сталою для довільного значення i;

в) $cov(\varepsilon_i, x_i)=0,\ i=\overline{1,n}$.

10. Загальна сума квадратів відхилень залежної змінної від середнього значення дорівнює:

а) $Q_e=\sum_{i=1}^{n}(y_i-\overline{y_x})^2$;

б) $Q_R=\sum_{i=1}^{n}(\overline{y_x}-\overline{y})^2$;

в) $Q=\sum_{i=1}^{n}(y_i-\overline{y})^2$.

11. Сума квадратів відхилень, зумовлена регресією, дорівнює:

а) $Q_e = \sum_{i=1}^{n}(y_i - \overline{y_x})^2$;

б) $Q_R = \sum_{i=1}^{n}(\overline{y_x} - \overline{y})^2$;

в) $Q = \sum_{i=1}^{n}(y_i - \overline{y})^2$.

12. Залишкова сума квадратів, що характеризує неврахowані фактори, дорівнює:

а) $Q_e = \sum_{i=1}^{n}(y_i - \overline{y_x})^2$;

б) $Q_R = \sum_{i=1}^{n}(\overline{y_x} - \overline{y})^2$;

в) $Q = \sum_{i=1}^{n}(y_i - \overline{y})^2$.

13. Щільність зв'язку між ознаками характеризують за допомогою:

а) коефіцієнта детермінації;
б) коефіцієнта кореляції;
в) коефіцієнта Стьюдента.

14. Оцінка адекватності регресійної моделі здійснюється за допомогою:

а) коефіцієнта детермінації;
б) коефіцієнта кореляції;
в) коефіцієнта Стьюдента.

15. Коефіцієнт кореляції обчислюється за формулою:

а) $r_{y/x} = \pm\sqrt{\dfrac{\sum_{i=1}^{n}\overline{y_x}^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$;

б) $r_{y/x} = \dfrac{\sum_{i=1}^{n}(\overline{y_x} - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$;

в) $r_{y/x} = \pm\sqrt{\dfrac{\sum\limits_{i=1}^{n}(\overline{y_x} - \overline{y})^2}{\sum\limits_{i=1}^{n}(y_i - \overline{y})^2}}$.

16. Коефіцієнт детермінації обчислюють за формулою:

а) $D = 1 + \dfrac{Q_e}{Q}$;

б) $D = \dfrac{Q_R}{Q} = 1 - \dfrac{Q_e}{Q}$;

в) $D = \dfrac{Q_e}{Q}$.

17. Якщо $D \in [0.45; 0.55]$, то для встановлення адекватності побудованої моделі реальному об'єкту використовують:

а) критерій Фішера;

б) критерій Снедекора;

в) критерій Кронекера.

18. Яка із перерахованих властивостей не є властивістю коефіцієнта кореляції?

а) якщо $r_{y/x} = \pm 1$, то кореляційна залежність лінійна, тобто всі емпіричні точки лежать на прямій;

б) $r_{y/x}\, \varepsilon\, [1, +\infty)$;

в) якщо $r_{y/x} > 0$, то кореляційний зв'язок прямий, при якому збільшення однієї змінної приводить до збільшення іншої змінної.

19. Динамічний (часовий) ряд – це:

а) таблиця значень траєкторії, для якої час змінюється дискретно;

б) таблиця значень траєкторії, для якої час змінюється неперервно;

в) будь-яка зростаюча крива.

20. Ланцюговий абсолютний приріст розраховують, як

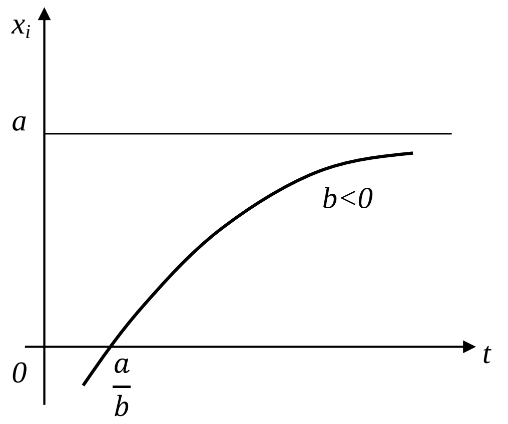
а) відношення поточного до базового (початкового) рівнів динамічного ряду;

б) різницю поточного та базового (початкового) рівнів динамічного ряду;

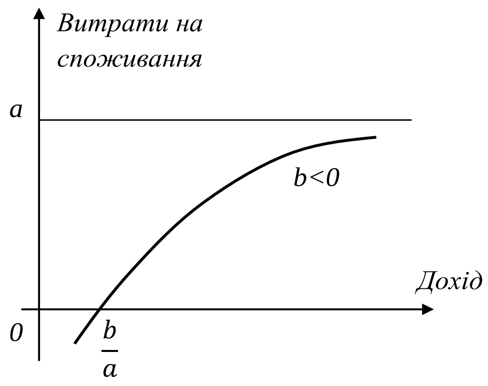
в) різницю поточного та базового (попереднього) рівнів динамічного ряду.

21. Яка із наведених кривих є кривою Філліпса?

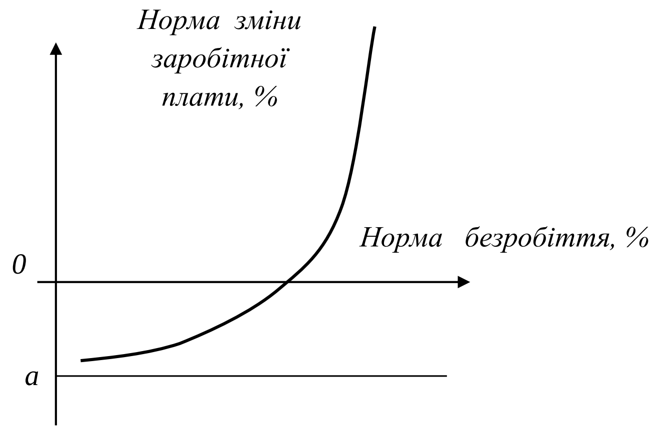
а)



б)



в)



22. Криву Енгеля використовують, щоб:

а) з’ясувати зв’язок між споживчими витратами на товари із загальними витратами або доходом;

б) розрахувати мінімальну заробітну плату;

в) з’ясувати залежність норми зміни заробітної плати від норми безробіття.

23. При моделюванні сезонних коливань економічних явищ використовують:

а) ряди Тейлора;

б) ряди Фур'є;

в) ряди Лорана.

24. Нехай статистика $t = \frac{r\sqrt{n-2}}{\sqrt{1-n^2}}$ має t-розподіл Стьюдента. Тоді коефіцієнт кореляції є значущим на рівні α, якщо:

а) $|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} < t_{1-\alpha;n-2}$;

б) $|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > t_{1-\alpha;n-2}$;

в) $|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = t_{1-\alpha;n-2}$.

25. Моделлю тренда називають:

а) функцію від часу вираження головної тенденції розвитку;

б) функцію регресії;

в) функцію Лаффера.

ТЕМА 4 МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

Економічні явища, як правило, визначаються великою кількістю факторів. Кожна економічна змінна залежить не від одного, а від багатьох факторів. Наприклад, валовий регіональний продукт залежить не лише від величини основних засобів, а й від величини оборотних фондів, величини інвестицій в основний капітал, кількості людей, зайнятих на підприємствах регіону, технологій, які використовуються, ефективності управлінських рішень тощо. Урожайність деяких культур залежить від родючості ґрунту, кількості внесених органічних і мінеральних добрив, сорту насіння, агротехнічного оброблення, природних умов тощо. Спільний вплив кількох факторів на одну результативну змінну досліджують за допомогою багатофакторних економетричних моделей.

4.1 Класична нормальна лінійна модель множинної регресії

Припустимо, що потрібно дослідити залежність змінної Y від змінних X_1, $X_2, \ldots,$ X_p. Позначимо i-те спостереження залежної змінної y_i, пояснювальних змінних $-$ x_{i1}, $x_{i2}, \ldots,$ x_{ip}. Тоді модель множинної регресії така:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i\,, \tag{4.1}$$

де ε_i – випадкова величина;

β_i – параметри моделі, $i = \overline{1,n}$, які відповідають припущенням (3.15) – (3.19).

Модель (4.1) називають класичною нормальною лінійною моделлю множинної регресії. Залежна змінна Y називається також *пояснюваною, ендогенною змінною,* незалежні змінні X_i – *пояснювальними, екзогенними змінними*.

Введення більшої кількості пояснювальних змінних ускладнює математичне оброблення даної моделі, тому виникла доцільність використання матричних позначень:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},\ X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix},\ \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},\ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{4.2}$$

Звертаємо увагу на те, що до матриці X введено додатковий стовпець, усі елементи якого дорівнюють 1, тобто умовно припускаємо, що в моделі

(4.1) вільний член β_0 множиться на фіктивну змінну x_{i0}, яка дорівнює 1 для всіх значень i.

Враховуючи позначення (4.2), модель (4.1) набуває вигляду:

$$Y = X\beta + \varepsilon . \quad (4.3)$$

Оцінкою моделі (4.1) за вибіркою є рівняння вигляду:

$$Y = Xb + e , \quad (4.4)$$

де $b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix}$, $e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$.

4.2 Оцінювання параметрів класичної регресійної моделі за методом найменших квадратів

За методом найменших квадратів параметр b обирається таким, щоб сума квадратів відхилень була мінімальною, тобто:

$$e^T e = \sum_{i=1}^{n} e_i^2 \to \min$$

або

$$Q_e = \sum_{i=1}^{n} (\overline{y_x} - y_i)^2 = \sum_{i=1}^{n} e_i^2 = (Y - Xb)^T (Y - Xb) \to min. \quad (4.5)$$

Врахувавши властивість транспонування добутку матриць $\left((AB)^T = B^T A^T\right)$, після розкриття дужок отримаємо:

$$Q_e = Y^T Y - b^T X^T Y - Y^T Xb + b^T X^T Xb.$$

Добуток $Y^T Xb$ – матриця розмірності $(1\times n)[n \times (p+1)] \times [(p+1) \times 1] =$ $= (1\times 1)$, тобто скалярна величина. Це означає, що даний добуток не змінюється при транспонуванні, тобто $Y^T Xb = (Y^T Xb)^T = b^T X^T Y$. Тому умова мінімізації набуває вигляду:

$$Q_e = Y^T Y - 2b^T X^T Y + b^T X^T Xb \to min. \quad (4.6)$$

Необхідною умовою екстремуму функції $Q(b_0,b_1,\dots,b_p)$ є

$$\frac{\partial Q}{\partial b}=\begin{pmatrix}\frac{\partial Q}{\partial b_0} & \frac{\partial Q}{\partial b_1} & \dots & \frac{\partial Q}{\partial b_p}\end{pmatrix}.$$

Для вектора частинних похідних можна довести такі твердження:

$$\frac{\partial}{\partial b}\left(b^T c\right)=c,\quad \frac{\partial}{\partial b}\left(b^T Ab\right)=2Ab,$$

де b та c – вектор-стовпці, A – симетрична матриця, в якій елементи, розташовані симетрично відносно головної діагоналі, рівні.

Справедливість наведених формул проілюструємо на прикладі. Нехай $b=\begin{pmatrix}b_1\\ b_2\end{pmatrix}$, $c=\begin{pmatrix}3\\ 4\end{pmatrix}$, $A=\begin{pmatrix}2 & 3\\ 3 & 5\end{pmatrix}$. Оскільки $b^T c=\begin{pmatrix}b_1 & b_2\end{pmatrix}\begin{pmatrix}3\\ 4\end{pmatrix}=3b_1+4b_2$ і $b^T Ab=\begin{pmatrix}b_1 & b_2\end{pmatrix}\begin{pmatrix}2 & 3\\ 3 & 5\end{pmatrix}\begin{pmatrix}b_1\\ b_2\end{pmatrix}=2b_1^2+6b_1b_2+5b_2^2$, то $\frac{\partial}{\partial b}\left(b^T c\right)=\frac{\partial}{\partial b}(3b_1+4b_2)=\begin{pmatrix}3\\ 4\end{pmatrix}=c$ та $\frac{\partial}{\partial b}\left(b^T Ab\right)=\frac{\partial}{\partial b}\left(2b_1^2+6b_1b_2+5b_2^2\right)=\begin{pmatrix}4b_1+6b_2\\ 6b_1+10b_2\end{pmatrix}=2\begin{pmatrix}2 & 3\\ 3 & 5\end{pmatrix}\begin{pmatrix}b_1\\ b_2\end{pmatrix}=2Ab$.

Тому, позначивши $c=X^TY$, а матрицю $A=X^TX$ (яка є симетричною), знайдемо

$$\frac{\partial Q}{\partial b}=-2X^TY^T+2X^TXb=0,$$

звідки одержуємо систему нормальних рівнянь в матричній формі для визначення вектора b:

$$X^TXb=X^TY. \tag{4.7}$$

Згідно з цим методом параметр b знаходиться за формулою:

$$b=(X^TX)^{-1}\cdot X^TY. \tag{4.8}$$

Тоді лінійна модель має вигляд: $\overline{Y}=X_0^Tb$ (4.9), де $X_0^T=\begin{pmatrix}1 & x_{10} & x_{20} & \dots & x_{p0}\end{pmatrix}$.

Зауваження! 1) Для зручності практичної роботи із системою (4.7) знайдемо матриці, які до неї входять. Під знаком Σ будемо розуміти знак $\sum_{i=1}^{n}$.

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1} x_{ip} \\ \dots & \dots & \dots & \dots \\ \sum x_{ip} & \sum x_{i1} x_{ip} & \dots & \sum x_{ip}^2 \end{pmatrix}. \tag{4.9}$$

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \dots \\ \sum y_i x_{ip} \end{pmatrix}. \tag{4.10}$$

В окремому випадку, використовуючи матричне рівняння (4.7) з врахуванням (4.9) та (4.10) для однієї пояснювальної змінної, нескладно одержати відому систему нормальних рівнянь (3.10):

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix} \leftrightarrow \begin{cases} nb_0 + b_1 \sum x_i = \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum y_i x_i \end{cases}.$$

2) Система (4.7) має розв’язок, якщо матриця $X^T X$ є невласною, тобто її визначник не дорівнює нулю.

Із другого зауваження можна одержати ще одну передумову множинного регресійного аналізу: *вектори значень пояснювальних змінних або стовпці матриці X повинні бути лінійно незалежними.*

Теорема Гаусса-Маркова, розглянута для парної регресійної моделі, є справедливою і для моделі множинної регресії: за умови виконання передумов множинного регресійного аналізу оцінювання методом найменших квадратів $b=(X^T X)^{-1} X^T Y$ є найбільш ефективним, тобто має дисперсію в класі лінійних несуміщених оцінок.

Приклад 4.1. Є такі дані (умовні) про видобуток вугілля за зміну на одного шахтаря Y (т), потужність пласта X_1 (м) та рівні механізації робіт X_2 (%), що характеризують видобуток вугілля на 10 шахтах (табл. 4.1)

Таблиця 4.1

i	x_{i1}	x_{i2}	y_i	i	x_{i1}	x_{i2}	y_i
1	8	5	5	6	8	8	6
2	11	8	10	7	9	6	6
3	12	8	10	8	9	4	5
4	9	5	7	9	8	5	6
5	8	7	5	10	12	7	8

Припускаючи, що між змінними Y, X_1, X_2 існує лінійна кореляційна залежність, знайти її аналітичний вираз.

Розв'язання. Позначимо

$$Y = \begin{pmatrix} 5 \\ 10 \\ 10 \\ 7 \\ 5 \\ 6 \\ 6 \\ 5 \\ 6 \\ 8 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ 1 & 12 & 8 \\ 1 & 9 & 5 \\ 1 & 8 & 7 \\ 1 & 8 & 8 \\ 1 & 9 & 6 \\ 1 & 9 & 4 \\ 1 & 8 & 5 \\ 1 & 12 & 7 \end{pmatrix}.$$

Тоді

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 11 & 12 & 9 & 8 & 8 & 9 & 9 & 8 & 12 \\ 5 & 8 & 8 & 5 & 7 & 8 & 6 & 4 & 5 & 7 \end{pmatrix} \cdot \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ 1 & 12 & 8 \\ 1 & 9 & 5 \\ 1 & 8 & 7 \\ 1 & 8 & 8 \\ 1 & 9 & 6 \\ 1 & 9 & 4 \\ 1 & 8 & 5 \\ 1 & 12 & 7 \end{pmatrix} = \begin{pmatrix} 10 & 94 & 63 \\ 94 & 908 & 603 \\ 63 & 603 & 417 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 11 & 12 & 9 & 8 & 8 & 9 & 9 & 8 & 12 \\ 5 & 8 & 8 & 5 & 7 & 8 & 6 & 4 & 5 & 7 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 10 \\ 10 \\ 7 \\ 5 \\ 6 \\ 6 \\ 5 \\ 6 \\ 8 \end{pmatrix} = \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix}.$$

Знайдемо матрицю, обернену до $X^T X$:

$$det\, X^T X = \begin{vmatrix} 10 & 94 & 63 \\ 94 & 908 & 603 \\ 63 & 603 & 417 \end{vmatrix} = 10 \begin{vmatrix} 908 & 603 \\ 603 & 417 \end{vmatrix} - 94 \begin{vmatrix} 94 & 603 \\ 63 & 417 \end{vmatrix} + 63 \begin{vmatrix} 94 & 908 \\ 63 & 603 \end{vmatrix} =$$

$$= 10 \cdot 15027 - 94 \cdot 1209 + 63 \cdot (-522) = 3738.$$

$A_{11} = \begin{vmatrix} 908 & 603 \\ 603 & 417 \end{vmatrix} = 15027$; $A_{21} = -\begin{vmatrix} 94 & 63 \\ 603 & 417 \end{vmatrix} = -1209$; $A_{31} = \begin{vmatrix} 94 & 63 \\ 908 & 603 \end{vmatrix} = -522$;

$A_{12} = -\begin{vmatrix} 94 & 603 \\ 63 & 417 \end{vmatrix} = -1209$; $A_{22} = \begin{vmatrix} 10 & 63 \\ 63 & 417 \end{vmatrix} = 201$; $A_{32} = -\begin{vmatrix} 10 & 63 \\ 94 & 603 \end{vmatrix} = -108$;

$A_{13} = \begin{vmatrix} 94 & 908 \\ 63 & 603 \end{vmatrix} = -522$; $A_{23} = -\begin{vmatrix} 10 & 94 \\ 63 & 603 \end{vmatrix} = -108$; $A_{33} = \begin{vmatrix} 10 & 94 \\ 94 & 908 \end{vmatrix} = 244$.

Таким чином, маємо:

$$\left(X^T X\right)^{-1} = \frac{1}{3738} \begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}.$$

$$b = \left(X^T X\right)^{-1} \cdot X^T Y = \frac{1}{3738} \begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix} \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix} = \begin{pmatrix} -3{,}5393 \\ 0{,}8539 \\ 0{,}3670 \end{pmatrix}.$$

Із урахуванням рівності (4.9)

$$\overline{Y} = -3{,}5393 + 0{,}854x_1 + 0{,}367x_2\,.$$

Одержане рівняння регресії показує, що при збільшенні потужності пласта X_1 (за незмінного значення X_2) лише на 1 м видобуток вугілля на одного шахтаря Y збільшується в середньому на 0,854 т, а при збільшенні рівня механізації робіт X_2 (за незмінного значення X_1) – в середньому на 0,367 т.

На практиці досить часто необхідне порівняння відокремленого впливу на залежну змінну різних пояснювальних змінних, коли останні виражають різними одиницями виміру. В цьому випадку використовують коефіцієнти еластичності E_j $\left(j = \overline{1, p}\right)$:

$$E_j = b_j \frac{\overline{x_j}}{\overline{y}}, \tag{4.11}$$

де b_j – відповідний коефіцієнт з рівняння регресії;

$\overline{x_j}$ – середнє арифметичне змінної x_j,;

$\overline{y}$ – середнє арифметичне ендогенної змінної Y.

Коефіцієнт еластичності E_j показує, на скільки відсотків зміниться в середньому Y, якщо x_j збільшити на 1%.

Приклад 4.2. За даними прикладу 4.1 порівняти відокремлений вплив на видобуток вугілля за зміну двох факторів – потужності пласта та рівня механізації робіт.

Розв'язання. Оскільки

$$\overline{x_1} = \frac{\sum_{i=1}^{10} x_{i1}}{10} = \frac{94}{10} = 9{,}4; \qquad \overline{x_2} = \frac{\sum_{i=1}^{10} x_{i2}}{10} = \frac{63}{10} = 6{,}3; \qquad \overline{y} = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{68}{10} = 6{,}8;$$

тому за формулами (4.11) коефіцієнти еластичності:

$$E_1 = 0{,}8539 \cdot \frac{9{,}4}{6{,}8} = 1{,}18; \qquad E_2 = 0{,}367 \cdot \frac{6{,}3}{6{,}8} = 0{,}34.$$

Таким чином, збільшення потужності пласта на 1% приводить до збільшення видобутку вугілля на 1,18%, а при збільшенні рівня механізації робіт на 1% видобуток вугілля збільшується на 0,34%. Останній результат показує, що на видобуток вугілля більший вплив має фактор “потужність пласту” порівняно з фактором “рівень механізації робіт”.

Приклад 4.3. Побудувати лінійну економетричну модель, що характеризує залежність між витратами обігу (Y), обсягом вантажообігу

(X_1) та фондомісткістю бази (X_2). Визначити коефіцієнти еластичності. Вихідні дані наведені в таблиці 4.2.

Таблиця 4.2

№	Витрати обігу	Обсяг вантажообігу	Фондомісткість бази
1	2,48	16,8	117,7
2	2,62	16,9	97,5
3	2,88	16,1	113,7
4	2,68	15,0	122,3
5	2,52	18,0	102,0
6	2,74	17,2	106,7
7	2,56	17,1	108,5
8	2,68	16,4	114,3
9	2,55	16,7	94,3

Розв'язання. Позначимо

$$Y = \begin{pmatrix} 2{,}48 \\ 2{,}62 \\ 2{,}88 \\ 2{,}68 \\ 2{,}52 \\ 2{,}74 \\ 2{,}56 \\ 2{,}68 \\ 2{,}55 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 16{,}8 & 117{,}7 \\ 1 & 16{,}9 & 97{,}5 \\ 1 & 16{,}1 & 113{,}7 \\ 1 & 15{,}0 & 122{,}3 \\ 1 & 18{,}0 & 102{,}0 \\ 1 & 17{,}2 & 106{,}7 \\ 1 & 17{,}1 & 108{,}5 \\ 1 & 16{,}4 & 114{,}3 \\ 1 & 16{,}7 & 94{,}3 \end{pmatrix}.$$

Тоді

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 16{,}8 & 16{,}9 & 16{,}1 & 15 & 18 & 17{,}2 & 17{,}1 & 16{,}4 & 16{,}7 \\ 117{,}7 & 97{,}5 & 113{,}7 & 122{,}3 & 102 & 106{,}7 & 108{,}5 & 114{,}3 & 94{,}3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 16{,}8 & 117{,}7 \\ 1 & 16{,}9 & 97{,}5 \\ 1 & 16{,}1 & 113{,}7 \\ 1 & 15{,}0 & 122{,}3 \\ 1 & 18{,}0 & 102{,}0 \\ 1 & 17{,}2 & 106{,}7 \\ 1 & 17{,}1 & 108{,}5 \\ 1 & 16{,}4 & 114{,}3 \\ 1 & 16{,}7 & 94{,}3 \end{pmatrix} =$$

$$=\begin{pmatrix} 9 & 150{,}2 & 977 \\ 150{,}2 & 2512{,}16 & 16266{,}1 \\ 977 & 16266{,}1 & 06762{,}6 \end{pmatrix}.$$

Застосувавши засоби *Mathcad,* знаходимо:

$$\left(X^T X\right)^{-1} = \begin{pmatrix} 171{,}3396 & -6{,}80699 & -0{,}53086 \\ -6{,}80699 & 0{,}29993 & 0{,}016595 \\ -0{,}53086 & 0{,}016595 & 0{,}002339 \end{pmatrix}.$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 16{,}8 & 16{,}9 & 16{,}1 & 15 & 18 & 17{,}2 & 17{,}1 & 16{,}4 & 16{,}7 \\ 117{,}7 & 97{,}5 & 113{,}7 & 122{,}3 & 102 & 106{,}7 & 108{,}5 & 114{,}3 & 94{,}3 \end{pmatrix} \cdot \begin{pmatrix} 2{,}48 \\ 2{,}62 \\ 2{,}88 \\ 2{,}68 \\ 2{,}52 \\ 2{,}74 \\ 2{,}56 \\ 2{,}68 \\ 2{,}55 \end{pmatrix} =$$

$$= \begin{pmatrix} 23{,}71 \\ 395{,}311 \\ 2576{,}513 \end{pmatrix}.$$

$$b = \left(X^T X\right)^{-1} \cdot X^T Y = \begin{pmatrix} 171{,}3396 & -6{,}80699 & -0{,}53086 \\ -6{,}80699 & 0{,}29993 & 0{,}016595 \\ -0{,}53086 & 0{,}016595 & 0{,}002339 \end{pmatrix} \begin{pmatrix} 23{,}71 \\ 395{,}311 \\ 2576{,}513 \end{pmatrix} = \begin{pmatrix} 3{,}826004 \\ -0{,}07058 \\ -0{,}00013 \end{pmatrix}.$$

Із врахуванням рівності (4.9)

$$\overline{Y} = 3{,}826004 - 0{,}07058 x_1 - 0{,}00013 x_2 .$$

Оскільки

$$\overline{x_1} = \frac{\sum_{i=1}^{9} x_{i1}}{9} = \frac{150{,}12}{9} = 16{,}68; \qquad \overline{x_2} = \frac{\sum_{i=1}^{9} x_{i2}}{9} = \frac{976{,}95}{9} = 108{,}55;$$

$$\bar{y} = \frac{\sum_{i=1}^{9} y_i}{9} = \frac{23{,}67}{9} = 2{,}63;$$

то за формулами (4.11) коефіцієнти еластичності:

$$E_1 = -0{,}07058 \cdot \frac{16{,}68}{2{,}63} \approx -0{,}45.$$

Збільшення обсягу вантажообігу на 1% веде до зменшення витрат обігу на 0,45%;

$$E_2 = -0{,}00013 \cdot \frac{108{,}55}{2{,}63} \approx -0{,}005.$$

При збільшенні фондомісткості бази на 1% витрати обігу зменшаться на 0,005%.

Коефіцієнт $b_0 = 3{,}826004$ характеризує граничні витрати обігу.

4.3 Коваріаційна матриця та її вибіркова оцінка

Перетворимо вектор оцінок параметрів (4.8) з урахуванням (4.3):

$$b = \left(X^T X\right)^{-1} X^T (X\beta + \varepsilon) = \left(X^T X\right)^{-1}\left(X^T X\right)\beta + \left(X^T X\right)^{-1} X^T \varepsilon = E\beta + \left(X^T X\right)^{-1} X^T \varepsilon$$

або

$$b = \beta + \left(X^T X\right)^{-1} X^T \varepsilon, \tag{4.12}$$

тобто оцінки параметрів (4.8), знайдені за вибіркою, будуть містити випадкові помилки.

Оскільки математичне сподівання оцінки b дорівнює параметру β

$$M(b) = M\left[\beta + \left(X^T X\right)^{-1} X^T \varepsilon\right] = M(\beta) + \left(X^T X\right)^{-1} X^T M(\varepsilon) = M(\beta) = \beta$$

($M(\varepsilon) = 0$ за припущенням регресійного аналізу (3.15)), то очевидно, що вектор b є несуміщеною оцінкою параметра β. Зрозумілим є той факт, що варіації оцінок параметрів визначають точність рівняння множинної регресії. Для їх вимірювання в множинному регресійному аналізі розглядають так звану *коваріаційну матрицю* вектора оцінок параметрів Σ_b, яка є матричним аналогом дисперсії однієї змінної:

$$\sum_b = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0p} \\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p0} & \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix},$$

де елементи σ_{ij} – коваріації (або кореляційні моменти) оцінок параметрів β_i та β_j. Коваріація двох змінних визначається як математичне сподівання добутку відхилення цих змінних від їх математичних сподівань. Тому

$$\sigma_{ij} = M\left[(b_i - M(b_i))(b_j - M(b_j))\right]. \qquad (4.13)$$

Каваріація характеризує як ступінь розсіювання значень двох змінних відносно їх математичних сподівань, так і взаємозв'язок цих змінних.

Оскільки оцінки параметрів, отримані методом найменших квадратів, є несуміщеними, то вираз (4.13) набуває вигляду:

$$\sigma_{ij} = M\left[(b_i - \beta_i)(b_j - \beta_j)\right].$$

Розглянуваши коваріаційну матрицю Σ_b, легко помітити, що на її головній діагоналі знаходяться дисперсії оцінок параметрів регресії, оскільки

$$\sigma_{jj} = M\left[(b_j - \beta_j)(b_j - \beta_j)\right] = M(b_j - \beta_j)^2 = \sigma_{b_j}^2. \qquad (4.14)$$

У скороченому вигляді коваріаційна матриця вектора оцінки параметрів Σ_b така:

$$\sum_b = M\left[(b-\beta)(b-\beta)^T\right]$$

(в останньому результаті легко переконатись, перемноживши вектори $(b-\beta)$ та $(b-\beta)^T$).

Враховуючи рівність (4.12) перетворюємо даний вираз:

$$\sum_b = M\left\{\left[\left(X^T X\right)^{-1} X^T \varepsilon\right]\left[\left(X^T X\right)^{-1} X^T \varepsilon\right]^T\right\} = M\left[\left(X^T X\right)^{-1} X^T \varepsilon\varepsilon^T X \left(X^T X\right)^{-1}\right] =$$

$$= \left(X^T X\right)^{-1} X^T M(\varepsilon\varepsilon^T) X \left(X^T X\right)^{-1}, \qquad (4.15)$$

оскільки елементи матриці X – невипадкові величини.

Матриця $M(\varepsilon\varepsilon^T)$ – коваріаційна матриця вектора збурень

$$\sum\nolimits_{\varepsilon} = M(\varepsilon\varepsilon^T) = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_n) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \dots & M(\varepsilon_n^2) \end{pmatrix},$$

в якій усі елементи, що не лежать на головній діагоналі, дорівнюють нулю в силу некорельованості збурень ε_i та ε_j між собою. А всі елементи, які лежать на головній діагоналі, виходячи із припущень кореляційного аналізу (3.15) та (3.16), дорівнюють одній і тій самій дисперсії σ^2:

$$M(\varepsilon_i^2) = M(\varepsilon_i - 0)^2 = D(\varepsilon_i^2) = \sigma^2.$$

Тому матриця

$$\sum\nolimits_{\varepsilon} = M(\varepsilon\varepsilon^T) = \sigma^2 E_n,$$

де E_n – одинична матриця n-го порядку (додаток И).

Врахувавши рівність (4.15), одержуємо, що коваріаційна матриця вектора оцінки параметрів:

$$\sum\nolimits_{b} = \left(X^T X\right)^{-1} X^T \sigma^2 E_n X \left(X^T X\right)^{-1} = \sigma^2 \left(X^T X\right)^{-1} \left(X^T E_n X\right)\left(X^T X\right)^{-1}$$

або

$$\sum\nolimits_{b} = \sigma^2 \left(X^T X\right)^{-1}. \tag{4.16}$$

Таким чином, *за допомогою оберненої матриці* $(X^T X)^{-1}$ *визначається не лише сам вектор* b*, але й дисперсії та коваріації його компонент.*

4.4 Доведення теореми Гаусса-Маркова. Оцінювання дисперсії збурень

Наразі у нас з’явилася можливість навести *доведення теореми Гаусса-Маркова*, сформульованої вище.

У попередньому підрозділі ми показали, що оцінка за методом найменших квадратів $b = \left(X^T X\right)^{-1} X^T Y$ є несуміщеною оцінкою для вектора параметрів β. Будь-яку іншу оцінку b_1 вектора β можна подати у вигляді

$$b_1 = \left[\left(X^T X\right)^{-1} X^T + C\right] Y,$$

де C – деяка матриця розміру $(p+1)\times n$.

Оскільки оцінки, що їх розглядають у теоремі, належать до класу несумісних, то $M(b_1)=\beta$ або $M(b_1)=M\left(\left[\left(X^TX\right)^{-1}X^T+C\right]Y\right)=\beta$.

Врахувавши, що матриця у квадратних дужках невипадкова, а за припущенням кореляційно-регресійного аналізу $M(\varepsilon)=0$, одержуємо

$$M(b_1)=\left[\left(X^TX\right)^{-1}X^T+C\right]M(Y)=\left[\left(X^TX\right)^{-1}X^T+C\right]X\beta=$$

$$=\left[\left(X^TX\right)^{-1}X^TX+CX\right]\beta=(E+CX)\beta=\beta,$$

звідки випливає, що CX=0.

Тоді

$$b_1-\beta=\left[\left(X^TX\right)^{-1}X^T+C\right]Y-\beta=\left[\left(X^TX\right)^{-1}X^T+C\right](X\beta+\varepsilon)-\beta=$$

$$=\left(X^TX\right)^{-1}X^TX\beta+CX\beta+\left[\left(X^TX\right)^{-1}X^T+C\right]\varepsilon-\beta=E\beta+\left[\left(X^TX\right)^{-1}X^T+C\right]\varepsilon-\beta=$$

$$=\beta+\left[\left(X^TX\right)^{-1}X^T+C\right]\varepsilon-\beta=\left[\left(X^TX\right)^{-1}X^T+C\right]\varepsilon.$$

За допомогою перетворень, аналогічних наведеним при одержанні формул (4.15) та (4.16), знайдемо, що коваріаційна матриця вектора оцінок b_1 дорівнює

$$\sum\nolimits_{b_1}=M\left[(b_1-\beta)(b_1-\beta)^T\right]=\sigma^2\left(X^TX\right)^{-1}+\sigma^2CC^T,$$

або, враховуючи (4.16),

$$\sum\nolimits_{b_1}=\sum\nolimits_b+\sigma^2CC^T.$$

Діагональні елементи матриці CC^T невід'ємні, оскільки це є суми квадратів елементів відповідних рядків цієї матриці. Оскільки діагональні елементи матриць $\sum_{b_1}$ та $\sum_b$ – дисперсії компонент векторів оцінок b_1 та b, то $\sigma^2_{b_{1i}}\geq\sigma^2_{b_i}$ $(i=1,2,\ldots,p+1)$. Це означає, що оцінки коефіцієнтів регресії, знайдених за методом найменших квадратів, мають найменшу дисперсію, *що й треба було довести*.

Розглянемо вектор залишків e, який дорівнює, з огляду на (4.4), $e=Y-Xb$. Із рівностей (4.3) та (4.8) маємо

$$e = (X\beta + \varepsilon) - X\left[\left(X^T X\right)^{-1} X^T (X\beta + \varepsilon)\right] =$$

$$= X\beta + \varepsilon - X\left(X^T X\right)^{-1}\left(X^T X\right)\beta - X\left(X^T X\right)^{-1} X^T \varepsilon =$$

$$= X\beta + \varepsilon - XE\beta - X\left(X^T X\right)^{-1} X^T \varepsilon$$

або

$$e = \varepsilon - X\left(X^T X\right)^{-1} X^T \varepsilon$$

(врахували, що добуток $\left(X^T X\right)^{-1} X^T X = E$, тобто дорівнює одиничній матриці E_{p+1} $(p+1)$-го порядку).

Знайдемо транспонований вектор залишків e^T. Оскільки при транспонуванні матриця $(X^T X)^{-1}$ не змінюється, тобто

$$\left[(X^T X)^{-1}\right]^T = \left[\left(X^T X^T\right)^T\right]^{-1} = (X^T X)^{-1},$$

то

$$e^T = \left[\varepsilon - X(X^T X)^{-1} X^T \varepsilon\right] = \varepsilon^T - \varepsilon^T X(X^T X)^{-1} X^T .$$

Тепер

$$M\left(e^T e\right) = M\left[\left(\varepsilon^T - \varepsilon^T X(X^T X)^{-1} X^T\right)\left(\varepsilon - \varepsilon X(X^T X)^{-1} X^T\right)\right] =$$

$$= M\left(\varepsilon^T \varepsilon\right) - M\left|\varepsilon^T X\left(X^T X\right)^{-1} X^T \varepsilon\right| - M\left|\varepsilon^T X\left(X^T X\right)^{-1} X^T \varepsilon\right| +$$

$$+ M\left|\varepsilon^T X\left(X^T X\right)^{-1}\left(X^T X\right)\left(X^T X\right)^{-1} X^T \varepsilon\right|.$$

Оскільки останні два доданки взаємно знищуються, то

$$M\left(e^T e\right) = M\left(\varepsilon^T \varepsilon\right) - M\left|\varepsilon^T X\left(X^T X\right)^{-1} X^T \varepsilon\right|. \qquad (4.17)$$

Перший доданок виразу (4.17)

$$M\left(\varepsilon^T \varepsilon\right) = M\left(\sum_{i=1}^{n} \varepsilon_i^2\right) = \sum_{i=1}^{n} M\left(\varepsilon_i^2\right) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2, \qquad (4.18)$$

оскільки за припущенням кореляційно-регресійного аналізу

$$M\left(\varepsilon_i^2\right)=M\left(\varepsilon_i-0\right)^2=D\left(\varepsilon_i\right)=\sigma^2.$$

Матриця $B=X\left(X^TX\right)^{-1}X^T$ симетрична, оскільки

$$B^T=\left[X\left(X^TX\right)^{-1}X^T\right]^T=X\left(X^TX\right)^{-1}X^T, \text{ тобто } B^T=B.$$

Тому $\varepsilon^TB\varepsilon$ є квадратичною формою $\sum\limits_{i,j=1}^{n}b_{ij}\varepsilon_i\varepsilon_j$; її математичне сподівання

$$M\left(\varepsilon^TB\varepsilon\right)=M\sum_{i,j=1}^{n}b_{ij}\varepsilon_i\varepsilon_j=\sum_{i,j=1}^{n}b_{ij}M(\varepsilon_i\varepsilon_j).$$

Останню суму можна розбити на дві складові суми елементів на головній діагоналі матриці *B* та поза нею:

$$M\left(\varepsilon^TB\varepsilon\right)=\sum_{i=1}^{n}b_{ii}M(\varepsilon_i{}^2)+\sum_{i,j=1\,(i\neq j)}^{n}b_{ij}M(\varepsilon_i\varepsilon_j).$$

Із припущення (3.17) кореляційно-регресійного аналізу випливає, що другий доданок останньої суми дорівнює нулю. Сума діагональних елементів матриці *B* утворює слід матриці *tr*(*B*). Тому

$$M\left(\varepsilon^TB\varepsilon\right)=\sum_{i=1}^{n}b_{ii}M(\varepsilon_i{}^2)=\sigma^2\sum_{i=1}^{n}b_{ii}=\sigma^2tr(B). \tag{4.19}$$

Врахувавши, що $B=X\left(X^TX\right)^{-1}X^T$, отримуємо

$$M\left(\varepsilon^TB\varepsilon\right)=\sigma^2tr(B)=\sigma^2tr\left(X\left(X^TX\right)^{-1}X^T\right)=\sigma^2tr\left(\left(X^TX\right)^{-1}X^TX\right)=$$

$$=\sigma^2tr\left(E_{p+1}\right)=\sigma^2(p+1),$$

оскільки слід матриці не змінюється при її транспонуванні, а слід одиничної матриці дорівнює порядку цієї матриці.

Із формул (4.17), (4.18) та (4.19) випливає:

$$M\left(e^Te\right)=M\left(\sum_{i=1}^{n}e_i^2\right)=n\sigma^2-\sigma^2(p+1)=(n-p-1)\sigma^2. \tag{4.20}$$

Рівність (4.20) означає, що несуміщена оцінка s^2 параметра σ^2 або вибіркова залишкова дисперсія s^2 визначається за формулою:

$$s^2 = \frac{e^T e}{n-p-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1}. \qquad (4.21)$$

У знаменнику (4.21) стоїть n–(p+1). Це означає, що (p+1) ступінь вільності втрачається при визначенні невідомих параметрів, число яких разом із вільним членом дорівнює (p+1).

4.5 Перевірка двофакторної регресії на адекватність за допомогою коефіцієнта детермінації. Критерій Фішера

Як і у випадку парної регресійної моделі сума квадратів відхилень залежної змінної від середнього може бути розвинена на дві складові:

$$Q = Q_R + Q_e,$$

де Q_R та Q_e – відповідно сума квадратів відхилень, зумовлених регресією, та залишкова сума квадратів, що характеризує вплив не врахованих у моделі факторів.

Знайдемо формули для пошуку відповідних сум квадратів.

$$Q = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = Y^T Y - n\bar{y}^2 \qquad (4.22)$$

(оскільки $\sum_{i=1}^{n} y_i^2 = y_1^2 + y_2^2 + ... + y_n^2 = (y_1 \quad y_2 \quad ... \quad y_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = Y^T Y$).

Із врахуванням (4.6) маємо

$$Q_e = \sum_{i=1}^{n} (y_i - \overline{y_x})^2 = Y^T Y - 2b^T X^T Y + b^T X^T X b = Y^T Y - b^T X^T Y \quad (4.23)$$

(оскільки згідно із рівністю (4.7) $X^T X b = X^T Y$).

Тоді

$$Q_R = Q - Q_e = Y^T Y - n\overline{y}^2 - \left(Y^T Y - b^T X^T Y\right) = b^T X^T Y - n\overline{y}^2. \qquad (4.24)$$

Рівняння множинної регресії є значущим (іншими словами – гіпотеза H_0 про рівність нулю параметрів регресійної моделі відкидається), якщо (враховуючи (3.22), при $m = p + 1$)

$$F = \frac{Q_R(n-p-1)}{Q_e p} > F_{\alpha;p;n-p-1}, \qquad (4.25)$$

де $F_{\alpha;p;n\text{-}p\text{-}1}$ — табличне значення F-критерію Фішера-Снедекора, а Q_R та Q_e визначаються за формулами (4.23) та (4.24).

У підрозділі 3.5 було введено коефіцієнт детермінації D як одну із найбільш ефективних оцінок адекватності регресійної моделі, міри якості рівняння регресії, характеристики його прогностичної сили.

$$D = \frac{Q_R}{Q} = \frac{b^T X^T Y - n\overline{y}^2}{Y^T Y - n\overline{y}^2} \qquad (4.26)$$

або

$$D = 1 - \frac{Q_e}{Q} = 1 - \frac{(Y - Xb)^T (Y - Xb)}{\left(Y - \overline{Y}\right)^T \left(Y - \overline{Y}\right)} = 1 - \frac{e^T e}{y^T y}, \qquad (4.27)$$

де $e = Y - Xb$, $\overline{Y} = \left(\overline{y},\ \ \overline{y},\ \ ...,\ \ \overline{y}\right)$, $y = \left(Y - \overline{Y}\right)$ – n-вимірні вектори;

$$e^T e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \overline{y_x})^2, \quad y^T y = \sum_{i=1}^{n} (y_i - \overline{y})^2 .$$

Нагадаємо, що *коефіцієнт детермінації D характеризує частку варіації залежної змінної, зумовленої регресією. Чим ближчий цей коефіцієнт до одиниці, тим краще регресія описує залежність між пояснювальними та залежною змінними*.

Разом із тим використання лише одного коефіцієнта детермінації для вибору найкращого рівняння регресії може виявитись недостатнім. На практиці зустрічаються випадки, коли неправильно визначена модель регресії може дати досить високий коефіцієнт детермінації.

Недоліком коефіцієнта детермінації є той факт, що він збільшується при додаванні нових пояснювальних змінних, хоча це не обов’язково означає покращення якості регресійної моделі. В цьому сенсі доцільніше використовувати *скоригований (адаптований, виправлений) коефіцієнт детермінації* $\overline{D}$, що визначається за формулою

$$\overline{D} = 1 - \frac{n-1}{n-p-1}(1-D) \tag{4.28}$$

або, врахувавши рівність (4.27),

$$\overline{D} = 1 - \frac{(n-1)e^T e}{(n-p-1)y^T y}. \tag{4.29}$$

З рівності (4.28) випливає, що чим більше число пояснювальних змінних *p*, тим менший коефіцієнт $\overline{D}$ порівняно із *D*. Скоригований коефіцієнт детермінації $\overline{D}$, на відміну від *D*, може зменшуватися при введенні в модель нових пояснювальних змінних, що не мають істотного впливу на залежну змінну. Однак навіть збільшення скоригованого коефіцієнта детермінації при введенні в модель нової пояснювальної змінної не завжди означає, що її коефіцієнт регресії є значущим (це відбувається лише в тому випадку, коли відповідне значення *t*-статистики більше за одиницю (за абсолютною величиною), тобто $|t| > 1$). Іншими словами, збільшення $\overline{D}$ ще не означає покращення якості регресійної моделі.

Якщо є відомим коефіцієнт детермінації *D*, то критерій значущості (4.25) рівняння регресії може бути записаний у вигляді (критерій Фішера):

$$F = \frac{D(n-p-1)}{(1-D)p} > F_{\alpha;k_1;k_2}, \tag{4.30}$$

де $k_1 = p$, $k_2 = n - p - 1$, оскільки в рівнянні множинної регресії разом із вільним членом оцінюється $m = p + 1$ параметрів, $F_{\alpha;\,k1;\,k2}$ – табличне значення функції Фішера-Снедекора.

Приклад 4.4. За даними прикладу 4.1 визначити множинний коефіцієнт детермінації та перевірити значущість одержаного рівняння регресії *Y* за X_1 та X_2 на рівні $\alpha = 0{,}05$. Перевірити значущість рівняння регресії за критерієм Фішера.

Розв'язання. Обчислимо добутки векторів (див. приклад 4.1):

$$b^T X^T Y = \begin{pmatrix} -3{,}54 & 0{,}854 & 0{,}367 \end{pmatrix} \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix} = -3{,}54 \cdot 68 + 0{,}854 \cdot 664 + 0{,}367 \cdot 445 = 489{,}65$$

та $Y^TY=\sum_{i=1}^{10} y_i^2=496$. Із таблиці 4.1 знаходимо також $\sum_{i=1}^{10} y_i=68$, звідки

$$\bar{y}=\frac{\sum_{i=1}^{10} y_i}{n}=\frac{68}{10}=6{,}8 \text{ (т)}.$$

Тепер за (4.26) множинний коефіцієнт детермінації дорівнює:

$$D=\frac{489{,}65-10\cdot 6{,}8^2}{496-10\cdot 6{,}8^2}=0{,}811.$$

Такий коефіцієнт детермінації свідчить про те, що варіація досліджуваної змінної Y – видобуток вугілля за зміну на одного робітника − на 81,1% залежить від потужності пласта X_1 та рівня механізації робіт X_2, 18,9% припадає на не враховані в моделі фактори. Отже, одержане в прикладі 4.1 рівняння регресії є значущим.

Перевіримо значущість рівняння регресії за критерієм Фішера. Фактичне значення критерію за (4.30):

$$F=\frac{0{,}811(10-2-1)}{(1-0{,}811)\cdot 2}=15{,}0$$

більше табличного $F_{0{,}05;2;7}=4{,}74$; визначеного на рівні значимості α = 0,05 при k_1 = 2 та k_2 = 10 – 2 – 1 = 7 ступенях вільності (додаток Д), тобто рівняння регресії є значущим і залежна змінна Y достатньо якісно описується внесеними в модель змінними X_1 та X_2.

Проробивши аналогічні розрахунки за даними прикладу 3.5 для однієї пояснювальної змінної X_1, в прикладі 3.6 було обчислено коефіцієнт детермінації D' = 0,75. Порівнюючи значення D та D' можна сказати, що внесення в модель другої пояснювальної змінної X_2 неістотно збільшило величину коефіцієнта детермінації, який визначає якість моделі.

За формулою (4.28) обчислимо скоригований коефіцієнт детермінації:

при $p=1$ $$\bar{D}=1-\frac{9}{8}(1-0{,}75)\approx 0{,}71875;$$

при $p=2$ $$\bar{D}=1-\frac{9}{8}(1-0{,}811)\approx 0{,}757.$$

Легко бачити, що хоча скоригований коефіцієнт детермінації і збільшився при внесенні в модель пояснювальної змінної X_2 (рівня механізації робіт), та це ще не свідчить про значущість коефіцієнта b_2.

Значущість коефіцієнта регресії b_j можна перевірити, якщо врахувати, що статистика $\frac{(b_j - \beta_j)}{s_{b_j}}$ має t-розподіл Стьюдента із $k = n - p - 1$ ступенями вільності. Тому b_j істотно відрізняється від нуля (гіпотеза H_0 про рівність параметра β_j нулю, тобто H_0: $\beta_j = 0$, відкидається) на рівні значимості α, якщо $|t| = \frac{|b_j|}{s_{b_j}} > t_{1-\alpha;n-p-1}$, де $t_{1-\alpha;n-p-1}$ – табличне значення t-критерію Стьюдента, визначене на рівні значимості α при числі ступенів вільності $k = n - p - 1$.

У загальній постановці гіпотеза H_0 про рівність параметра β_j заданому числу β_{j0}, тобто H_0: β_{j0}=0, відкидається, якщо

$$|t| = \frac{|b_j - \beta_{j0}|}{s_{b_j}} > t_{1-\alpha;n-p-1}. \tag{4.31}$$

Тому довірчий інтервал для параметра β_j:

$$b_j - t_{1-\alpha;n-p-1} \cdot s_{b_j} \le \beta_j \le b_j + t_{1-\alpha;n-p-1} \cdot s_{b_j}. \tag{4.32}$$

Приклад 4.5. За даними прикладу 4.3 визначити множинний коефіцієнт детермінації та перевірити значущість одержаного рівняння регресії Y за X_1 та X_2 на рівні $\alpha = 0{,}05$.

Розв'язання. Обчислимо добутки векторів (див. приклад 4.3):

$$b^T X^T Y = (3{,}826004 \quad -0{,}07058 \quad -0{,}00013)\begin{pmatrix} 23{,}71 \\ 395{,}311 \\ 2576{,}513 \end{pmatrix} = 62{,}48935$$

та $Y^T Y = \sum_{i=1}^{9} y_i^2 = 62{,}5881$.

Із таблиці 4.2 знаходимо також $\bar{y} = \frac{\sum_{i=1}^{9} y_i}{9} = \frac{23{,}67}{9} = 2{,}63$

Тепер за (4.26) множинний коефіцієнт детермінації

$$D = \frac{62{,}48935 - 9 \cdot 2{,}63^2}{62{,}5881 - 9 \cdot 2{,}63^2} = \frac{0{,}23725}{0{,}336} \approx 0{,}7\,.$$

Такий коефіцієнт детермінації свідчить про те, що варіація досліджуваної змінної Y – витрати обігу − на 70% залежить від обсягу вантажообігу X_1 та фондомісткості бази X_2, 30% припадає на невраховані в моделі фактори. Отже, одержане в прикладі 4.3 рівняння регресії є значущим.

4.6 Моделювання нелінійної множинної регресії. Виробнича функція Кобба-Дугласа. Коефіцієнти часткової еластичності

До цих пір ми розглядали лінійні регресійні моделі. Однак співвідношення між соціально-економічними явищами та процесами далеко не завжди можна виразити лінійними функціями, оскільки при цьому можуть виникати невиправдано великі помилки.

Так, наприклад, нелінійними є *виробничі функції* (залежності між обсягом виробничої продукції та основними факторами виробництва – працею, капіталом і т. п.), *функції попиту* (залежність між попитом на товари чи послуги та їх цінами або доходом) та інші.

Для оцінювання параметрів нелінійних моделей використовують два підходи.

Перший підхід заснований на *лінеаризації* моделі і полягає в тому, що за допомогою відповідних перетворень досліджувану залежність подають у вигляді *лінійного* співвідношення між *перетвореними* змінними.

Другий підхід зазвичай застосовують у випадку, коли підібрати відповідне лінеаризоване перетворення не вдалося. В цьому випадку застосовують методи *нелінійної оптимізації* на базі вихідних змінних.

Для лінеаризації моделі в межах першого підходу можуть використовуватися як моделі нелінійні за змінними, так і нелінійні за параметрами.

Якщо модель *нелінійна за змінними*, то введенням нових змінних її можна звести до лінійної моделі, для оцінювання параметрів якої можна використовувати звичайний метод найменших квадратів.

Так, наприклад, якщо нам потрібно оцінити параметри регресійної моделі

$$y_i = \beta_0 + \beta_1 \sqrt[3]{x_{i1}} + \beta_2 \ln x_{i2} + \varepsilon_i\,, \;\; i = \overline{1,n}\,,$$

то, зробивши заміну $Z_1 = \sqrt[3]{X_1}$, $Z_2 = ln\, X_2$, отримаємо лінійну модель

$$y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \varepsilon_i\,, \qquad i = \overline{1,n}\,,$$

параметри якої знаходять методом найменших квадратів за формулою (4.8).

Слід, однак, відмітити і недолік такої заміни, пов'язаний із тим, що ми отримаємо вектор оцінок b не з умови мінімізації суми квадратів відхилень для вихідних змінних, а з умови мінімізації квадратів відхилень для перетворених змінних. У зв'язку з цим необхідне певне уточнення одержаних оцінок.

Більш складною проблемою є *нелінійність* моделі *за параметрами*, оскільки безпосереднє застосування методу найменших квадратів для їх оцінювання неможливе. До таких моделей можна віднести, наприклад, *мультиплікативну (степеневу)* модель

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} \varepsilon_i, \ i = \overline{1,n}, \tag{4.33}$$

експоненціальну модель

$$y_i = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}} \varepsilon_i, \ i = \overline{1,n} \tag{4.34}$$

та ін.

В окремих випадках, шляхом необхідної заміни, ці моделі можна звести до лінійної форми. Так, моделі (4.33) та (4.34) можуть бути зведені до лінійних шляхом логарифмування обох частин рівнянь. Тоді, наприклад, модель (4.33) стане такою:

$$ln\, y_i = ln\, \beta_0 + \beta_1 ln\, x_{i1} + \beta_2 ln\, x_{i2} + ln\, \varepsilon_i, \ i = \overline{1,n}. \tag{4.35}$$

Відмітимо, що до моделі (4.35) вже можна застосовувати звичайні методи дослідження лінійної регресії, розглянуті вище. Однак слід зауважити, що критерії значущості параметрів, які застосовують для нормальної лінійної регресії, потребують, щоб нормальний закон розподілу в моделях (4.33), (4.34) мав логарифм вектора збурень ε (тобто $\ln \varepsilon \approx N_n\left(0, \sigma^2 E_n\right)$), а зовсім не ε. Іншими словами, вектор збурень ε повинен мати логарифмічно нормальний розподіл.

Зауважимо, що до моделі

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} + \varepsilon_i, \ i = \overline{1,n}, \tag{4.36}$$

яку розглядають як альтернативу моделі (4.33), викладені вище методи дослідження лінійної регресії вже стають неможливими, оскільки модель (4.36) не можна звести до лінійного вигляду. В цьому випадку використовують спеціальні (ітеративні) процедури оцінювання параметрів.

Як приклад використання лінеаризувального перетворення регресії розглянемо ***виробничу функцію Кобба-Дугласа***

$$Y = AK^{\alpha} L^{\beta},$$

де Y – обсяг виробництва;

K – витрати капіталу;

L – витрати праці.

Показники α та β є коефіцієнтами часткової еластичності обсягу виробництва Y відповідно до витрат капіталу та праці. Це означає, що при збільшенні одних лише витрат капіталу (праці) на 1% обсяг виробництва збільшиться на α% (β%).

Зауважимо, що *коефіцієнт часткової еластичності* $E_{x_i}(y)$ функції $y = f(x_1, x_2, ..., x_n)$ відносно змінної x_i $\left(i = \overline{1,n}\right)$ називають границею відношення відносного часткового приросту функції до відносного приросту цієї змінної, коли останній прямує до нуля, тобто

$$E_{x_i}(y) = \lim_{\Delta x_i \to 0} \left(\frac{\Delta_{x_i} y}{y} \div \frac{\Delta x_i}{x_i} \right) = \lim_{\Delta x_i \to 0} \left(\frac{\Delta_{x_i} y}{\Delta x_i} \cdot \frac{x_i}{y} \right) = \frac{x_i}{y} \cdot \frac{\partial y}{\partial x_i}.$$

Для функції Кобба-Дугласа маємо:

$$E_K(y) = \frac{K}{AK^{\alpha} L^{\beta}} \cdot AL^{\beta} \alpha K^{\alpha - 1} = \alpha; \qquad E_L(y) = \frac{L}{AK^{\alpha} L^{\beta}} \cdot A\beta L^{\beta - 1} K^{\alpha} = \beta.$$

Враховуючи вплив випадкових збурень, функцію Кобба-Дугласа можна записати у вигляді

$$Y = AK^{\alpha} L^{\beta} \varepsilon. \tag{4.37}$$

Таку мультиплікативну (степеневу) модель легко звести до лінійної шляхом логарифмування обох частин рівняння (4.37). Тоді для i-го спостереження отримаємо

$$\ln y_i = \ln A + \alpha \ln K_i + \beta \ln L_i + \ln \varepsilon_i, \ i = \overline{1,n}. \tag{4.38}$$

Якщо в моделі (4.37) $\alpha + \beta = 1$ (тобто модель така, що при розширенні масштабу виробництва – збільшенні витрат капіталу та праці в певну кількість разів обсяг виробництва збільшується в те саме число разів) функцію Кобба-Дугласа подають у вигляді

$$Y = AK^{\alpha} L^{1-\alpha} \varepsilon$$

або

$$\frac{Y}{L} = A\left(\frac{K}{L}\right)^{\alpha} \varepsilon \,. \qquad (4.39)$$

Таким чином, отримуємо залежність продуктивності праці $\left(\frac{Y}{L}\right)$ від її капіталоозброєності $\left(\frac{K}{L}\right)$. Для оцінювання параметрів моделі (4.39) шляхом логарифмування зводимо її (для i -го спостереження) до вигляду

$$ln\left(\frac{Y}{L}\right)_i = ln\,A + \alpha\, ln\left(\frac{K}{L}\right)_i + ln\,\varepsilon_i\,, \quad i = \overline{1,n}\,. \qquad (4.40)$$

Функція Кобба-Дугласа з урахуванням технічного прогресу є такою:

$$Y = A_0 K^{\alpha} L^{\beta} e^{\theta t}, \qquad (4.41)$$

де A_0 – параметр масштабування і початкової ефективності виробництва (загальна базова ефективність);

t – час;

θ – параметр, який відображає темп приросту обсягу виробництва завдяки технічному прогресу (автономний темп).

Модель (4.41) зводиться до лінійного вигляду аналогічно моделі (4.37).

Ступінь однорідності функції Кобба-Дугласа визначають сумою $\alpha+\beta$. Для статичної функції приймають $\alpha+\beta$ =1 (постійна віддача ресурсів), для динамічної функції – $\alpha+\beta$ >1. Остання умова відображає зростання загальної ефективності виробничих факторів в економічній динаміці.

Здійснимо нормування коефіцієнтів еластичності ресурсів:

$$\frac{\beta}{\alpha+\beta} = \gamma\,, \qquad \frac{\alpha}{\alpha+\beta} = 1-\gamma\,,$$

тоді динамічна функція набуває вигляду:

$$Y = A_0 e^{\theta t}\left(L^{\gamma} K^{1-\gamma}\right)^{\alpha+\beta}. \qquad (4.42)$$

Прологарифмуємо обидві частини рівняння (4.41):

$$ln\,Y = ln\,A_0 + \theta t + \alpha\, ln\,K + \beta\, ln\,L\,.$$

Отриману рівність продиференціюємо за t. Враховуючи, що для траєкторії $Q(t)\dfrac{d\ln Q(t)}{dt}=\tilde{\rho}(t)$ – неперервний темп приросту, отримаємо рівність:

$$\rho_Y=\theta+\alpha\rho_K+\beta\rho_L, \tag{4.43}$$

де ρ_Y – темп приросту національного доходу;

ρ_L, ρ_K – темпи приросту затрат праці та основних фондів, відповідно.

Рівняння (4.43) показує, що темпи приросту національного доходу є сумою автономного темпу і зваженої суми темпів приросту виробничих факторів.

Якщо $\rho_Y > \rho_L$ і $\rho_Y > \rho_K$, то це означає збільшення ефективності обох виробничих факторів (зростання ефективності праці та зростання фондовіддачі).

У разі $\rho_K > \rho_Y > \rho_L > 0$ $\theta > 0$ може бути лише при $\alpha < 1$, тобто за умови зменшення фондовіддачі.

Приклад 4.6. Є дані (умовні) про продуктивність праці $\left(\dfrac{Y}{L}\right)$ та капіталоозброєність $\left(\dfrac{K}{L}\right)$ для n = 10 підприємств Вінницької області (табл. 4.3). Оцінити виробничу функцію Кобба-Дугласа у вигляді (4.39) на рівні β = 0,2.

Таблиця 4.3 – Умовні дані для прикладу 4.6

№	$\frac{Y}{L}$	$\frac{K}{L}$
1	0,147619	7,005952
2	0,15503	5,769231
3	0,178882	7,062112
4	0,178667	8,153333
5	0,14	5,666667
6	0,159302	6,203488
7	0,149708	6,345029
8	0,163415	6,969512
9	0,152695	5,646707
10	0,135948	7,24183

Розв'язання. Від вихідних значень змінних перейдемо до їх натуральних логарифмів та, використовуючи метод найменших квадратів, розрахуємо оцінки параметрів моделі (4.40). Для цього складемо систему нормальних рівнянь, скориставшись допоміжною розрахунковою таблицею (табл. 4.4).

Таблиця 4.4

№ параметра	$\ln\frac{Y}{L}$	$\ln\frac{K}{L}$	$\left(\ln\frac{K}{L}\right)^2$	$\left(\ln\frac{Y}{L}\right)\left(\ln\frac{K}{L}\right)$
1	-1,910543	1,946767	3,789901	-3,719382
2	-1,86433	1,752499	3,071252	-3,267236
3	-1,720369	1,954728	3,820963	-3,362855
4	-1,720369	2,098386	4,403224	-3,609999
5	-1,966113	1,73466	3,009045	-3,410537
6	-1,838851	1,825033	3,330746	-3,355964
7	-1,89712	1,847667	3,413874	-3,505246
8	-1,814005	1,941615	3,76987	-3,5221
9	-1,877317	1,731124	2,996792	-3,24987
10	-1,9951	1,979897	3,919994	-3,950094
$\sum$	**-18,60412**	**18,81238**	**35,52566**	**-34,95328**

У нашому випадку

$$\begin{cases} 10\ln A + 18{,}81238\alpha = -18{,}60412 \\ 18{,}81238\ln A + 35{,}52566\alpha = -34{,}95328 \end{cases}.$$

Розв’язавши систему методом Гаусса, знаходимо:

$$\alpha \approx 0{,}34; \qquad \ln A = -2{,}5;\ A \approx 0{,}08195.$$

Таким чином, виробнича функція Кобба-Дугласа набуває вигляду:

$$\overline{\left(\frac{Y}{L}\right)} = 0{,}08195\left(\frac{K}{L}\right)^{0{,}34}.$$

Коефіцієнт еластичності α = 0,34 говорить про те, що при зміні капіталоозброєності праці на 1% продуктивність праці на підприємствах збільшується в середньому на 0,34%.

Легко показати, що рівняння $\overline{\frac{Y}{L}} = -2{,}5 + 0{,}34 ln\left(\frac{K}{L}\right)$ є значущим, оскільки за допомогою t-критерію Стьюдента значення t-відношення t = 1,5 більше за табличне значення $t_{1-0,3;8} = 1{,}4$ (додаток Ж).

4.7 Часткова кореляція

Якщо змінні корелюють одна з одною, то на значення коефіцієнта кореляції частково впливають інші змінні. У зв’язку з цим часто виникає необхідність дослідити часткову кореляцію між змінними при вилученні (елімінуванні) впливу однієї або кількох змінних.

Вибірковим частковим коефіцієнтом кореляції (або просто частковим коефіцієнтом кореляції) між змінними X_i та X_j при фіксованих значеннях решти (p-2) змінних називають вираз

$$r_{ij.1,2,\dots,p} = \frac{-q_{ij}}{\sqrt{q_{ii}q_{jj}}}, \qquad (4.44)$$

де q_{ii} та q_{jj} – алгебраїчні доповнення елементів r_{ii} та r_{jj} матриці вибіркових коефіцієнтів кореляції

$$q_p = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix},$$

а r_{ij} визначають за формулами (3.23) – (3.25). Зокрема, у випадку трьох змінних (n = 3) із (4.44) випливає, що

$$r_{ij.k} = \frac{r_{ij}{}^2 - r_{ik}r_{jk}}{\sqrt{\left(1-r_{ik}{}^2\right)\left(1-r_{jk}{}^2\right)}}. \qquad (4.45)$$

Пояснимо отриману формулу (4.45). Припустимо, що є звичайна регресійна модель $x_i = \beta_0 + \beta_1 x_j + \beta_2 x_k + \varepsilon_i$ та необхідно оцінити кореляцію між залежною змінною X_i та пояснювальною змінною X_j при виключенні (елімінуванні) впливу іншої пояснювальної змінної X_k. Із цією метою знайдемо рівняння парної регресії X_i за X_k ($\overline{x_i} = b_0 + b_1 x_k$) та X_j за X_k ($\overline{x_i} = b_0' + b_1' x_k$), а потім виключимо вплив змінної X_k, вибравши залишки $e_{x_i} = x_i - \overline{x_i}$ та $e_{x_j} = x_j - \overline{x_j}$. Очевидно, що коефіцієнти кореляції між залишками e_{x_i} та e_{xj} відображають щільність часткової кореляції між змінними X_i та X_j при вилученні змінної X_k. Можна показати, що знайдений за формулою (3.23) звичайний коефіцієнт кореляції між залишками e_{x_i} та e_{x_j} дорівнює частковому коефіцієнту кореляції $r_{ij.k}$, визначеному за формулою (4.45).

Частковий коефіцієнт кореляції $r_{ij.1,2,...,p}$, як і парний коефіцієнт r_{ij}, може набувати значень від -1 до +1. Окрім того, частковий коефіцієнт кореляції, обчислений за вибіркою обсягу n, має такий самий розподіл, як і звичайний коефіцієнт кореляції r_{ij}, обчислений за $n' = n - p + 2$ спостереженнями. Тому значущість часткового коефіцієнта кореляції $r_{ij.1,2,...,p}$ оцінюють так само, як і звичайного коефіцієнта кореляції r, але при цьому обирають $n' = n - p + 2$.

Приклад 4.7. Для дослідження залежності між продуктивністю праці (X_1), віком (X_2) та виробничим стажем (X_3) було проведено вибірку із 100 робітників однієї і тієї ж спеціальності. Обчислені парні коефіцієнти кореляції виявились значущими та склали: r_{12} = 0,2; r_{13} = 0,41; r_{23} = 0,82. Обчислити часткові коефіцієнти кореляції та оцінити їх значущість на рівні α = 0,05.

Розв'язання. За формулою (4.45) визначимо часткові коефіцієнти кореляції

$$r_{12.3} = \frac{0{,}2^2 - 0{,}41 \cdot 0{,}82}{\sqrt{(1-0{,}41)^2 (1-0{,}82)^2}} = -0{,}26;$$

$$r_{13.2} = \frac{0{,}41^2 - 0{,}2 \cdot 0{,}82}{\sqrt{(1-0{,}2)^2 (1-0{,}82)^2}} = 0{,}44;$$

$$r_{23.1} = \frac{0{,}82^2 - 0{,}2 \cdot 0{,}41}{\sqrt{(1-0{,}41)^2 (1-0{,}2)^2}} = 0{,}83.$$

Оцінимо значущість коефіцієнта $r_{12.3}$. Значення статистики t-критерію при $n' = n - p + 2 = 100 - 3 + 2 = 99$ (за абсолютною величиною)

$$|t| = \frac{|-0{,}26|\sqrt{99-2}}{\sqrt{1-(-0{,}26)^2}} = 2{,}65$$

є більшим за табличне $t_{0,95;97}$ = 1,99 (додаток Г) отже, частковий коефіцієнт кореляції є значущим. Аналогічно встановлюємо значущість інших часткових коефіцієнтів кореляції.

Порівнюючи часткові коефіцієнти кореляції з відповідними парними коефіцієнтами, помічаємо, що найбільші зміни відбулися з коефіцієнтом кореляції між продуктивністю праці (X_1) та віком робітників (X_2) (змінилось не лише його значення, але й знак: r_{12} = 0,2; $r_{12.3}$ = -0,26, причому обидва коефіцієнти є значущими).

Таким чином, між продуктивністю праці (X_1) та віком робітників (X_2) існує прямий кореляційний зв’язок (r_{12} = 0,2). Якщо ж усунути (елімінувати) вплив змінної “виробничий стаж” (X_3), то в чистому вигляді продуктивність праці (X_1) знаходиться в зворотному за напрямом зв’язку з віком робітників (X_2) ($r_{12.3}$ = -0,26). Цей факт легко пояснити, якщо розглядати вік лише як працеспроможність організму на певному етапі його життєдіяльності. Аналогічно можна пояснити й інші часткові коефіцієнти кореляції.

4.8 Багатофакторні моделі економічного зростання

Багатофакторна модель економічного зростання виражає залежність взаємодії динаміки обсягу виробництва від низки виробничих факторів із урахуванням зміни їх якості, ефективності використання, загальних наслідків науково-технічного прогресу та вдосконалення організації виробництва:

$$y(t) = f(X_t, A_t, t), \tag{4.46}$$

де X_t – вектор фізичних обсягів виробничих ресурсів;
A_t – вектор параметрів, які характеризують якість та ефективність ресурсів.

Функція (4.46) є динамічним варіантом виробничої функції з взаємозамінними ресурсами, але, на відміну від статичної функції, її параметри змінюються у часі й вона може містити змінну часу. Модель (4.46) будують, використовуючи емпіричні динамічні ряди обсягів виробництва і виробничих ресурсів, тобто як багатофакторне рівняння регресії.

За функцією (4.46) визначають середні μ_{it} і граничні v_{it} показники ефективності використання ресурсів:

$$\mu_{it} = \frac{f\left(X_t, A_t, t\right)}{x_{it}}, \tag{4.47}$$

$$v_{it} = \frac{\partial f\left(X_t, A_t, t\right)}{\partial x_{it}}. \tag{4.48}$$

Перевагами показників ефективності (4.47), (4.48) є врахування поєднання всіх основних виробничих факторів. За їх допомогою можна дослідити симбіоз виробничих факторів, що приводять, наприклад, до найбільшого зростання продуктивності праці або найбільшої економії деяких природних ресурсів тощо.

Ознакою підвищення ефективності використання виробничих факторів у динаміці виробництва є нерівності:

$$\frac{d\mu_{it}}{dt} > 0, \qquad \frac{d\nu_{it}}{dt} > 0.$$

Якщо середня ефективність *i*-го ресурсу зростає, то темп приросту виробництва буде вищим за темп приросту відповідного ресурсу.

В основі багатофакторного аналізу значення екстенсивних та інтенсивних факторів економічного зростання лежить розкладання повного приросту функції (4.46) на складові та економічна інтерпретація цього розкладання. Проте формально повний приріст функції *n* змінних як суму *n* доданків, які відповідають приростам аргументів (якщо вони не прямують до нуля), подати не можна. З огляду на це неможливо отримати точні однозначні оцінки внеску кожного фактора у приріст функції. Економічний зміст цього висновку очевидний: ефект взаємодії факторів не є сумою ефектів дії кожного фактора окремо, його досягають лише у сукупній взаємодії усіх факторів (властивість емерджентності будь-якої кібернетичної системи).

У разі використання багатофакторних моделей приріст обсягу виробництва умовно розкладають на складові таким чином:

$$\Delta f(X, A, t) = \Delta f(X) + \Delta f(A) + \Delta f(t), \tag{4.49}$$

де $\Delta f(X) = \sum_{i \in M} \Delta f(x_i)$, $\Delta f(x_i)$ – приріст, отриманий за рахунок збільшення фізичного обсягу i-го ресурсу; $\Delta f(A) = \sum_{j \in N} \Delta f(a_j)$, $\Delta f(a_j)$ – приріст, який отриманий за рахунок підвищення ефективності використання *j*-го ресурсу; $\Delta f(t)$ – приріст, отриманий за рахунок загального економічного прогресу виробництва, який відображає сукупну взаємодію всіх факторів.

Відповідно:

$\lambda_{екс} = \dfrac{\Delta f(X)}{\Delta f(X, A, t)}$ – частка екстенсивних факторів зростання;

$\lambda_{інт} = \dfrac{\Delta f(A)}{\Delta f(X, A, t)}$ – частка інтенсивних факторів зростання;

$\lambda_t = \dfrac{\Delta f(t)}{\Delta f(X, A, t)}$ – частка факторів загального економічного прогресу (часто її приєднують до $\lambda_{інт}$).

При теоретичних дослідженнях багатофакторних моделей часто використовують неперервний (граничний) варіант формули (4.49):

$$dy(t)=\sum_{i\in M}\frac{\partial f(X,A,t)}{\partial x_i}dx_i+\sum_{j\in M}\frac{\partial f(X,A,t)}{\partial a_j}da_j+\frac{\partial f(X,A,t)}{\partial t}dt, \qquad (4.50)$$

тобто повний диференціал функції $y(t)=f(X,A,t)$.

Приклад 2.8. Дослідити вплив екстенсивних та інтенсивних факторів зростання валового випуску підприємства залежно від взаємодії двох виробничих ресурсів: витрат праці та величини виробничих фондів. Вихідні дані, необхідні для проведення аналізу використання ресурсів, наведено в таблиці 4.5.

Таблиця 4.5

Показник	Базовий період	Поточний період
Валовий випуск підприємства	26879	31259
Кількість працівників, осіб	3680	4160
Продуктивність праці, шт./осіб	7,3041	7,5142
Величина основних виробничих фондів, грн	51067	53014
Фондовіддача основних виробничих фондів, грн/грн	0,5263	0,5896

Розв'язок. Припустимо, що обсяг валового випуску підприємства описують степеневою виробничою функцією, яка має такий вигляд:

$$y=x_1^{5/8}a_1^{5/8}x_2^{3/8}a_2^{3/8},$$

де y – обсяг валового випуску продукції;

x_1 – витрати праці (кількість працівників);

a_1 – продуктивність праці;

x_2 – величина основних виробничих фондів;

a_2 – фондовіддача основних виробничих фондів.

За даний період валовий випуск продукції підприємства зріс на 4380 (31259 - 26879) одиниць.

Приріст витрат ресурсу x_1 (витрати праці) дорівнює

$$\Delta x_{1t}=4160-3680=480.$$

Приріст ефективності використання ресурсу a_1 (продуктивності праці) становить

$$\Delta\mu_{1t}=7{,}5142-7{,}3041=0{,}2101.$$

Приріст витрат ресурсу x_2 (величина основних виробничих фондів) становить

$$\Delta x_{2t} = 53014 - 41067 = 1947 .$$

Приріст ефективності використання ресурсу a_2 (фондовіддачі основних виробничих фондів) дорівнює

$$\Delta \mu_{2t} = 0{,}5896 - 0{,}5263 = 0{,}0633 .$$

За методом ланцюгових підстановок приріст валового випуску підприємства становить:

– за рахунок приросту фонду робочого часу (екстенсивна складова x_1)

$$\Delta f(x_{1t}) = x_{1зв}^{\frac{5}{8}} a_{1баз}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} - x_{1баз}^{\frac{5}{8}} a_{1баз}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} = 4160^{\frac{5}{8}} \cdot 7{,}3041^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot$$
$$\cdot 0{,}5263^{\frac{3}{8}} - 3680^{\frac{5}{8}} \cdot 7{,}3041^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 2140{,}6681;$$

– за рахунок приросту продуктивності праці (інтенсивна складова a_1)

$$\Delta f(a_{1t}) = x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} - x_{1зв}^{\frac{5}{8}} a_{1баз}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} =$$
$$4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot \cdot 0{,}5263^{\frac{3}{8}} - 4160^{\frac{5}{8}} \cdot 7{,}3041^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 518{,}9359;$$

– за рахунок приросту величини основних виробничих фондів (екстенсивна складова x_2)

$$\Delta f(x_{2t}) = x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2зв}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} - x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2баз}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} = 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 53014^{\frac{3}{8}} \cdot$$
$$\cdot 0{,}5263^{\frac{3}{8}} - 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 51067^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 417{,}3947;$$

– за рахунок приросту фондовіддачі основних виробничих фондів (інтенсивна складова a_2)

$$\Delta f(a_{2t}) = x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2зв}^{\frac{3}{8}} a_{2зв}^{\frac{3}{8}} - x_{1зв}^{\frac{5}{8}} a_{1зв}^{\frac{5}{8}} x_{2зв}^{\frac{3}{8}} a_{2баз}^{\frac{3}{8}} = 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 53014^{\frac{3}{8}} \cdot 0{,}5896^{\frac{3}{8}} -$$
$$- 4160^{\frac{5}{8}} \cdot 7{,}5142^{\frac{5}{8}} \cdot 53014^{\frac{3}{8}} \cdot 0{,}5263^{\frac{3}{8}} = 1303{,}0014.$$

Перевіримо формулу розкладання загального приросту на складові:

2140,6681+518,9359+417,3947+1303,0014=4380.

Частки екстенсивних та інтенсивних факторів зростання валового випуску продукції наведено в таблиці 4.6.

Результати дослідження економічного зростання валового випуску продукції показують, що відносно виробничого фактора витрат праці на підприємстві переважає екстенсивний тип розвитку, а відносно виробничого фактора витрат основних виробничих фондів переважає інтенсивний тип розвитку.

Таблиця 4.6

Приріст валового випуску продукції за рахунок	Частка загального приросту
фонду робочого часу	0,4887 (48,87%)
продуктивності праці	0,1185 (11,85%)
величини основних виробничих фондів	0,0953 (9,53%)
фондовіддачі основних виробничих фондів	0,2975 (29,75%)

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Запишіть класичну нормальну лінійну модель множинної регресії у звичайній та матричній формах.
2. Охарактеризуйте процес оцінювання параметрів класичної регресійної моделі методом найменших квадратів.
3. Що характеризує коефіцієнт еластичності та як його можна обчислити?
4. Що називають коваріаційною матрицею? Запишіть її. Що пояснює коефіцієнт коваріації?
5. Охарактеризуйте коваріаційну матрицю вектора збурень.
6. Що таке коваріаційна матриця?
7. Сформулюйте та доведіть теорему Гаусса-Маркова.
8. Поясніть, що є дисперсією збурень.
9. Охарактеризуйте перевірку двофакторної регресії на адекватність за допомогою коефіцієнта детермінації.
10. Охарактеризуйте критерій Фішера.
11. Що називають скоригованим коефіцієнтом детермінації і як його обчислюють?
12. Вкажіть недолік коефіцієнта детермінації.
13. Як оцінити значущість коефіцієнта регресії b_j?
14. У чому полягає метод лінеаризації нелінійних множинних моделей? Вкажіть недоліки цього методу.

15. Знайдіть довірчий інтервал для параметра b_j лінійної множинної регресії.
16. Охарактеризуйте виробничу функцію Кобба-Дугласа.
17. Що називають коефіцієнтом часткової еластичності?
18. Запишіть функцію Кобба-Дугласа з урахуванням технічного прогресу.
19. За допомогою якого показника визначають ступінь однорідності функції Кобба-Дугласа.
20. Запишіть динамічну функцію Кобба-Дугласа, нормувавши коефіцієнт еластичності.
21. За якою формулою можна оцінити темп приросту національного доходу?
22. Що називають частковим коефіцієнтом кореляції? Поясніть формулу для обчислення цього коефіцієнта.
23. Що виражає багатофакторна модель економічного зростання?
24. Запишіть середні та граничні показники ефективності використання ресурсів та вкажіть їх переваги. Що можна дослідити за їх допомогою?
25. Вкажіть ознаку ефективності використання виробничих факторів у динаміці виробництва.
26. Як можна розкласти на складові приріст обсягу виробництва?
27. Запишіть формулу для пошуку частки екстенсивних факторів зростання.
28. Опишіть залежність для пошуку частки інтенсивних факторів зростання.
29. Сформулюйте модель для оцінювання частки факторів загального економічного прогресу.
30. Запишіть неперервний варіант формули розкладання на складові приросту обсягу виробництва.

ВПРАВИ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Вправа 4.1. Є такі дані: виробіток литва на одного працівника X_1 (т), браковане литво X_2 (%), собівартість 1 т литва Y (грн) для 25 цехів заводів (табл. 4.7):

Необхідно: 1) знайти рівняння множинної регресії Y за X_1 та X_2, оцінити значущість цього рівняння та його коефіцієнтів на рівні $\alpha = 0{,}05$; 2) знайти множинний коефіцієнт детермінації та пояснити його зміст; 3) порівняти відокремлений вплив на залежну змінну кожної з пояснювальних змінних, використовуючи коефіцієнт еластичності; 4) знайти 95%-ві довірчі інтервали для коефіцієнтів регресії.

Таблиця 4.7 – Умовні дані для завдання 4.1

i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i
1	14,6	4,2	239	10	25,3	0,9	198	19	17,0	9,3	282
2	135,	6,7	254	11	56,0	1,3	170	20	33,1	3,3	196
3	21,5	5,5	262	12	40,2	1,8	173	21	30,1	3,5	186
4	17,4	7,7	251	13	40,6	3,3	197	22	65,2	1,0	176
5	44,8	1,2	158	14	75,8	3,4	172	23	22,6	5,2	238
6	111,9	2,2	101	15	27,6	1,1	201	24	33,4	2,3	204
7	20,1	8,4	259	16	88,4	0,1	130	25	19,7	2,7	205
8	28,1	1,4	186	17	16,6	4,1	251				
9	22,3	4,2	204	18	33,4	2,3	195				

Вправа 4.2. Є такі дані про річні ставки місячних доходів за трьома акціями за шість місяців:

Акція	Доходи по місяцях, %					
A	5,4	5,3	4,9	4,9	5,4	6,0
B	6,3	6,2	6,1	5,8	5,7	5,7
C	9,2	9,2	9,1	9,0	8,7	8,6

Є підстави припускати, що доходи Y акції C залежать від доходів X_1 та X_2 акцій A та B. Необхідно:

1) знайти множинний коефіцієнт детермінації та пояснити його зміст;

2) скласти рівняння регресії Y за X_1 та X_2;

3) перевірити значущість рівняння регресії на рівні $\alpha = 0,2$;

4) оцінити середній дохід акції C, якщо доходи акцій A та B склали відповідно 5,5% та 6%.

Вправа 4.3. Із метою дослідження впливу факторів X_1 − середньомісячної кількості профілактичного налаштування автоматичної лінії та X_2 – середньомісячного числа обривів нитки на показник Y – середньомісячну характеристику якості тканини (в балах) за даними 37 підприємств легкої промисловості було обчислено парні коефіцієнти кореляції: $r_{y1} = 0,105$; $r_{y2} = 0,024$ та $r_{12} = 0,996$. Визначити часткові коефіцієнти кореляції $r_{y1.2}$ та $r_{y2.1}$ на оцінити їх значущість на 5%-му рівні.

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Завдання 4.1. Побудувати лінійну економетричну модель, що характеризує залежність між витратами обігу, обсягом вантажообігу та фондомісткістю бази. Визначити коефіцієнти еластичності та детермінації. Вихідні дані наведені в таблицях відповідно до варіанта.

Варіант 1

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,72	15,6	106,3
2	3,04	13,5	128,5
3	2,84	15,3	118,0
4	2,89	14,9	121,2
5	2,58	15,1	120.0
6	2,64	16,1	118,4
7	2,52	16,7	108,4
8	2,75	15,4	110,0
9	2,63	17,1	105,9

Варіант 2

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,58	15,1	120,0
2	2,64	16,1	118,4
3	2,52	16,7	108,4
4	2,75	15,4	110,0
5	2,63	17,1	105,9
6	2,48	16,8	117,7
7	2,62	16,9	97,5
8	2,88	16.1	113,7
9	2,68	15,0	122,3

Варіант 3

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,48	16.8	117,7
2	2,62	16,9	97,5
3	2,88	16,1	113,7
4	2,68	15,0	122,3
5	2,52	18,0	102,0
6	2,74	17,2	106,7
7	2,56	17.1	108,5
8	2,68	16,4	114,3
9	2,55	16.7	94,3

Варіант 4

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,75	16,8	110,0
2	2,63	15,5	105,9
3	2,48	17,0	117,7
4	2,62	16,8	97,5
5	2,88	16,9	113,7
6	2,68	16,1	122,3
7	2,56	15,0	102,0
8	2,74	18,0	106,7
9	2,60	17,2	108,5

Варіант 5

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,92	14,1	87,8
2	2,64	17,2	72,0
3	2,79	17,1	72,4
4	2,67	17,8	69,5
5	2,68	16,2	75,0
6	2,85	17,2	70-6
7	2,40	16,8	73,4
8	2.91	14,8	80,7
9	2,29	19,6	62.2

Варіант 6

№	***Витрати***	Вантажо-обіг	Фондо-міскість
1	2,45	17,1	71,3
2	2,38	19,5	61,7
3	3,04	12,5	96,2
4	2,67	16,5	72,9
5	2,70	16,0	75,0
6	2,65	16,1	74,6
7	2,79	16,2	74,1
8	2,49	18,0	66,9
9	3,27	11,4	98,6

Варіант 7

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,85	17,2	70,6
2	2,40	16,8	73,4
3|	2,91	14,8	80,7
4	2,29	19,6	62,2
5	3,27	11,4	98,6
6	2,45	17,1	71,3
7	2,38	19,5	61,7
8	3,04	12,5	96,2
9	2,67	16.5	72,9

Варіант 8

№	***Витрати***	Вантжо-обіг	Фондо-міскість
1	2,93	18,1	71,2
2	2,50	17,2	73,4
3	2,95	14,9	81,2
4	2,39	20,1	63.7
5	3,25	11,4	96,6
6	2,65	17,1	72,2
7	2,42	19,5	61,7
8	3,14	17,5	96,2
9	2,75	16,7	72,9

Варіант 9

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,14	17,4	80,3
2	2,94	13,8	102,5
3	2,67	15,0	94.3
4	2,44	18,6	76,0
5	2,83	16,2	87.3
6	2,92	15,7	90,1
7	2,61	17.9	82.8
8	2,72	15,3	96,9
9	2,68	16,3	83,7

Варіант 10

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,67	15,0	94,0
2	2,45	18,6	78,0
3	2,86	16.2	87.5
4	2.90	15,7	90,2
5	2,60	17.9	84,8
6	2,72	16,3	95,9
7	2,68	17,7	91,0
8	2,50	16,8	84,7
9	2,74	17,5	88,2

Варіант 11

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,83	13,8	68.0
2	2,75	14,8	64,3
3	2,40	16,9	55,1
4	2,30	16,8	55,5
5	2,47	14,8	63.3
6	2,45	17,9	52,7
7	2,48	17,6	53,7
8	2,41	15,7	60,2
9	2,34	15,2	62.2

Варіант 12

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,69	14,9	69,4
2	2,48	16,1	58,7
3	2,11	19,7	62,3
4	2,82	14,0	83,8
5	2,43	17,1	68,5
6	2,34	18,2	64,5
7	2,48	17,4	67,6
8	2,69	16,1	72,9
9	2,36	18,8	62,4

Варіант 13

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,35	16,8	56.5
2	2,48	15,0	64,3
3	2,45	17,5	53,7
4	2,47	17,6	54,7
5	2,42	15,7	60,2
6	2,34	15.2	62,4
7	2,70	14.9	69.5
8	2,48	16,1	58.7
9	2,15	19,7	62.3

Варіант 14

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,46	19,0	65,4
2	2,70	16,3	73,9
3	2,58	17,5	68,5
4	2,34	18,4	64,5
5	2,43	17.2	69,3
6	2,84	15,0	83,8
7	2.15	19,8	62,3
8	2,49	16.1	58,7
9	2,60	14,9	69,4

Варіант 15

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,27	32,1	112,5
2	1,94	31,0	116,4
3	2,32	32,4	111,6
4	2,49	33,2	108,9
5	2,57	31,2	116,5
6	2,01	34,8	104,5
7	1,87	35,4	102,7
8	2,39	33,0	110,2
9	2,18	34,8	104,7

Варіант 16

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,17	33,3	109,4
2	1,80	36,1	101,1
3	2,36	38,3	102,6
4	2,50	30,6	128,5
5	2.27	32,1	122,5
6	2,33	37,6	105,2
7	2,51	34,8	114,8
8	2,40	34,2	116,0
9	2,50	34,2	116,0

Варіант 17

№	***Витрати***	Вантажо-обіг	Фондо-місткіст
1	2,95	29,4	152,0
2	2,55	35,4	126,2
3	2,26	39,7	112,6
4	2,49	37,1	120,5
5	2,17	35,7	125,3
6	2,38	40,2	111,3
7	2,22	39,4	112,2
8	2,64	43,7	121,2
9	2,63	38,4	126.4

Варіант 18

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,32	38,8	114,0
2	2,19	39,9	103,1
3	2,83	30,1	153,8
4	2,75	31,7	146,0
5	2,59	37,2	124,8
6	2,27	39,7	103,6
7	2,05	36,9	119,0
8	1,95	38,2	108,7
9	2,08	40,1	106,5

Варіант 19

№	***Витрат и***	Вантажо-обіг	Фондо-місткість
1	3,95	29,4	152,0
2	3,55	35,4	126,2
3	3,26	39,7	112,6
4	3,49	37,1	120,5
5	3,17	35,7	125,3
6	3,38	40,2	111,3
7	3,22	39,4	112,2
8	3,64	43,7	121,2
9	3,63	38,4	126.4

Варіант 20

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,32	39,8	114,0
2	2,19	40,9	103,1
3	2,83	30,1	153,8
4	2,75	32,7	146,0
5	2,59	33,2	124,8
6	2,27	39,7	103,6
7	2,05	36,3	119,0
8	1,95	28,2	108,7
9	2,08	40,1	106,5

Варіант 21

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,95	29,41	153,0
2	2,55	35,42	127,2
3	2,26	39,73	113,6
4	2,49	37,14	121,5
5	2,17	35,75	126,3
6	2,38	40,24	112,3
7	2,22	39,43	113,2
8	2,64	43,72	122,2
9	2,63	38,41	127.4

Варіант 22

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,32	38,81	115,0
2	2,19	39,92	104,1
3	2,83	30,13	154,8
4	2,75	31,74	147,0
5	2,59	37,25	125,8
6	2,27	39,74	104,6
7	2,05	36,93	119,0
8	1,95	38,22	109,7
9	2,08	40,11	107,5

Варіант 23

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,08	32,5	81,6
2	1,99	33,4	79.4
3	1,96	37,8	69,5
4	2,18	35,8	85.4
5	1,91	34,2	84,3
6	2,37	37,2	71,4
7	1,92	38,2	78.1
8	2,15	29,4	90,8
9	2,41	37.2	72,1

Варіант 24

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	2,20	34,5	83,6
2	1,96	35,0	76,5
3	2,25	43,7	76,9
4	2,69	31.9	104,6
5	2,24	37.3	72,3
6	2,43	40,9	66,3
7	2,32	38,8	69,6
8	2,60	35,7	75,6
9	2,70	43,2	62,4

Варіант 25

№	***Витрати***	Вантажо-обіг	Фондо-місткість
1	4,95	32,1	69,4
2	4,55	31,0	58,7
3	4,26	32,4	62,3
4	4,49	33,2	83,8
5	4,17	31,2	68,5
6	4,38	34,8	64,5
7	4,22	35,4	67,6
8	4,64	33,0	72,9
9	4,63	34,8	62,4

ТЕСТИ

1. Модель множинної регресії має вигляд $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i$. Назвіть параметри і випадкову величину цієї моделі:

а) ε_i – випадкова величина, β_i – параметри моделі;

б) ε_i – параметр моделі, β_i – випадкові величини;

в) y_i, x_i – параметри моделі, ε_i – випадкова величина;

г) y_i, x_i – параметри моделі, β_i – випадкові величини.

2. У моделі виду $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i$ залежною та незалежною змінною є:

а) залежна змінна – X_i, незалежна змінна – Y;

б) залежна змінна – Y, незалежна змінна – X_i;

в) залежна змінна – Y, незалежна змінна відсутня;

г) залежна змінна – ε_i, незалежна змінна – X_i;

3. Залежну змінну Y регресійної моделі називають:

а) екзогенною;

б) пояснюваною;

в) ендогенною;

г) усі варіанти правильні.

4. Незалежні змінні X_i часто називають:

а) пояснювальними;

б) неспостережуваними;

в) екзогенними;

г) ендогенними.

5. За методом найменших квадратів параметр b обирається таким, щоб:

а) добуток квадратів відхилень був мінімальним;

б) сума квадратів відхилень була максимальною, тобто $e^T e = \sum_{i=1}^{n} e_i^2 \to \max$;

в) сума квадратів відхилень була мінімальною, тобто $e^T e = \sum_{i=1}^{n} e_i^2 \to \min$;

г) добуток квадратів відхилень був максимальним.

6. Параметр b регресійної моделі знаходиться за формулою:

а) $b = (X^T X)^{-1} \cdot X^T Y$;

б) $b = (X^T X)^{-1} \cdot X \;\; Y^T$;

в) $b = (X^{-1} X)^T \cdot X^T Y$;

г) немає правильного варіанта.

7. Система $X^T X b = X^T Y$ має розв'язок, якщо:

а) матриця $X^T X$ є невласною;

б) визначник матриці $X^T X$ не дорівнює нулю;

в) визначник матриці $X^T X$ дорівнює нулю;

г) визначник матриці $X^T X$ не дорівнює одиниці.

8. Рівняння $\overline{Y} = 2{,}3 + 0{,}4x_1 + 0{,}7x_2$ описує залежність між ціною одиниці продукції Y (грн) та витратами на робочу силу X_1 (грн) і на сировину X_2 (грн). Що показує дане рівняння?

а) при збільшенні витрат на робочу силу X_1 (за незмінного значення X_2) на 1 грн. ціна одиниці продукції Y зростає на 0,4 грн, а при збільшенні на 1 грн витрат на сировину X_2 (за незмінного значення X_1) ціна зростає на 0,7 грн;

б) при збільшенні витрат на робочу силу X_1 та сировину X_2 на 1%, ціна одиниці продукції Y зростає на (0,4+0,7)%;

в) при збільшенні витрат на робочу силу X_1 та сировину X_2 на 1%, ціна одиниці продукції Y спадає на (0,4+0,7)%;

г) при сталому рівні витрат на робочу силу X_1 та сировину X_2, ціна одиниці продукції Y спадає на 2,3.

9. Для порівняння відокремленого впливу на залежну змінну різних пояснювальних змінних, коли останні виражають різними одиницями виміру, використовують:

а) коефіцієнт детермінації;

б) коефіцієнт кореляції;

в) критерій Фішера;

г) коефіцієнти еластичності.

10. Коефіцієнти еластичності розраховують за формулою:

а) $E_j = \frac{\overline{x_j}}{\overline{y}}$;

б) $E_j = b_j \frac{\overline{x_j}}{\overline{y}}$;

в) $E_j = -(b_j)\frac{\overline{x_j}}{\overline{y}}$;

г) $E_j = b_j$.

11. Коефіцієнт еластичності E_j показує:

а) на скільки відсотків зміниться в середньому Y, якщо x_j збільшити на 1%;

б) у скільки разів зміниться в середньому Y, якщо x_j збільшити на 1%;

в) на скільки відсотків зміниться в середньому X, якщо Y збільшити на 1%;

г) жодної правильної відповіді.

12. Коваріація характеризує:

а) ступінь розсіювання значень двох змінних відносно їх математичних сподівань;

б) взаємозв'язок змінних моделі;

в) відокремлений вплив на залежну змінну різних пояснювальних змінних;

г) усі варіанти правильні.

13. При проведенні кореляційно-регресійного аналізу за допомогою оберненої матриці $(X^{\mathrm{T}}X)^{-1}$ визначають:

а) лише вектор b;

б) дисперсії та коваріації компонент вектора b;

в) не лише сам вектор b, але й дисперсії та коваріації його компонент;

г) вектор b та коваріації його компонент.

14. Несуміщена оцінка s^2 параметра σ^2 або вибіркова залишкова дисперсія s^2 визначається за формулою:

а) $s^2 = \frac{e^T e}{n - p - 1}$;

б) $s^2 = \frac{e^T e - 1}{n - p}$;

в) $s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1}$;

г) усі варіанти правильні.

15. Рівняння множинної регресії є значущим (іншими словами – гіпотеза H_0 про рівність нулю параметрів регресійної моделі відкидається), якщо:

а) $F = \frac{Q_R(n-p-1)}{Q_e p} > F_{\alpha;p;n-p-1}$, де $F_{\alpha;p;n\text{-}p\text{-}1}$ – табличне значення F - критерію;

б) $F = \frac{Q_R(n-p-1)}{Q_e p} < F_{\alpha;p;n-p-1}$, де $F_{\alpha;p;n\text{-}p\text{-}1}$ – табличне значення F -критерію;

в) $F = \frac{Q_R(n-p)}{Q_e p} > F_{\alpha;p;n-p-1}$, де $F_{\alpha;p;n\text{-}p\text{-}1}$ – табличне значення F -критерію;

г) немає правильного варіанта.

16. Коефіцієнт детермінації D:

а) характеризує частку варіації залежної змінної, зумовленої регресією;

б) характеризує частку варіації залежної змінної, не зумовленої моделлю;

в) чим ближчий до одиниці, тим краще регресія описує залежність між пояснювальними та залежною змінними;

г) чим ближчий до одиниці, тим гірше регресія описує залежність між пояснювальними та залежною змінними;

17. Скоригований (адаптований, виправлений) коефіцієнт детермінації $\overline{D}$ визначається за формулою:

а) $\overline{D} = 1 - \frac{n}{n-p}(1-D)$;

б) $\overline{D} = 1 - \frac{n-1}{n-p-1}(1-D)$;

в) $\overline{D} = 1 - \frac{(n)e^T e}{(n-p)y^T y}$;

г) $\overline{D} = 1 - \frac{(n-1)e^T e}{(n-p-1)y^T y}$.

18. Виробнича функція Кобба-Дугласа має вигляд:

а) $Y = AK^{\alpha}L^{\beta}$, де Y – обсяг виробництва, K – витрати капіталу, L – витрати праці;

б) $K = AY^{\alpha}L^{\beta}$, де K – витрати капіталу, Y – обсяг виробництва, L – витрати праці;

в) $L = A^{\alpha}KY^{\beta}$, де Y – обсяг виробництва, K – витрати капіталу, L – витрати праці;

г) $Y = A^{\alpha}K^{\beta}L^{\gamma}$, де Y – витрати праці, K – витрати капіталу, L – обсяг виробництва.

19. Вибірковим частковим коефіцієнтом кореляції між змінними X_i та X_j при фіксованих значеннях решти (p-2) змінних називають вираз:

а) $r_{ij.1,2,...,p} = \frac{-q_{ij}}{\sqrt{q_{ii}q_{jj}}}$, де q_{ii} та q_{jj} – алгебраїчні доповнення елементів r_{ii} та r_{jj} матриці вибіркових коефіцієнтів кореляції;

б) $r_{ij.1,2,...,p} = \frac{q_{ij}}{\sqrt{q_{ii}q_{jj}}}$, де q_{ii} та q_{jj} – алгебраїчні доповнення елементів r_{ii} та r_{jj} матриці вибіркових коефіцієнтів кореляції;

в) $r_{ij.1,2,...,p} = \frac{-q_{ij}}{\sqrt{q_{ij}}}$, де q_{ii} та q_{ij} – алгебраїчні доповнення елементів r_{ii} та r_{jj} матриці вибіркових коефіцієнтів кореляції;

г) немає правильного варіанта.

20. Багатофакторна модель економічного зростання виражає:

а) залежність динаміки прибутковості та рентабельності господарської діяльності підприємства від обсягу виробництва у динаміці;

б) залежність взаємодії динаміки обсягу виробництва від низки виробничих факторів із урахуванням зміни їх якості, ефективності використання, загальних наслідків науково-технічного прогресу та вдосконалення організації виробництва;

в) залежність взаємодії динаміки обсягу виробництва від низки виробничих факторів без урахуванням зміни їх якості та ефективності використання, проте враховуючи загальні наслідки науково-технічного прогресу та вдосконалення організації виробництва;

г) усі варіанти правильні.

21. Спільний вплив кількох факторів на одну результативну змінну досліджують за допомогою:

а) порогової логіки;

б) теорії ігор;

в) багатофакторних економетричних моделей;

г) усі варіанти правильні.

22. Модель вигляду $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i$ називають:

а) класичною нормальною лінійною моделлю множинної регресії;

б) специфічною лінійною моделлю;

в) класичною лінійною моделлю однофакторної регресії.

г) класичною нормальною нелінійною моделлю множинної регресії.

23. Використовуючи матричні позначення, розпишіть модель вигляду $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i$:

а) $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$;

б) $Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$,

$\beta = \begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_p \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{pmatrix}$;

в) $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$;

г) немає правильної відповіді.

24. При описі регресійної моделі матричним способом до матриці X введено додатковий стовпець, усі елементи якого дорівнюють 1, оскільки:

а) висувається гіпотеза про наявність додаткової змінної, вплив якої не враховується у моделі;

б) умовно припускається, що в регресійній моделі вільний член β_0 множиться на фіктивну змінну x_{i0}, яка дорівнює 1 для всіх значень i;

в) сума усіх параметрів регресійної моделі дорівнює 1;

г) умовно припускається, що в регресійній моделі кожне значення ε множиться на фіктивну змінну x_{i0}, яка дорівнює 1 для всіх значень I.

25. За методом найменших квадратів:

а) параметр b обирається таким, щоб сума квадратів відхилень була мінімальною;

б) невідомий вектор X обирається таким, щоб сума квадратів їх відхилень від свого середнього значення була мінімальною;

в) параметр b обирається таким, щоб сума квадратів відхилень була максимальною;

г) невідомі X обираються такими, щоб сума квадратів їх відхилень від свого середнього значення була максимальною.

26. Що означає даний вираз при проведенні регресійного аналізу: $Q_e = \sum_{i=1}^{n} (\overline{y_x} - y_i)^2 \to \min$:

а) значення змінної y_i мають бути такими, щоб їх сума була мінімальною;

б) параметри моделі шукаються такими, щоб сума квадратів відхилень була мінімальною;

в) сума квадратів різниці середнього значення змінної від усієї сукупності значень має бути мінімальною;

г) усі варіанти правильні.

27. Необхідною умовою екстремуму функції $Q(b_0, b_1, \ldots, b_p)$ є :

а) $\frac{\partial Q}{\partial b} = \begin{pmatrix} \frac{\partial Q}{\partial b_0} & \frac{\partial Q}{\partial b_1} & \cdots & \frac{\partial Q}{\partial b_p} \end{pmatrix}$;

б) $\frac{\partial b}{\partial Q} = \begin{pmatrix} \frac{\partial b_0}{\partial Q} & \frac{\partial b_1}{\partial Q} & \cdots & \frac{\partial b_p}{\partial Q} \end{pmatrix}$;

в) $\frac{\partial Q}{\partial b} = \begin{pmatrix} \frac{\partial Q}{\partial b_1} & \frac{\partial Q}{\partial b_2} & \cdots & \frac{\partial Q}{\partial b_p} \end{pmatrix}$;

г) жодної правильної відповіді.

28. Передумовою множинного регресійного аналізу є таке:

а) вектори значень пояснювальних змінних або стовпці матриці X повинні бути лінійно залежними;

б) вектори значень пояснювальних змінних або стовпці матриці X повинні бути лінійно незалежними;

в) вектори значень пояснювальних змінних або рядки матриці X повинні бути лінійно незалежними;

г) вектори значень пояснювальних змінних або рядки матриці X повинні бути лінійно залежними.

29. Оцінивши параметри класичної лінійної кореляційно регресійної моделі, було отримано матрицю параметрів $b = \begin{pmatrix} -3{,}5393 \\ 0{,}8539 \\ 0{,}3670 \end{pmatrix}$. Який вигляд матиме модель?

а) $\overline{Y} = -3{,}5393x_1 + 0{,}854x_2 + 0{,}367x_3$;

б) $\overline{Y} = 3{,}5393x_1 + 0{,}854x_2 + 0{,}367x_3$;

в) $\overline{Y} = 3{,}5393 - 0{,}854\,x_1 - 0{,}367\,x_2$;

г) $\overline{Y} = -3{,}5393 + 0{,}854\,x_1 + 0{,}367\,x_2$.

30. Відома лінійна економетрична модель, що характеризує залежність між витратами обігу (Y), обсягом вантажообігу (X_1) та фондомісткістю бази (X_2). Що характеризує параметр b_0 у такій моделі:

а) максимальні витрати обігу;

б) граничні витрати обігу;

в) граничний вантажообіг та фондомісткість разом;

г) жодної правильної відповіді.

31. Елементи σ_{ij} матриці $\sum_b = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0p} \\ \sigma_{10} & \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p0} & \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$ у регресійному аналізі називаються:

а) варіації параметрів β_i та β_j;

б) кореляційні моменти;

в) дисперсійні моменти;

г) жодної правильної відповіді.

32. Коваріація двох змінних визначається як:

а) математичне сподівання суми відхилення цих змінних від їх математичних сподівань;

б) математичне сподівання добутку відхилення математичних сподівань змінних від їх початкових значень;

в) математичне сподівання добутку відхилення цих змінних від їх математичних сподівань;

г) математичне сподівання суми відхилення математичних сподівань змінних від їх початкових значень.

33. У двофакторній регресії сума квадратів відхилень, зумовлених регресією Q_R, розраховується за залежністю:

а) $Q_R = b^T X^T Y - n\overline{y}^2$;

б) $Q_R = Y^T Y - b^T X^T Y$;

в) $Q_R = b^T X^T Y + n\overline{y}^2$

г) $Q_R = b^T X^T Y$.

34. У двофакторній регресії залишкова сума квадратів, що характеризує вплив не врахованих в моделі факторів Q_e, визначається за формулою:

а) $Q_e = Y^T Y + b^T X^T Y$;

б) $Q_e = Y^T Y - b^T X^T Y$;

в) $Q_e = b^T X^T Y - n\bar{y}^2$;

г) $Q_e = Y^T Y - b^T X^T$.

35. Коефіцієнт детермінації D як одна з найбільш ефективних оцінок адекватності регресійної моделі розраховується за залежністю:

а) $D = \dfrac{b^T X^T Y + n\bar{y}^2}{Y^T Y + n\bar{y}^2}$;

б) $D = \dfrac{b^T X^T Y - n\bar{y}^2}{Y^T Y - n\bar{y}^2}$;

в) $D = \dfrac{bX^T Y - n\bar{y}^2}{Y^T Y + n\bar{y}^2}$;

г) $D = \dfrac{b^T X^T Y}{Y^T Y}$.

ТЕМА 5 АВТОКОРЕЛЯЦІЯ, ГЕТЕРОСКЕДАСТИЧНІСТЬ ТА МУЛЬТИКОЛІНЕАРНІСТЬ У КОРЕЛЯЦІЙНОМУ АНАЛІЗІ

5.1 Сутність, причини та наслідки автокореляції

При побудові кореляційно-регресійних моделей обов'язковою умовою є незалежність будь-якого ε_i значення випадкової величини від значення ε_j: $cov(\varepsilon_i,\varepsilon_j) = 0$, оскільки це свідчить про відсутність кореляції між будь-якими значеннями випадкової величини.

Отже, **автокореляція** – це явище кореляції значень результуючої змінної, що виникає унаслідок залежності значень випадкової величини ε_i, $i = \overline{1,n}$, у різних спостереженнях. Це називають терміном **автокореляція випадкових відхилень** або просто **автокореляція**.

Сутність автокореляції можна пояснити за допомогою такого прикладу. Нехай досліджується взаємозв'язок між попитом y на шуби та доходом x, що зображено на рис. 5.1.

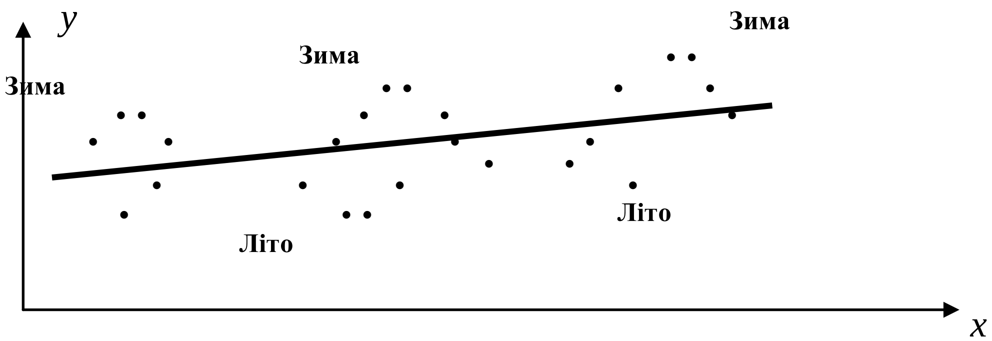


Рисунок 5.1 – Залежність між попитом y на шуби та доходом x

Як видно з рисунка 5.1, така трендова залежність може бути представлена лінійно. При цьому зазначимо, що фактичні емпіричні точки зазвичай перевищують трендову лінію взимку і є нижчими влітку. Така ситуація спостерігається у макроекономіці під впливом циклів ділової активності.

Оскільки виникнення автокореляції спостерігається частіше під час використання динамічних рядів, то для нумерації спостережень використовуватимемо символ, що позначає момент часу спостереження, а обсяг вибірки позначатимемо символом T $(t = \overline{1,T})$.

Розрізняють поняття *автокореляції* та *серійної кореляції*.

Автокореляція виникає у разі наявності залежності між значеннями випадкових відхилень в межах однієї вибірки. Автокореляція між двома послідовними значеннями випадкової величини (із запізненням в один лаг) $\mathrm{cov}(\varepsilon_{t-1}, \varepsilon_t) \neq 0$ є ***автокореляцією першого порядку***. Якщо є залежність між значеннями випадкових відхилень однієї вибірки із запізненням у два лаги $\text{cov}(\varepsilon_{t-1}, \varepsilon_t) \neq 0$, $\text{cov}(\varepsilon_{t-2}, \varepsilon_t) \neq 0$, то це ***автокореляція другого порядку***.

Серійною кореляцією називають залежність між значеннями випадкових відхилень двох різних вибірок.

Основними ***причинами, що уможливлюють появу автокореляції***, є такі:

1. *Помилки специфікації.* Неврахування в кореляційно-регресійній моделі будь-якої важливої факторної ознаки або неправильний вибір форми залежності здебільшого зумовлює систематичні відхилення точок спостережень від лінії регресії, суттєву перевагу послідовних відхилень однакового знаку над сусідніми відхиленнями протилежних знаків, що і є типовим при позитивній автокореляції;

2. *Інерція.* Це явище виникає при аналізі тих економічних характеристик, зміна яких є інертною (наприклад, інфляція, безробіття, ВВП), що пояснюється циклічним розвитком ділової активності в країні. Дійсно, економічний підйом спричиняє зростання зайнятості, скорочення інфляції, збільшення ВВП, що триває, доки зміна кон'юнктури ринку і низки економічних показників не вплинуть на уповільнення зростання, його припинення, а після цього і спадання цих показників;

3. *Ефект павутиння.* У різних сферах економіки показники реагують на зміну економічних умов із певним запізненням (часовим лагом). Наприклад, пропозиція цукрової продукції реагує на зміну ціни із запізненням, що дорівнює періоду визрівання цукрового буряку. Як правило, висока ціна сільськогосподарської продукції у попередньому періоді спричинює її перевиробництво у наступному, а отже, ціна на неї спадає;

4. *Усереднення даних.* Значення показників за тривалі часові періоди здебільшого аналізують усередненням їх підінтервалів. Це призводить до певного згладжування коливань, які були всередині даного інтервалу, що і може спричинити автокореляцію.

Серед ***основних наслідків*** автокореляції виділяють такі:

1. Оцінки параметрів, залишаючись лінійними та незміщеними, перестають бути ефективними, у них зникають властивості найкращих лінійних оцінок;

2. Дисперсії оцінок параметрів є зміщеними, що призводить до збільшення t-статистик, до визнання статистично значущими тих факторних ознак, які насправді такими не є;

3. Оцінка непоясненої дисперсії за кореляційно-регресійною моделлю з k факторами $S^2 = \frac{\sum_{t=1}^{T} \varepsilon_t^2}{T - k - 1}$ є зміщеною оцінкою істинного значення σ^2 (переважно знижує істинне значення);

4. Значення t- і F-статистик коефіцієнтів регресії і детермінації можуть бути неправильними, що погіршує прогнозні якості побудованих регресійних моделей.

5.2 Тестування автокореляції. Графічний метод

Графічний метод передбачає побудову послідовно-часових графіків, які зображають відхилення e_t в різні моменти часу t. Тоді на осі абсцис відкладають моменти часу отримання статистичних даних (або порядкові номери спостережень), а на осі ординат – відхилення e_t (рис. 5.2).

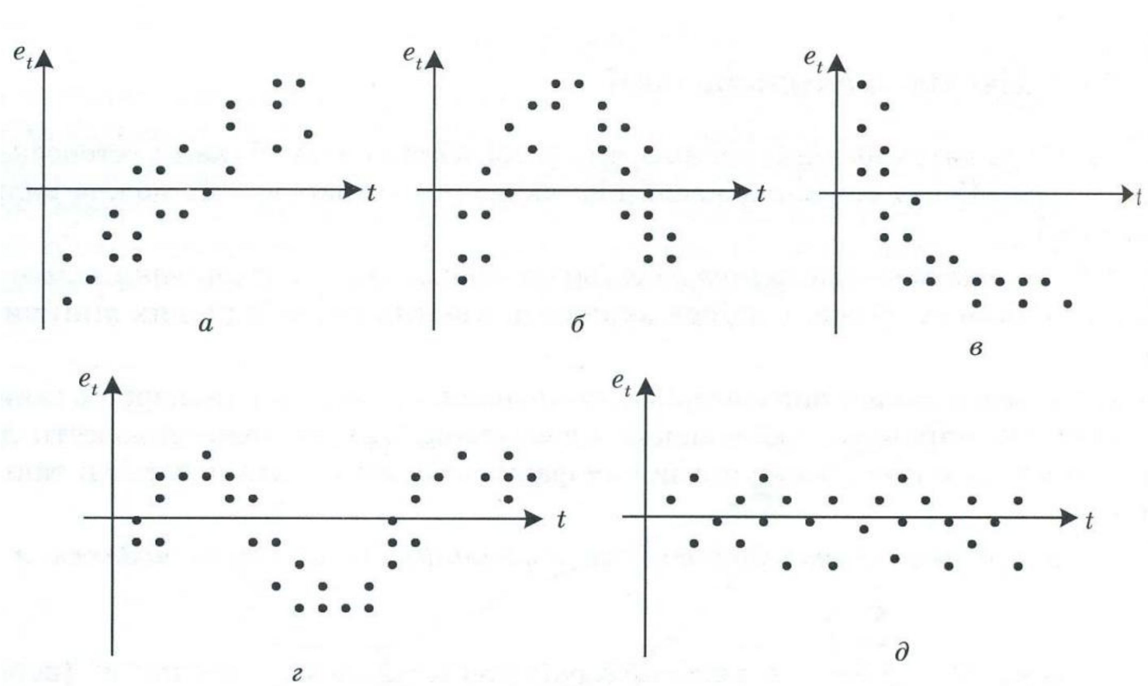


Рисунок 5.2 – Варіанти залежностей відхилень e_t від моментів часу t

Враховуючи наявність систематичності зміни відхилень на рис. 5.2, *а – г* між ними наявна автокореляція, оскільки, наприклад, на рис. 5.2, *б* відхилення спочатку здебільшого від'ємні, потім додатні, потім знову від'ємні. Отже, між ними є певний зв'язок. У цьому прикладі (5.2, *б*) наявна додатна автокореляція випадкових відхилень, яку можна спостерігати наочно, якщо цей рисунок доповнити графіком залежності e_t від e_{t-1}, як показано на рис. 5.3.

Судячи з рис. 5.3, переважна кількість точок зосереджена в I і III четвертях декартової системи координат, що і підтверджує додатну залежність між сусідніми значення випадкових відхилень.

Відсутність систематичності зміни відхилень на рис. 5.2, *д* свідчить про відсутність автокореляції.

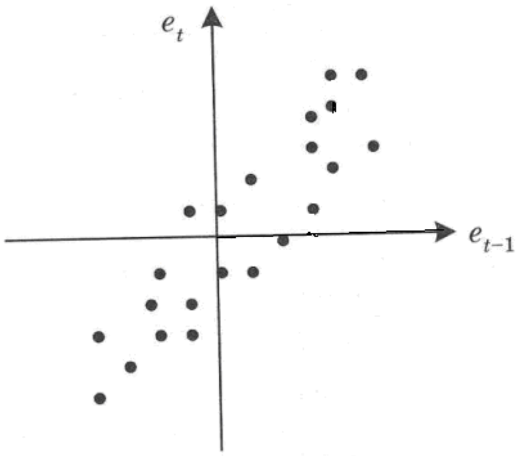


Рисунок 5.3 – Залежність значень випадкових відхилень e_t від e_{t-1}

Приклад 5.1. Перевірити за допомогою графічного методу наявність автокореляції випадкових відхилень, що отримані на базі такої кореляційно-регресійної залежності, яка складена за допомогою даних табл. 5.1.

Розв'язання. Спочатку на базі даних табл. 5.1 побудуємо відповідну однофакторну лінійну регресійну модель вигляду: $\hat{y} = \beta_0 + \beta_1 x + \varepsilon$.

Таблиця 5.1 – Значення реальних та оцінених ознак, їх відхилень

№	y_t	x_t	$\hat{y}_t$	e_t
1	15	29	1.623	13.377
2	18	42	8.790	9.210
3	18	44	9.893	8.107
4	27	63	20.368	6.632
5	40	109	45.730	-5.730
6	36	102	41.870	-5.870
7	42	112	47.384	-5.384
8	38	104	42.973	-4.973
9	31	84	31.946	-0.946
10	40	110	46.281	-6.281
11	41	111	46.832	-5.832
12	40	111	46.832	-6.832
13	38.12	102.4	42.091	-3.971
14	42.16	112.8	47.825	-5.665
15	62.4	148.8	67.673	-5.273
16	69.02	160	73.848	-4.828
17	84.46	183.2	86.639	-2.179
18	98.9	200	95.902	2.998
19	123.04	228.8	111.780	11.260
20	124.4	229.6	112.221	12.179
Σ	1028.5	2386.6	1028.5	0

Отже, таку кореляційно-регресійну модель можна оцінити за допомогою МНК так:

$$\hat{y} = -14{,}3663 + 0{,}551339x.$$

Зобразимо послідовно-часовий графік, який зображає випадкові відхилення e_t в моменти часу t (рис. 5.4).

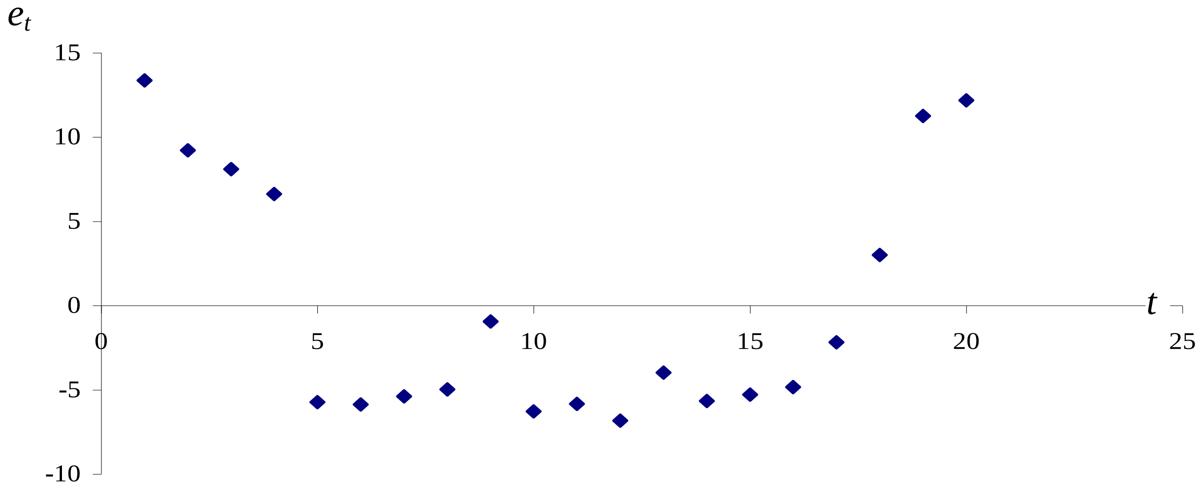


Рисунок 5.4 – Послідовно-часовий графік випадкових відхилень

На рис. 5.4 можна побачити, що є залежність між випадковими відхиленнями e_t, зокрема спочатку відхилення додатні, після цього вони стають від'ємними, а потім знову додатними. У цьому разі можна стверджувати, що наявна автокореляція випадкових відхилень для вибірки.

Розглянемо графік залежності e_t від e_{t-1}, що зображено на рис. 5.5.

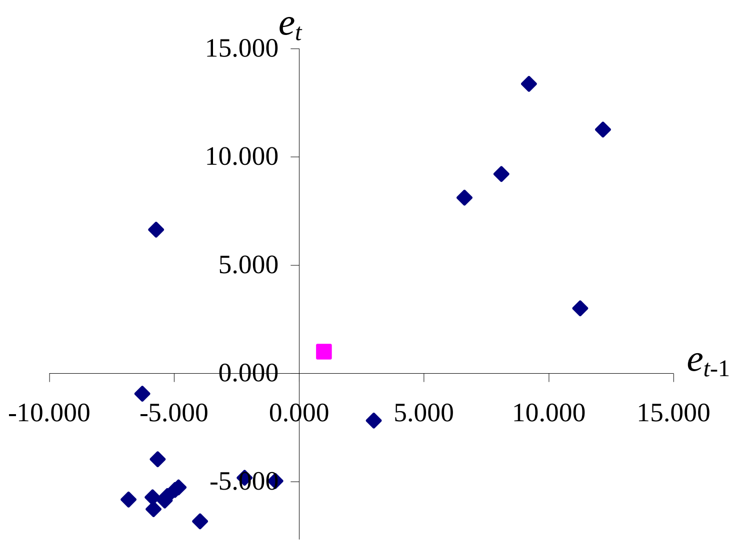


Рисунок 5.5 – Залежність поточних значень e_t випадкових відхилень від попередніх e_{t-1}

Оскільки здебільшого точки на цьому графіку розташовано в I і III четвертях, то можна підтвердити наявність додатної залежності між суміжними значення випадкових відхилень.

5.3 Тестування автокореляції. Критерій Дарбіна-Уотсона

Одним із найчастіше застосовуваних критеріїв тестування автокореляції першого порядку вважають ***критерій Дарбіна-Уотсона***.

Для перевірки статистичної незалежності випадкових відхилень e_t між собою необхідно визначити коефіцієнт кореляції між сусідніми значеннями e_t.

Сусідніми значеннями випадкових відхилень e_t є *сусідні* в часі (якщо розглядають динамічні ряди) або значення e_t, знайдені на підставі проранжованих за зростанням значень факторної ознаки x_t (у разі перехресної вибірки). Для цих величин розраховують коефіцієнт кореляції, який називають коефіцієнтом автокореляції першого порядку:

$$r_{e_t e_{t-1}} = \frac{\sum_{t=2}^{T}((e_t - E(e_t)) \cdot (e_{t-1} - E(e_{t-1})))}{\sqrt{\sum_{t=1}^{T}(e_t - E(e_t))^2 \cdot \sum_{t=2}^{T}(e_{t-1} - E(e_{t-1}))^2}} = \frac{\sum_{t=2}^{T}(e_t \cdot e_{t-1})}{\sqrt{\sum_{t=1}^{T} e_t^2 \cdot \sum_{t=2}^{T} e_{t-1}^2}}. \qquad (5.1)$$

Зауважимо, що математичне сподівання $E(e_t)=0$.

Проте, більш інформативним показником на практиці є не коефіцієнт автокореляції першого порядку, а ***d-статистика*** – ***коефіцієнт Дарбіна-Уотсона*** (***DW***):

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}. \qquad (5.2)$$

Зробимо припущення, що при великих значеннях T: $\sum_{t=1}^{T} e_t^2 \approx \sum_{t=2}^{T} e_{t-1}^2$, тоді між коефіцієнтом автокореляції першого порядку і коефіцієнтом DW можна встановити таку взаємозалежність:

$$\sum_{t=2}^{T}(e_t - e_{t-1})^2 = \sum_{t=2}^{T}(e_t^2 - 2e_t e_{t-1} + e_{t-1}^2) = \sum_{t=2}^{T} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1} + \sum_{t=2}^{T} e_{t-1}^2 \approx 2\sum_{t=1}^{T} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1}. \qquad (5.3)$$

Враховуючи (5.3), залежність (5.2) можна переписати так:

$$DW \approx \frac{2\sum_{t=1}^{T} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} = 2(1 - r_{e_t e_{t-1}}). \tag{5.4}$$

Проаналізуємо цю залежність: якщо $e_t = e_{t-1}$, то $r_{e_t e_{t-1}} = 1$ і $DW = 0$; якщо $e_t = -e_{t-1}$, то $r_{e_t e_{t-1}} = -1$ і $DW = 4$. В усіх інших випадках значення коефіцієнта Дарбіна-Уотсона лежить у межах від 0 до 4, тобто **$0 < DW < 4$**.

До аналогічного результату можна дійти і шляхом таких міркувань. Якщо $e_t \approx e_{t-1}$, то кожний доданок $(e_t - e_{t-1})$ чисельника дробу (5.2) близький до нуля і коефіцієнт DW дорівнює нулю.

Коли точки спостережень по черзі відхиляються в різні боки від лінії регресії $(e_t \approx -e_{t-1})$, то $e_t - e_{t-1} \approx 2e_t$, а це означає, що:

$$DW = \frac{\sum_{t=1}^{T}(2e_t)^2}{\sum_{t=1}^{T} e_t^2} = 4\frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} e_t^2} = 4. \tag{5.5}$$

Такий випадок означає від’ємну автокореляцію випадкових відхилень e_t.

Якщо ж відхилення e_t є випадковими, то вважатимемо, що у частині прикладів знаки випадкових величин e_t є, наприклад, додатними, а в іншій частині – від'ємними. Тоді ***необхідною умовою незалежності випадкових відхилень*** e_i є ***випадок***, коли ***коефіцієнт Дарбіна-Уотсона дорівнює 2***.

Для перевірки наявності автокореляції в моделі використовують таблицю критичних точок розподілу Дарбіна-Уотсона.

Розглянемо зони автокореляційного зв'язку за критерієм Дарбіна-Уотсона, подані на рис. 5.6.

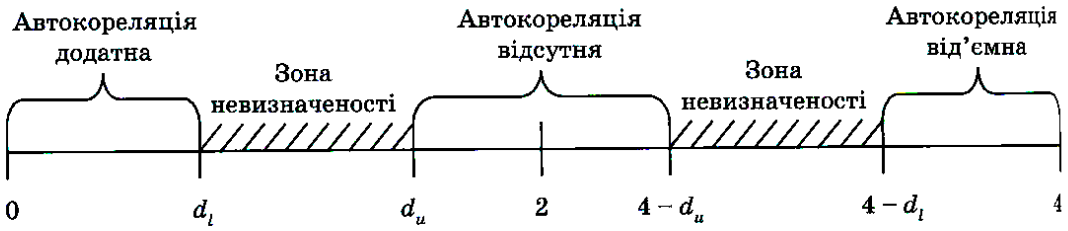


Рисунок 5.6 – Зони автокореляційного зв'язку за критерієм Дарбіна-Уотсона

За таблицею розподілу Дарбіна-Уотсона для заданого рівня значущості α, кількості спостережень T і кількості факторних ознак k визначають два значення: d_l – нижню межу і d_u – верхню межу критичної області за критерієм Дарбіна-Уотсона. При цьому можливі такі випадки (див. рис. 5.6):

- якщо $0 \le DW < d_l$, то наявна додатна автокореляція;
- якщо $d_l \le DW < d_u$, то значення DW потрапляє в зону невизначеності й не можна зробити висновок ні про наявність, ні про відсутність автокореляції;
- якщо $d_u \le DW < 4 - d_u$, то автокореляція відсутня;
- якщо $4 - d_u \le DW < 4 - d_l$, то значення DW потрапляє в зону невизначеності й не можна зробити висновок ні про наявність, ні про відсутність автокореляції;
- якщо $4 - d_l \le DW \le 4$, то наявна від'ємна автокореляція.

Для використання критерію Дарбіна-Уотсона потрібно дотримуватися таких умов:

1) цей критерій використовують лише для моделей, що мають вільний член кореляційно-регресійної моделі;

2) припускають, що випадкові величини ε_t визначають за такою ітераційною схемою:

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + \upsilon_t, \quad -1 < \rho < 1, \tag{5.6}$$

де ρ – коефіцієнт автокореляції;

υ_t – випадкова величина, для якої виконуються всі припущення кореляційно-регресійного аналізу:

$$E(\upsilon_t) = 0; \sigma^2_{\upsilon_t} = \sigma^2; \operatorname{cov}(\upsilon_t, \upsilon_{t+s}) = 0; s \ne 0. \tag{5.7}$$

Модель (5.6) називають *авторегресійною моделлю Маркова*[1] *першого порядку AR*(1), і вона описує автокореляцію першого порядку;

3) статистичні дані повинні мати однакову періодичність, тобто не має бути пропусків у спостереженнях;

4) критерій DW не застосовують для кореляційно-регресійних моделей, що містять у переліку екзогенних змінних лагове значення результуючої змінної, тобто для таких авторегресійних моделей:

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + \lambda y_{t-1} + \varepsilon_t. \tag{5.8}$$

Приклад 5.2. Перевірити кореляційно-регресійну модель з прикладу 5.1 (табл. 5.1) на наявність автокореляції за допомогою критерію Дарбіна-Уотсона.

[1] Марков Андрій Андрійович (1856 – 1922) – російський математик, професор Санкт-Петербурзького університету, заклав основи теорії випадкових процесів, зокрема дослідив окремий вид випадкових процесів, які отримали назву "марковські ланцюги".

Розв'язання. На основі даних табл. 5.1 розрахуємо у табл. 5.2 необхідні для визначення критерію Дарбіна-Уотсона суми квадратів випадкових відхилень згідно з залежністю (5.2).

Таблиця 5.2 – Дані для обчислення критерію Дарбіна-Уотсона

e_t	$(e_t - e_{t-1})^2$	e_t^2
13.377		178.957
9.210	17.367	84.825
8.107	1.216	65.730
6.632	2.177	43.983
-5.730	152.809	32.829
-5.870	0.020	34.460
-5.384	0.237	28.984
-4.973	0.169	24.730
-0.946	16.215	0.895
-6.281	28.460	39.451
-5.832	0.201	34.016
-6.832	1.000	46.681
-3.971	8.188	15.767
-5.665	2.869	32.089
-5.273	0.153	27.804
-4.828	0.198	23.309
-2.179	7.017	4.748
2.998	26.807	8.991
11.260	68.251	126.785
12.179	0.844	148.324
Σ	334	1003

Використовуючи (5.2), оцінимо критерій Дарбіна-Уотсона так:

$$DW = \frac{334}{1003} = 0{,}333.$$

Знайдемо із таблиці розподілу Дарбіна-Уотсона для рівня значущості $\alpha = 0{,}05$ та кількості спостережень $n = 20$ значення критичних точок: $d_l = 1{,}201$ та $d_u = 1{,}411$. Оскільки ж розраховане значення потрапляє в інтервал $[0;\ d_l)$, $(0{,}33 < 1{,}201)$, то з імовірністю $p = 0{,}95$ можна стверджувати, що у вибірковій сукупності наявна додатна автокореляція.

5.4 Сутність гетероскедастичності, її наслідки

Для побудови якісної кореляційно-регресійної моделі однією з необхідних умов є однакова дисперсія випадкових величин – ***гомоскедастичність***, тобто:

$$D(\varepsilon_i / x_i) = \sigma_i^2 = const, \ i = \overline{1, n}, \tag{5.9}$$

де $D(\varepsilon_i / x_i)$ – умовна дисперсія випадкової величини ε_i (дисперсія випадкової величини ε_i, зумовлена значенням фактора x_i).

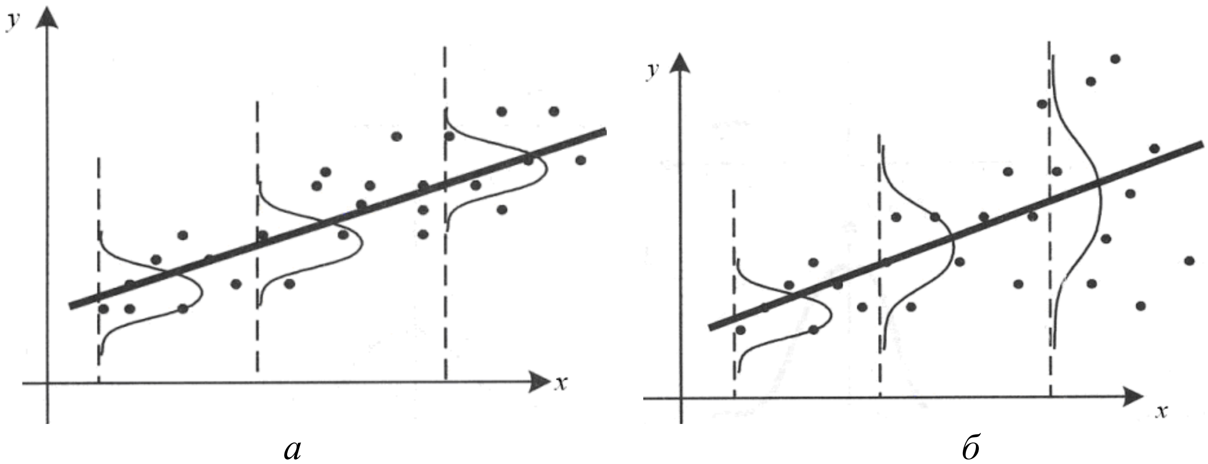
Отже, умову гомоскедастичності можна це записати так: $\sigma_i^2 = \sigma_j^2$ для всіх *i* та *j*. Сутність такого явища як гомоскедастичність виявляється у незалежності варіації значень кожної ε_i навколо її математичного сподівання від значень *x*, тобто дисперсія кожної ε_i залишається сталою незалежно від малих чи великих значень фактора: σ_ε^2 не є функцією від *x*.

У протилежному випадку, коли умовна дисперсія випадкових величин не є сталою, тобто значення σ_ε^2 залежать від *x*, то наявною є **гетероскедастичність**:

$$D(\varepsilon_i / x_i) = \sigma_i{}^2 \neq const, \tag{5.10}$$

тобто при гетероскедастичності $\sigma_i^2 \neq \sigma_j^2$.

Розглянемо на рис. 5.7 два таких кореляційних поля.



а *б*

Рисунок 5.7 – Кореляційні поля та дисперсії відхилень регресійної моделі

В обох випадках зі зростанням *x* збільшується середнє значення *y*. Однак, якщо на рис. 5.7, *а* дисперсія споживання залишається сталою для різних рівнів x, то на рис. 5.7, *б* при аналогічній залежності середнього y від x дисперсія y не залишається сталою, а збільшується зі зростанням x.

Розкид значень y викликає розкид точок спостережень відносно лінії регресії, що й визначає дисперсію випадкової величини. Динамічне змінення дисперсій випадкових величин ε для даного випадку розглянуто на рис. 5.8. На рис. 5.8, *а* (випадок гомоскедастичності) дисперсії ε_i постійними, а на рис. 5.8, *б* (при гетероскедастичності) дисперсії ε_i змінюються (зростають).

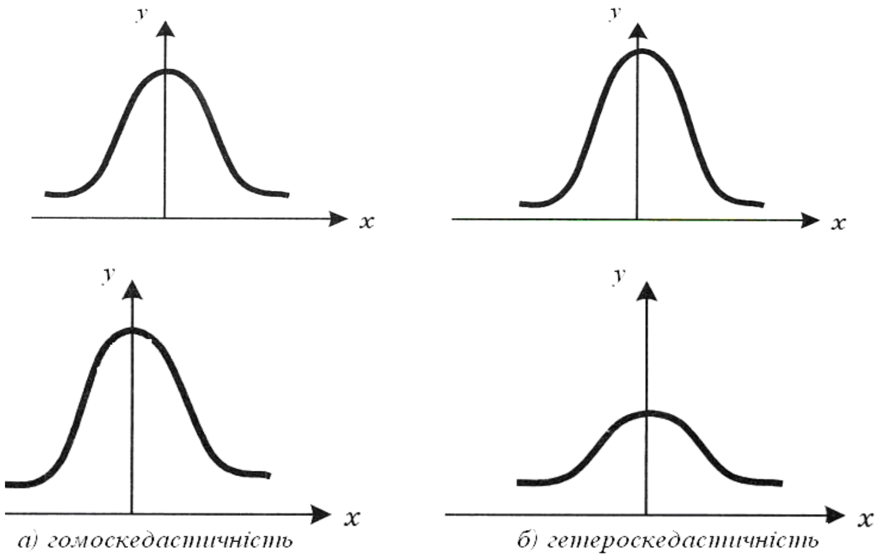


Рисунок 5.8 – Зміна дисперсії відхилень при гомо- та гетероскедатичності

Гетероскедастичність у більшості виникає через ефект масштабу у варіаційних рядах, що описують економічні суб'єкти, які функціонують на різних рівнях – від макро- (країна), мезо- (галузь) до мікрорівня (споживач, домогосподарство, підприємство) і відповідно мають різні доходи, розміри і т. п. На відміну від варіаційних рядів, у динамічних рядах гетероскедастичність не спостерігається, оскільки вони зазвичай розглядають показники одного і того самого економічного суб'єкта за низку періодів (наприклад, ВВП, чистий експорт, темпи інфляції у деякому регіоні). Слід зауважити, що проблема гетероскедастичності може виникти під час збільшення або зменшення рівнів динамічного ряду.

Наслідки гетероскедастичності. У разі відсутності гомоскедастичності, виникають такі наслідки при застосуванні МНК:

1. Незміщеність і лінійність оцінок коефіцієнтів моделі;
2. Оцінки не матимуть найменшої дисперсії порівняно з іншими оцінками невідомого параметра. Збільшення дисперсії оцінок знижує ймовірність отримання максимально точних оцінок;
3. Зміщеність дисперсії оцінок параметрів регресії. Наприклад, для парної кореляційно-регресійної моделі за умови гомоскедастичності дисперсії оцінок розраховують за формулами:

$$D_{b_0} = \sigma_\varepsilon^2 \cdot \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \overline{x})^2}; \qquad D_{b_1} = \sigma_\varepsilon^2 \cdot \frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{5.11}$$

Зміщення відбувається внаслідок того, що непояснена дисперсія кореляційно-регресійної моделі $\sigma_\varepsilon^2 = \frac{\sum_{i=1}^{n}\varepsilon_i^2}{n-k-1}$ (k – кількість факторних ознак), яку використовують під час обчислення оцінок дисперсій коефіцієнтів b_0 та b_1, буде зміщеною;

4. Усі висновки, отримані на підставі відповідних t- і F-статистик, а також інтервальні оцінки будуть ненадійними. Отже, статистичні висновки, одержані на підставі стандартної перевірки якості оцінок, можуть бути помилковими й призводити до неправильно зроблених висновків на основі побудованої кореляційно-регресійної моделі. Ймовірно, що стандартні помилки коефіцієнтів будуть занижені, а отже, t-статистики будуть завищені. Це може призвести до визнання статистично незначущих параметрів регресії статистично значущими;
5. Зростання довірчих інтервалів параметрів моделі. Якщо наявна гетероскедастичність, дисперсію $D^*_{d_1}$ параметра b_1 розраховують за формулою:

$$D^*_{b_1} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sigma_i^2}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2}. \tag{5.12}$$

Дисперсія параметра b_1 у цьому разі при $\sigma_i^2 > \sigma_\varepsilon^2$ перевищує аналогічну дисперсію при відсутності гетероскедастичності: $D^*_{b_1} > D_{b_1}$, що впливає на зростання довірчого інтервалу.

Причину неефективності оцінок параметрів, отриманих за допомогою методу найменших квадратів при гетероскедатичності, пояснимо на прикладі парної лінійної кореляційно-регресійної моделі.

На рис. 5.9 показано, що кожному конкретному значенню x_i відповідає значення y_i з деякої множини, яка має свій закон розподілу та несталу дисперсію. За критерій методу найменших квадратів беремо мінімізацію суми квадратів відхилень: $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \to min$.

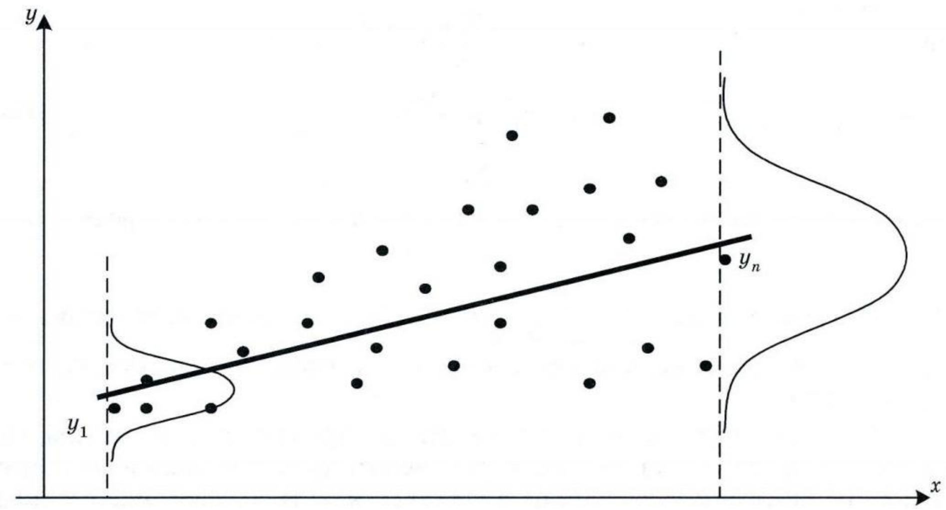


Рисунок 5.9 – Несталість дисперсій при гетероскедастичності

У даному прикладі кожне значення e_i^2 у цій сумі однаково важливе, незалежно від того, отримано воно на підставі розподілу з малою (наприклад, e_1^2) або великою (наприклад, e_n^2) дисперсією. Проте це суперечить логіці, оскільки точка, отримана на підставі розподілу з малою дисперсією, точніше визначає напрям лінії регресії. Тому вона важливіша, ніж точка, отримана на підставі розподілу з великою дисперсією. Саме тому методи оцінювання, що враховують значення точок спостережень, дають можливість одержувати точніші (ефективніші) оцінки. Урахування значення точок спостережень характерне, наприклад, для методу зважених найменших квадратів.

5.5 Тестування гетероскедастичності. Графічний аналіз випадкових відхилень

Інколи на підставі знань про характер статистичних даних появу проблеми гетероскедастичності можна передбачати і спробувати її усунути ще на етапі специфікації кореляційно-регресійної моделі, провівши глибокий аналіз досліджуваної проблеми. Проте значно частіше цю проблему доводиться вирішувати після побудови кореляційно-регресійної моделі.

Виявлення гетероскедастичності у кожному разі є досить складним завданням, оскільки для знання дисперсій відхилень $\sigma^2(e_i) = \sigma_i^2$ потрібно знати закон розподілу випадкової величини ε, що відповідає вибраному значенню x_i. Дуже часто на практиці для кожного конкретного значення x_i визначають лише одне значення y_i, що не дає можливості оцінити дисперсію випадкових величин ε для даного x_i.

Для тестування гетероскедастичності використовують різноманітні підходи: як графічний, так і аналітичні методи тестування.

Графічний аналіз допомагає не лише виявити гетероскедастичність, а й зробити висновок про саму форму зв'язку, що особливо корисно при трансформації наявних даних для побудови моделі з гомоскедастичністю помилок. Для цього спочатку будують регресійну модель на базі припущення про відсутність гетероскедастичності, а потім за допомогою дослідження квадратів залишків e_i^2 з'ясовують, чи мають вони якусь систематичність. Звичайно, e_i^2 є лише оцінками невідомих ε_i^2, проте вони можуть успішно використовуватися, особливо за умови великих вибірок даних.

Розглянемо різні види графіків, зображені на рис. 5.10. На цьому рисунку квадрати залишків розсіяні навколо оціненої лінії регресії.

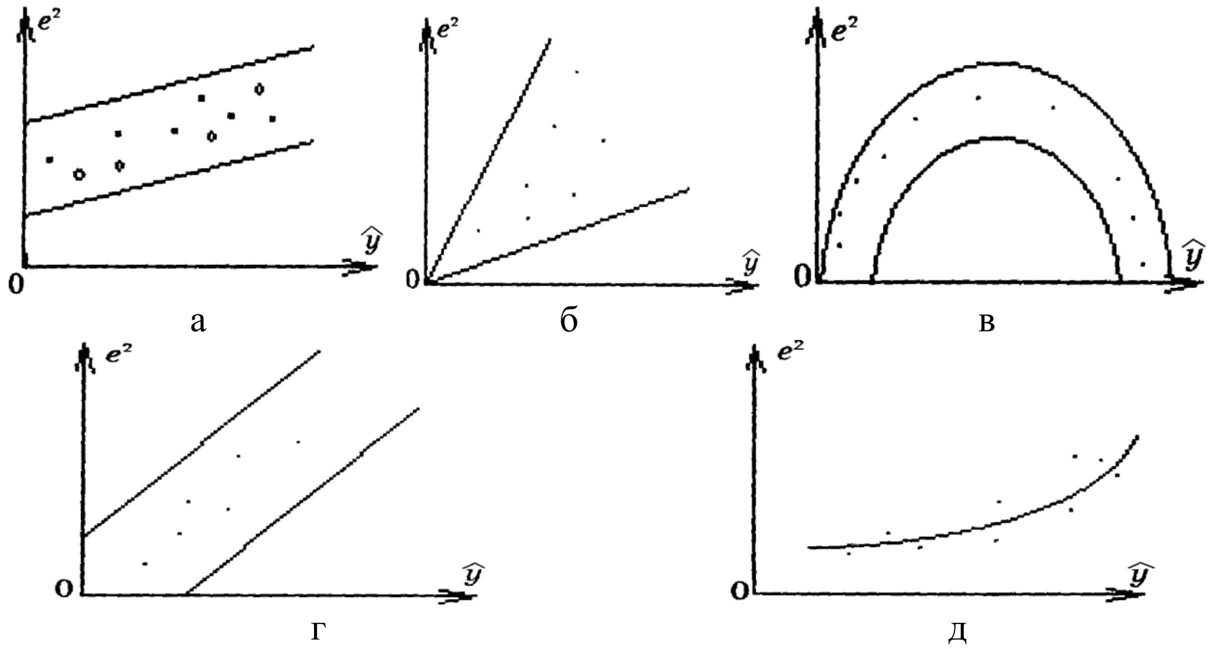


Рисунок 5.10 – Типи залежності квадратів залишків від оцінених значень $\hat{y}$

Графіки дозволяють з'ясувати, чи систематично оцінене середнє значення залежної змінної, пов'язане з квадратом залишків. Зокрема, на рис. 5.10, а зображено ситуацію, за якої систематичний зв'язок відсутній, тобто гетероскедастичності немає. На рис. 5.10, г показано лінійний зв'язок, а на 5.10, в − квадратичну залежність.

Завдяки графічному методу можна проаналізувати залежність не лише між y та e^2, а й між будь-яким фактором x та e^2. Графіки залишаються такими ж при наявності гетероскедастичності. У цьому разі по осі абсцис відкладають значення факторної ознаки x (або значення лінійної комбінації факторних ознак $\hat{y}_i = b_0 + b_1 x_1 + \cdots + b_k x_k$), а по осі ординат – відхилення e або їх квадрати e^2. Такі графіки дають можливість проаналізувати, чи систематично залежать квадрати випадкових відхилень e^2 від упорядкованих значень факторної ознаки (або оцінених середніх значень результуючої змінної $\hat{y}$).

Зазначимо, що графічний аналіз випадкових відхилень є зручним і достатньо надійним у разі парної кореляційно-регресійної моделі. Застосування графічного аналізу випадкових відхилень, щоб тестувати гетероскедастичність множинної кореляційно-регресійної моделі, можливе для кожної з факторних ознак x_j, $j=\overline{1,k}$ окремо.

Перевіримо наявність гетероскедастичності графічним методом за даними прикладу 5.1. Впорядкуємо дані відносно x_i (рис. 5.11) та $\hat{y}_i$ (рис. 5.12) за зростанням. Відобразимо їх залежності від квадрата залишків на графіках (рис. 5.11 та 5.12).

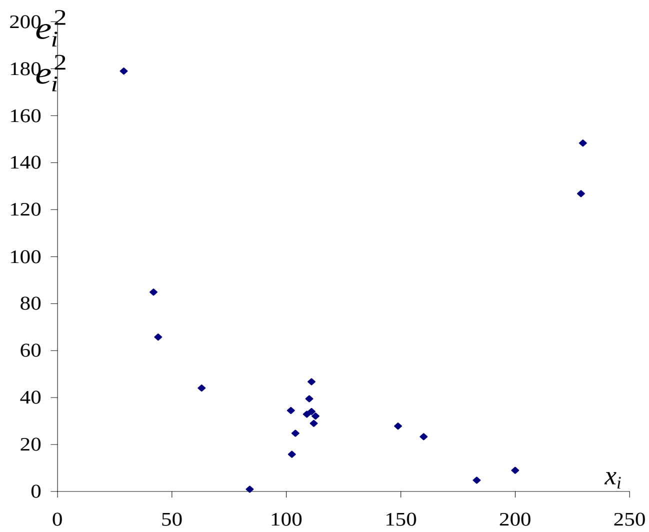


Рисунок 5.11 – Залежність впорядкованої факторної ознаки від залишків

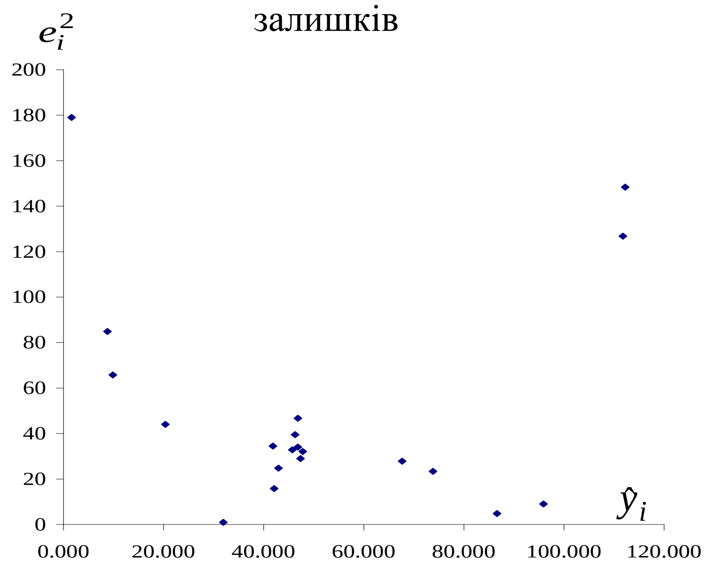


Рисунок 5.12 – Залежність впорядкованої оціненої ознаки від залишків

Аналізуючи графіки, зображені на рис. 5.11 та 5.12, можна в обох випадках побачити параболічну залежність між квадратом залишків та факторною (або оціненою) ознаками, що свідчить про наявність гетероскедастичності у даних спостереженнях.

5.6 Тестування гетероскедастичності. Аналітичні методи

Більш точними вважаються аналітичні методи тестування. Нині немає єдиного універсального методу визначення гетероскедастичності. Щоб її протестувати можна використовувати три групи методів:

1. Методи, що здійснюють тестування гетероскедастичності у довільних неперервних функціях $\sigma^2(e_i)$;

2. Методи, що тестують монотонні функції $\sigma^2(e_i)$;

3. Методи, які визначають гетероскедастичність у певних функціях $\sigma^2(e_i)$, тобто таких, для яких задано форму залежності та кількість параметрів.

Проілюструємо застосування методів ***першої групи*** для визначення гетероскедастичності у кореляційно-регресійних моделях. Одним із таких методів є **критерій Бартлета**:

$$b^c = \frac{n \ln b^s}{1 + \dfrac{\sum\limits_{l=1}^{m} \dfrac{1}{n_l} + \dfrac{1}{n}}{3(m-1)}}, \text{ при цьому } b^s = \frac{\dfrac{1}{m}\sum\limits_{l=1}^{m} s_l^2}{\sqrt[m]{s_1^2 \cdot s_2^2 \cdot s_2^2 \cdot \cdots \cdot s_m^2}},$$

де n – сукупна кількість спостережень;

n_l – кількість спостережень в l-му інтервалі, $l = \overline{1, m}$;

m – кількість інтервалів спостережень;

s_l^2 – дисперсія випадкових відхилень.

Значення критерію b^c порівнюють із табличним значенням $\chi^2_{кр}$ – розподілу, що береться для відповідного рівня значущості α та кількості ступенів вільності $\nu = m$ - 1.

Якщо табличне значення $b^c \geq \chi^2_{кр}$, то з ймовірністю ρ = 1-α нульову гіпотезу H_0 слід відхилити, тобто наявна гетероскедастичність випадкових відхилень, якщо навпаки, то відхилення в кореляційно-регресійній моделі є гомоскедастичними.

Приклад 5.3. За допомогою критерію Бартлета здійснити тестування гетероскедастичності для кореляційно-регресійної моделі з прикладу 5.1, (табл. 5.1).

Проранжувавши x за зростанням, сформуємо 4 інтервали відповідних їм значень випадкових відхилень, як вказано у табл. 5.3. На базі даних таблиці 5.1 і 5.3 розрахуємо дисперсії випадкових відхилень s_l^2 (табл. 5.3).

Таблиця 5.3 – Залежність дисперсії випадкових відхилень від x

Межі інтервалів	Кількість спостережень в інтервалі, n_l	Дисперсія випадкових відхилень, s_l^2
[29 – 84]	5	27,4172845
(84 – 104]	3	0,902469
(104 – 112,8]	6	0,2702588
(112,8 – 229,6]	6	61,3124715

Визначимо спочатку значення b^s:

$$b^s = \frac{\frac{1}{4}\left(27,412845 + 0,902469 + 0,2702588 + 613,124715\right)}{\sqrt[4]{27,412845 \cdot 0,902469 \cdot 0,2702588 \cdot 613,124715}} = 4,9947681.$$

Значення критерію Бартлета дорівнює:

$$b^c = \frac{20}{1 + \frac{\frac{1}{5} + \frac{1}{3} + \frac{1}{6} + \frac{1}{6} + \frac{1}{20}}{3(4-1)}} \cdot \ln(4,9947681) = 29,194.$$

Порівняємо отримане значення із табличним для рівня значущості 95% та кількості ступенів вільності ν = 4 – 1 = 3. Табличне значення становить $\chi^2_{кр} = 7{,}81$. Отже, оскільки 29,194 > 7,81 ($b^c \geq \chi^2_{кр}$), то з ймовірністю ρ = 1 - 0,05 = 0,95 нульову гіпотезу H_0 слід відхилити, тобто наявна гетероскедастичність випадкових відхилень.

До ***методів другої групи*** відносять тест рангової кореляції Спірмена та тест Голдфельда-Квандта.

Тест рангової кореляції Спірмена. До методів ***другої групи*** належить тест рангової кореляції Спірмена. При його використанні припускають, що дисперсія випадкових відхилень ε монотонно збільшується або зменшується зі зростанням значень незалежної змінної x. Отже, за МНК абсолютні значення відхилень $|e_i|$ і x_i є корельованими. Їх ранжують (присвоюють найменшому значенню одиничний ранг, наступному – ранг, що дорівнює 2, і т. п.), упорядковуючи за зростанням.

Далі оцінюють ***коефіцієнт рангової кореляції Спірмена***:

$$r_{x,e} = 1 - 6 \cdot \frac{\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{5.13}$$

де n – кількість спостережень;

d_i – різниця між рангами x_i та e_i (наприклад, якщо x_{10} стоїть на 12-му місці за зростанням величини серед усіх спостережень x_i, а e_{10} – на 15-му місці, то d_{10} = 12-15=-3).

Нульову гіпотезу H_0 визначають так: *коефіцієнт рангової кореляції $\rho_{x,\varepsilon}$ між факторною ознакою x та випадковою величиною ε дорівнює нулю:* $\rho_{x,\varepsilon}$ = 0, а *альтернативну гіпотезу H_1* формулюють протилежно: *коефіцієнт рангової кореляції $\rho_{x,\varepsilon}$ між факторною ознакою x та випадковою величиною ε не дорівнює нулю:* $\rho_{x,\varepsilon} \neq 0$.

Відомо, що коли коефіцієнт рангової кореляції $\rho_{x,\varepsilon}$ для генеральної сукупності дорівнює нулю, то критерій перевірки гіпотези визначається так:

$$t = \frac{r_{x,e}\sqrt{n-2}}{\sqrt{1-r_{x,e}^2}} \tag{5.14}$$

і має розподіл Стьюдента з кількістю ступенів вільності $\nu = n - 2$.

Для перевірки нульової гіпотези знаходять у таблицях розподілу Стьюдента (для даного рівня значущості α та кількості ступенів вільності $\nu = n$-2) критичне значення статистики $t_\alpha^{кр}$.

При цьому, якщо розраховане значення t не менше за критичне значення $t_\alpha^{кр}$, то H_0 гіпотезу про нульове значення коефіцієнта рангової кореляції $\rho_{x,\varepsilon}$ з довірчою ймовірністю p = 1 - α потрібно відхилити, тобто наявна гетероскедастичність і, навпаки, якщо $t < t_\alpha^{кр}$, то гетероскедастичність є відсутньою.

Для випадку множинної кореляційно-регресійної моделі гіпотезу про коефіцієнт рангової кореляції $\rho_{x,\varepsilon}$, що дорівнює нулю, перевіряють за допомогою t-статистики для кожної з результативних ознак x_i окремо.

Приклад 5.4. Протестувати наявність гетероскедастичності на основі тесту рангової кореляції Спірмена для кореляційно-регресійної моделі, розглянутої у прикладі 5.1.

Розв'язання. Значенням випадкових відхилень e_i та факторної ознаки x_i, впорядковуючи її за зростанням, присвоїмо відповідно ранги 1, 2, ..., 20. Ранги випадкових відхилень e_i та факторної ознаки x_i, різниці між цими рангами d_i та квадрати різниць d_i^2 наведено у табл. 5.4.

Використовуючи (5.13), знайдемо значення коефіцієнта рангової кореляції Спірмена: $r_{x,e} = 1 - 6 \cdot \dfrac{\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$, Таблиця 5.4 – Дані для розрахунку тесту Спірмена

x_i	Ранг x_i	$\lvert e_i \rvert$	Ранг e_i	d_i	d_i^2
29	1	13.377	20	-19	361
42	2	9.200	17	-15	225
44	3	8.107	16	-13	169
63	4	6.632	14	-10	100
84	5	5.730	10	-5	25
102	6	5.870	12	-6	36
102.4	7	5.384	8	-1	1
104	8	4.973	6	2	4
109	9	0.946	1	8	64
110	10	6.281	13	-3	9
111	11.5	5.832	11	0.5	0.25
111	11.5	6.832	15	-3.5	12.25
112	13	3.971	4	9	81
112.8	14	5.665	9	5	25
148.8	15	5.273	7	8	64
160	16	4.828	5	11	121
183.2	17	2.179	2	15	225
200	18	2.998	3	15	225
228.8	19	11.260	18	1	1
229.6	20	12.179	19	1	1
2386.6	210	127.517	210	0	1749.5

$$\left| r_{x,e} \right| = 1 - 6 \cdot \frac{1749.5}{20(20^2 - 1)} = 0{,}3154.$$

Отже, зв'язок між випадковою змінною *e* і факторною ознакою *x* є слабким.

Для перевірки нульової гіпотези H_0 обчислимо значення *t*-статистики на базі залежності (5.14):

$$t = \frac{0{,}3154\sqrt{20-2}}{\sqrt{1-0{,}3154^2}} = 1{,}41.$$

Порівняємо табличне значення $t_{кр}$ при ступені вільності df = 20-2 =18 (для кількості спостережень n = 20) та рівні значимості 0,9 із розрахованим значенням *t*. Табличне значення складає $t_{кр}$ = 1,33.

Оскільки $t > t_{кр}$, то H_0 гіпотеза про нульове значення коефіцієнта рангової кореляції $\rho_{x,\varepsilon}$ відкидається (коефіцієнт рангової кореляції є статистично значимим), тобто наявною є гетероскедастичність залишків регресійної моделі.

Параметричний тест Гольдфельда-Квандта. Гольдфельд і Квандт у своєму методі пропонують розглянути випадок, коли дисперсія залишків зростає пропорційно квадрату незалежної змінної $M(ee') = \sigma_e^2 x_{ij}^2$. Тоді тест полягає у виконанні таких етапів.

Етап 1. Впорядкувати спостереження за зростанням незалежної змінної X.

Етап 2. Вилучити c спостережень, що містяться у центрі векторів X та Y:

$$c = 4 \cdot n/15,$$

де n – кількість спостережень.

Етап 3. Побудувати дві регресійні моделі на базі методу найменших квадратів на основі новоутворених вибірок спостережень, які відповідно мають обсяг:

$$n_{1(2)} = (n - c)/2,$$

причому розмір новоутворених сукупностей $n_{1(2)}$ має бути більшим за кількість змінних моделі.

Етап 4. Визначити суму квадратів залишків за кожною із моделей S_1 та S_2:

$$S_1 = e_1'e_1; \quad S_2 = e_2'e_2,$$

де e_1 – залишки за першою моделлю;
e_2 – залишки за другою моделлю.

Етап 5. Визначити співвідношення R^*:

$$R^* = S_1 / S_2.$$

Це співвідношення, якщо модель є гомоскедастичною, відповідає F-розподілу зі ступенями вільності відповідно $\nu_{1(2)}$, що визначаються як $(n - c - 2m)/2$. Тоді розраховане значення R^* порівнюють із табличним (для відповідного рівня значимості) значенням F-розподілу зі ступенями вільності $\nu_1=(n - c - 2m)/2$ та $\nu_2=(n - c - 2m)/2$. При цьому, якщо $R^* \le F_{табл.,\nu_1,\nu_2}$, то спостережувані дані є гомоскедастичними.

Приклад 5.5. На основі параметричного тесту Гольдфельда-Квандта перевірити відсутність гетероскедастичності за даними прикладу 5.1.

Розв'язання. На етапі 1 впорядкуємо спостереження за зростанням незалежної змінної x_i, як вказано у табл. 5.5.

Таблиця 5.5 – Відсортовані за зростанням x дані прикладу 5.1

№	y_i	x_i
1	15	29
2	18	42
3	18	44
4	27	63
5	31	84
6	36	102
7	38.12	102.4
8	38	104
9	40	109
10	40	110
11	41	111
12	40	111
13	42	112
14	42.16	112.8
15	62.4	148.8
16	69.02	160
17	84.46	183.2
18	98.9	200
19	123.04	228.8
20	124.4	229.6

На другому етапі вилучимо 6 спостережень (6 $\approx$ 4·20/15), що містяться у центрі векторів X та Y (спостереження від 8-го до 13-го).

На третьому етапі побудуємо дві економетричні моделі на основі вибірок обсягом відповідно n_1 = (20 - 6)/2 = 7 та n_2 = (20 - 6)/2 = 7, тобто за вибірками (1-7) та (14-20). Для цього скористаємося даними кореляційних таблиць 5.6 та 5.7.

Отже, перша модель набуває вигляду: $\hat{y} = 5{,}572 + 0{,}31x$, а друга $\hat{y} = -43{,}03 + 0{,}72x$.

На четвертому етапі S_1 та S_2 візьмемо як суму квадратів залишків (зі стовпців з e_t^2), розрахованих у вищевикладених табл. 5.6 та 5.7.

На етапі 5 визначимо співвідношення R^* так: R^* = 56,365 / 8,01878 = = 7,029078.

Таблиця 5.6 – Дані для розрахунків за першою моделлю

№	y_t	x_t	$\hat{y}_t$	e_t	$x_t y_t$	x^2	$(\hat{y}_t-\bar{y})^2$	$(y_t-\bar{y})^2$	e_t^2
1	15	29	14.533	0.467	435	841	135.19	124.55	0.2182
2	18	42	18.550	-0.550	756	1764	57.914	66.586	0.30233
3	18	44	19.168	-1.168	792	1936	48.89	66.586	1.36385
4	27	63	25.039	1.961	1701	3969	1.2571	0.7056	3.84638
5	31	84	31.528	-0.528	2604	7056	28.812	23.426	0.27849
6	36	102	37.090	-1.090	3672	10404	119.46	96.826	1.18737
7	38.12	102.4	37.213	0.907	3903	10486	122.17	143.04	0.82217
Σ	183.1	466.4	183.120	0.000	13863	36456	513.7	521.72	**8.01878**

Таблиця 5.7 – Дані для розрахунків за другою моделлю

№	y_t	x_t	$\hat{y}_t$	e_t	$x_t y_t$	x_t^2	$(\hat{y}_t-\bar{y})^2$	$(y_t-\bar{y})^2$	e_t^2
14	42.16	112.8	37.836	4.324	4755.6	12724	2352.7	1951.87	18.698
15	62.4	148.8	63.645	-1.245	9285.1	22141	515.08	573.124	1.5491
16	69.02	160	71.674	-2.654	11043	25600	215.09	299.982	7.044
17	84.46	183.2	88.306	-3.846	15473	33562	3.8667	3.5344	14.795
18	98.9	200	100.350	-1.450	19780	40000	196.29	157.754	2.1039
19	123.04	228.8	120.998	2.042	28152	52349	1201.1	1346.89	4.1717
20	124.4	229.6	121.571	2.829	28562	52716	1241.2	1448.56	8.0029
∑	604.38	1263	604.380	0.000	117051	239093	5725.4	5781.72	**56.365**

На останньому етапі за таблицями розподілу Фішера визначимо значення F-критерію для рівня значимості 0,95 зі ступенями вільності відповідно $\nu_1 = (20 - 6 - 2\cdot 2)/2 = 5$ та $\nu_2 = (20 - 6 - 2\cdot 2)/2 = 5$. Воно для рівня значимості 0,95 становить $F_{0,95,\,5,\,5} = 5,05$. Порівняємо значення критерію R^* (7,029078) із $F_{0,95,\,5,\,5}$ (5,05). Оскільки $R^* > F_{0,95,\,5,\,5}$, то спостережувані дані є гетероскедастичними.

Тест Парка. Цей тест належить до ***третьої групи методів***. Розглянемо механізм тестування гетероскедастичності за допомогою тесту Парка у однофакторній лінійній регресійній моделі вигляду $y = \beta_0 + \beta_1 x + \varepsilon$.

Припустімо, що дисперсія $\sigma_i^2 = \sigma^2(e_i)$ є функцією i-го значення x_i факторної ознаки x. Парк запропонував таку функціональну залежність:

$$y^2 = k^2 x^{\beta} \exp(\upsilon), \tag{5.15}$$

де k, β – деякі невідомі параметри;

υ – випадкова величина.

Прологарифмувавши (5.15), отримаємо:

$$\ln(y^2) = \ln(k^2) + \beta \ln x + \upsilon. \qquad (5.16)$$

Якщо метод Парка застосовують практично, то в моделі (5.16) замість y^2 беруть квадрат випадкового відхилення e^2.

Алгоритм реалізації тесту Парка є таким.

1. Оцінити вибіркову однофакторну лінійну регресійну модель.

2. Знайти відхилення e_i.

3. Визначити для кожного спостереження $\ln e_i^2 = \ln(y_i - \hat{y}_i)^2$.

4. Побудувати кореляційно-регресійну модель вигляду:

$$\ln(e^2) = \alpha + \beta \cdot \ln x + \upsilon, \text{ або } z = \alpha + \beta u + \upsilon, \qquad (5.17)$$

де $z = \ln(e^2)$, $u = \ln x$, $\alpha = \ln(k^2)$.

Нехай відповідна вибіркова модель має вигляд:

$$\hat{z} = a_0 + a_1 u,$$

де a_0 – оцінка вільного члена α моделі (5.17);

a_1 – оцінка коефіцієнта регресії β моделі (5.17).

У разі множинної кореляційно-регресійної моделі (5.17) будують для кожної факторної ознаки x_j.

5. На базі t-статистики перевірити статистичну значущість коефіцієнта a_1 моделі (5.17):

$$t_{a_1} = \frac{|a_1|}{S_{a_1}}.$$

Якщо коефіцієнт a_1 статистично значущий, то це означає, що між $\ln(e_i^2)$ та $\ln(x_i)$ є зв'язок, тобто у вибірковій сукупності наявна гетероскедастичність. У іншому випадку спостерігається гомоскедастичність.

Приклад 5.5. На основі тесту Парка перевірити наявність гетероскедастичності за даними прикладу 5.1.

Розв'язання.

1. Вибіркова парна лінійна регресійна модель має вигляд:

$$\hat{y} = -14{,}3663 + 0{,}551339x.$$

2. Скористаємося визначеними у табл. 5.2 e_i^2 і занесемо їх у відповідні комірки табл. 5.8.

3. Розраховані для кожного спостереження $\ln e_i^2$ та $\ln(x_i)$ занесемо до табл. 5.8.

Таблиця 5.8 – Значення допоміжних змінних для тесту Парка

$z_i = \ln(e_i^2)$	$u_i = \ln(x_i)$	$z_i u_i$	u_i^2	$\hat{z}_i$	$(\hat{z}_i - \bar{z})^2$	$(z_i - \bar{z})^2$
5.19	3.36729583	17.47	11.339	4.0492	0.39382	3.116679
4.44	3.737669618	16.6	13.97	3.8685	0.199682	1.038194
4.19	3.784189634	15.84	14.32	3.8458	0.179913	0.583381
3.78	4.143134726	15.68	17.166	3.6707	0.062023	0.131153
-0.11	4.430816799	-0.49	19.632	3.5304	0.011814	12.47988
3.54	4.624972813	16.37	21.39	3.4356	0.000195	0.013935
2.76	4.628886713	12.77	21.427	3.4337	0.000145	0.440399
3.21	4.644390899	14.9	21.57	3.4262	2.02E-05	0.045631
3.49	4.691347882	16.38	22.009	3.4033	0.000339	0.004868
3.68	4.700480366	17.27	22.095	3.3988	0.000523	0.06421
3.53	4.709530201	16.61	22.18	3.3944	0.000744	0.011037
3.84	4.709530201	18.1	22.18	3.3944	0.000744	0.177724
3.37	4.718498871	15.89	22.264	3.39	0.001002	0.003003
3.47	4.725616339	16.39	22.331	3.3865	0.001234	0.002204
3.33	5.002603122	16.63	25.026	3.2514	0.028989	0.009305
3.15	5.075173815	15.98	25.757	3.216	0.042299	0.074418
1.56	5.210578452	8.117	27.15	3.1499	0.073835	3.474234
2.20	5.298317367	11.63	28.072	3.1071	0.09893	1.502514
4.84	5.43284826	26.31	29.516	3.0415	0.144525	2.018819
5.00	5.436338664	27.18	29.554	3.0398	0.145822	2.489341
68.43	**93.07222057**	**315.6**	**438.95**	**68.433**	**0.551162**	**27.68093**

4. Скористаємося розрахунками, наведеними у табл. 5.8, щоб побудувати однофакторну лінійну кореляційно-регресійну модель, в якій $z = \ln(e^2)$ – результуюча, а $u = \ln(x)$ – незалежна змінні:

$$\hat{z} = 5{,}692005 - 0{,}49 \cdot u; \quad R^2 = 0{,}02.$$

5. Обчислимо стандартну помилку параметрів вибіркової моделі – коефіцієнта регресії a_1 – S_{a_1}. Для цього:

5.1. Обчислимо незміщену оцінку дисперсії залишків, скориставшись такою залежністю:

$$\sigma_\varepsilon^2 = \frac{\mathrm{Y}^\mathrm{T} \cdot \mathrm{Y} - \hat{\mathrm{A}}^\mathrm{T} \cdot \mathrm{X}^\mathrm{T} \cdot \mathrm{Y}}{n - m}.$$

Розрахуємо елементи числівника: $\mathrm{Y}^\mathrm{T} \cdot \mathrm{Y} = 72970$; $\hat{\mathrm{A}}^\mathrm{T} \cdot \mathrm{X}^\mathrm{T} \cdot \mathrm{Y} = 71960$;

$$\sigma_\varepsilon^2 = \frac{72970 - 71960}{18} = 56{,}111.$$

5.2. Визначимо дисперсію оцінки $\sigma_{a_1}^2$:

$$var(a_1) = \sigma_{a_1}^2 = \sigma_{\varepsilon}^2 \cdot c_{22},$$

де c_{22} – елемент головної діагоналі матриці $(X^T \cdot X)^{-1}$:

$$(X^T \cdot X)^{-1} = \begin{pmatrix} 0{,}277 & -1{,}902 \cdot 10^{-3} \\ -1{,}902 \cdot 10^{-3} & 1{,}594 \cdot 10^{-5} \end{pmatrix},$$

тоді $var(a_1) = 56{,}111 \cdot 0{,}00001594 = 0{,}0008944$.

Отже, стандартна похибка оцінки параметра a_1 визначається як $S_{a_1} = \sqrt{var(a_1)}$ і становить:

$$S_{a_1} = \sqrt{var(a_1)} = \sqrt{0{,}0008944} = 0{,}0299.$$

Перевіримо статистичну значущість параметра a_1 вибіркової кореляційно-регресійної моделі за допомогою t-статистики.

Емпіричне значення t-статистики:

$$t_{a_1} = \frac{|-0{,}49|}{0{,}0299} = 16{,}388.$$

Із таблиць розподілу Стьюдента при заданому рівні значущості α = = 0,05 та кількості ступенів вільності $\nu = 20 - 2 = 18$ знаходимо табличне значення t-статистики $t_{\alpha}^{кр} = 1{,}734$.

Оскільки розраховане емпіричне значення t-статистики для параметра a_1 є більшим за табличне значення (16,388 > 1,734), то з імовірністю p = 1 - 0,05 = 0,95 можна стверджувати про наявність гетероскедастичності у вибірковій сукупності.

5.7 Способи вилучення гетероскедастичності

Сенс вилучення гетероскедастичності, яку зокрема можна встановити за допомогою одного з вищеописаних тестів, полягає у зміні моделі таким чином, щоб відхилення набули постійної дисперсії. Це уможливлює подальше оцінювання параметрів трансформованої моделі за методом найменших квадратів. Трансформація моделі виявляється у зміненні первісної форми моделі. Форма залежності між дисперсією $\sigma_{\varepsilon_i}^2$ та значеннями незалежних змінних визначає спосіб, яким це здійснюється.

На прикладі простої лінійної регресії розглянемо можливі випадки трансформації моделі. Нехай наявна початкова модель $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, яка відповідає всім класичним припущенням лінійної регресії, окрім гетероскедастичності ($D(\varepsilon_i / x_i) = \sigma_i^2 \neq const$) випадкової величини ε_i.

Випадок 1. Припустимо, що гетероскедастичність має форму $E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = k^2 \cdot x^2$ (де k — скінченна константа), тобто дисперсія ε зростає пропорційно x^2. Виражаючи коефіцієнт пропорційності k^2, маємо $k^2 = \dfrac{\sigma_{\varepsilon_i}^2}{x^2}$. Отже трансформація моделі полягає у діленні початкової моделі на $\sqrt{x^2} = x$. Трансформована таким чином модель має вигляд:

$$\frac{y_i}{x_i} = b_o \frac{1}{x_i} + b_1 + \frac{\varepsilon_i}{x_i}. \tag{5.18}$$

Зауважимо, що параметр при змінній $1/x_i$ у трансформованій моделі є перетином у початковій моделі, тоді як перетин трансформованої моделі є нахилом початкової.

Нове змінене значення випадкової величини ε/x_i є гомоскедастичним, оскільки

$$E\left(\frac{\varepsilon_i}{x_i}\right)^2 = \frac{1}{x_i^2} E(\varepsilon_i^2) = \frac{1}{x_i^2} \sigma_{\varepsilon_i}^2$$

(за одним із припущень, X — множина фіксованих значень в усіх вибірках).

Проте у нашому прикладі припускалося, що $\sigma_{\varepsilon_i}^2 = k^2 \cdot x_i^2$, тому $E\left(\dfrac{\varepsilon_i}{x_i}\right)^2 = \dfrac{1}{x_i^2} \cdot k^2 \cdot x_i^2 = k^2$. Отже, нова випадкова змінна має скінченну сталу дисперсію k^2, що уможливлює застосування класичного методу найменших квадратів до оцінювання параметрів трансформованої моделі.

Випадок 2. Припустимо, що гетероскедастичність має форму

$$E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = k^2 \cdot x_i\,. \tag{5.19}$$

Трансформація полягає в діленні початкової моделі на $\sqrt{x}$, тобто трансформована модель набуває вигляду:

$$\frac{y_i}{\sqrt{x_i}} = b_o \frac{1}{\sqrt{x_i}} + b_1 \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}}.$$

Для трансформованої моделі з урахуванням того, що $\sigma_{\varepsilon_i}^2 = k^2 \cdot x_i$, випадкова величина $\varepsilon / \sqrt{x}$ є гомоскедастичною із сталою дисперсією k^2, оскільки

$$E\left(\frac{\varepsilon_i}{\sqrt{x_i}}\right)^2 = \frac{1}{x_i} \cdot E\left(\sigma_{\varepsilon_i}^2\right) = \frac{1}{x_i}\left(\sigma_{\varepsilon_i}^2\right) = k^2.$$

Випадок 3. Припустимо, що форма гетероскедастичності є такою:

$$E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = k^2 (a_0 + a_1 x_i)^2. \tag{5.20}$$

Припустима трансформація полягає в діленні початкової моделі на $\sqrt{(a_0 + a_1 x_i)^2} = a_0 + a_1 x_i$, тобто

$$\frac{y_i}{a_0 + a_1 x_i} = \beta_o \frac{1}{a_0 + a_1 x_i} + \beta_1 \frac{x_i}{a_0 + a_1 x_i} + \frac{\varepsilon_i}{a_0 + a_1 x_i}.$$

Оскільки:

$$E\left(\frac{\varepsilon_i}{a_0 + a_1 x_i}\right)^2 = \frac{1}{(a_0 + a_1 x_i)^2} E(\varepsilon_i^2) = \frac{k^2}{(a_0 + a_1 x_i)^2} (a_0 + a_1 x_i)^2 = k^2,$$

то нова випадкова величина є гомоскедастичною із сталою дисперсією, що дорівнює k^2.

Приклад 5.6. Вилучити гетероскедастичність для моделі, запропонованої у прикладі 5.1

Розв'язання. У прикладі 5.1 було оцінено параметри моделі без урахування наявності гетероскедастичності таким чином:

$$\hat{y} = -14{,}3663 + 0{,}551339x. \tag{5.21}$$

Визначимо коефіцієнт детермінації, розрахувавши відповідні дані, що наведено у табл. 5.9, так:

$$D = R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{20} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{20} (y_i - \bar{y})^2}, \text{ отже, } D = \frac{19071{,}051}{20074{,}411} = 0{,}95.$$

Таблиця 5.9 – Дані для визначення D у моделі (5.21)

№	y_i	x_i	$\hat{y}_i$	$(\hat{y}_i - \overline{y})^2$	$(y_i - \overline{y})^2$
1	15	29	1.623	2480.286	1326.781
2	18	42	8.790	1817.748	1117.231
3	18	44	9.893	1724.939	1117.231
4	27	63	20.368	964.533	596.581
5	40	109	45.730	32.437	130.531
6	36	102	41.870	91.292	237.931
7	42	112	47.384	16.332	88.831
8	38	104	42.973	71.437	180.231
9	31	84	31.946	379.424	417.181
10	40	110	46.281	26.461	130.531
11	41	111	46.832	21.092	108.681
12	40	111	46.832	21.092	130.531
13	38.12	102.4	42.091	87.127	177.023
14	42.16	112.8	47.825	12.962	85.840
15	62.4	148.8	67.673	263.996	120.451
16	69.02	160	73.848	502.789	309.584
17	84.46	183.2	86.639	1240.028	1091.311
18	98.9	200	95.902	1978.162	2253.876
19	123.04	228.8	111.780	3642.738	5128.708
20	124.4	229.6	112.221	3696.175	5325.351
Σ	1028.5	2386.6	1028.5	19071.051	20074.411

Коефіцієнт детермінації для цієї моделі становить $D = R^2 = 0,95$, що свідчить про суттєвий вплив x на y. Зауважимо, що залишкова дисперсія σ_e^2 становить 55,74.

Для вилучення гетероскедастичності у моделі (5.21) зробимо припущення щодо її форми.

1. Розглянемо спочатку ***випадок 1***, коли гетероскедастичність має форму $\sigma_{\varepsilon_i}^2 = k^2 \cdot x_i^2$.

Для цього випадку припустима трансформація початкової моделі є такою:

$$\frac{y_i}{x_i} = b_o \frac{1}{x_i} + b_1 + \frac{\varepsilon_i}{x_i}.$$

Розрахуємо трансформовані змінні $y^* = \dfrac{y_i}{x_i}$ та $x^* = \dfrac{1}{x_i}$ у табл. 5.10.

Таблиця 5.10 – Розрахунок параметрів у першій трансформованій моделі

№	$y_i* = y_i/x_i$	$x_i*=1/x_i$	$\hat{y}_i^*$	$\left(\hat{y}_i^* - \bar{y}\right)^2$	$\left(y_i^* - \bar{y}\right)^2$
1	0.51724	0.0345	0.433579	0.000278	0.010068
2	0.42857	0.0238	0.425931	0.000082	0.000136
3	0.40909	0.0227	0.425155	0.000068	0.000061
4	0.42857	0.0159	0.420243	0.000011	0.000136
5	0.36697	0.0092	0.415443	0.000002	0.002493
6	0.35294	0.0098	0.415894	0.000001	0.004091
7	0.375	0.0089	0.415267	0.000003	0.001756
8	0.36538	0.0096	0.415759	0.000001	0.002654
9	0.36905	0.0119	0.417399	0.000000	0.002290
10	0.36364	0.0091	0.415383	0.000002	0.002837
11	0.36937	0.009	0.415324	0.000002	0.002259
12	0.36036	0.009	0.415324	0.000002	0.003197
13	0.37227	0.0098	0.415867	0.000001	0.001992
14	0.37376	0.0089	0.415221	0.000003	0.001861
15	0.41935	0.0067	0.413684	0.000010	0.000006
16	0.43138	0.0063	0.413347	0.000013	0.000209
17	0.46103	0.0055	0.412780	0.000017	0.001947
18	0.4945	0.005	0.412451	0.000020	0.006021
19	0.53776	0.0044	0.412000	0.000024	0.014607
20	0.54181	0.0044	0.411989	0.000024	0.015602
Σ	**8.338**	**0.224**	**8.338**	**0.000565**	**0.074226**

$$\frac{\hat{y}_i}{x_i} = 0{,}40887 + 0{,}71662 \cdot \frac{1}{x_i}$$

або
$$\hat{y}_i^* = 0{,}40887 + 0{,}71662 \cdot x_i^*. \qquad (5.22)$$

Коефіцієнт детермінації для моделі (5.22) становить $D = R^2 = 0{,}0076$, через те, що індивідуальні спостереження у процесі трансформації набувають ваги. Проте, оскільки за умов гетероскедастичності коефіцієнт детермінації не є інформативним критерієм, то слід скористатися іншим – залишковою дисперсією σ_e^2. Залишкова дисперсія σ_e^2 для моделі (5.22) становить 0,004, що є набагато меншим, ніж у попередній класично побудованій моделі (5.21).

2. Розглянемо ***випадок 2***, коли гетероскедастичність має форму $E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = k^2 \cdot x_i$. Для цього випадку припустима трансформація початкової моделі є такою:

$$\frac{y_i}{\sqrt{x_i}} = b_o \frac{1}{\sqrt{x_i}} + b_1 \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}}.$$

Розрахуємо трансформовані змінні $y* = \frac{y_i}{\sqrt{x_i}}$ та $x* = \frac{1}{\sqrt{x_i}}$ у табл. 5.11.

Таблиця 5.11 – Розрахунок параметрів у другій трансформованій моделі

№	$y_i* = y_i/\sqrt{x_i}$	$x_i*=1/\sqrt{x_i}$	$\hat{y}_i^*$	$\left(\hat{y}_i^* - \bar{y}\right)^2$	$\left(y_i^* - \bar{y}\right)^2$
1	2.78543	0.1857	1.052628	11.726080	2.861290
2	2.77746	0.1543	2.328690	4.615085	2.888316
3	2.7136	0.1508	2.472901	4.016274	3.109448
4	3.40168	0.126	3.479682	0.994573	1.156237
5	3.83131	0.0958	4.707515	0.053153	0.416877
6	3.56453	0.099	4.576131	0.009834	0.832535
7	3.96863	0.0945	4.760014	0.080117	0.258407
8	3.72621	0.0981	4.615020	0.019059	0.563638
9	3.38238	0.1091	4.165810	0.096818	1.198122
10	3.81385	0.0953	4.725253	0.061647	0.439721
11	3.89155	0.0949	4.742751	0.070642	0.342713
12	3.79663	0.0949	4.742751	0.070642	0.462853
13	3.76706	0.0988	4.584000	0.011457	0.503960
14	3.96959	0.0942	4.773659	0.088027	0.257426
15	5.11544	0.082	5.268654	0.626771	0.407652
16	5.45651	0.0791	5.387402	0.828896	0.959509
17	6.24005	0.0739	5.597769	1.256202	3.108483
18	6.99329	0.0707	5.726672	1.561767	6.331872
19	8.13426	0.0661	5.913656	2.064081	13.375843
20	8.20984	0.066	5.918342	2.077567	13.934320
Σ	**89.539**	**2.029**	**89.539299**	**30.328691**	**53.409224**

$$\frac{y_i}{\sqrt{x_i}} = 8{,}60101 \frac{1}{\sqrt{x_i}} + \text{-}40{,}649 \frac{x_i}{\sqrt{x_i}}. \qquad (5.23)$$

Коефіцієнт детермінації для моделі (5.23) становить $D = R^2 = 0{,}57$, а залишкова дисперсія σ_e^2 дорівнює 1,28, що є також набагато меншим, ніж у класично побудованій моделі (5.21).

Отже, моделі (5.22) і (5.23) дозволяють вилучити гетероскедастичність та суттєво знизити залишкову дисперсію, яка в таких випадках є більш інформативним критерієм для порівняння якості моделей, ніж критерій детермінації.

5.8 Сутність мультиколінеарності та її наслідки

Мультиколінеарність – це явище, що виникає під час побудови множинної кореляційно-регресійної моделі у разі наявності лінійного взаємозв'язку між двома (чи більше) факторними ознаками.

У разі функціональної залежності між факторними ознаками виникає явище **досконалої мультиколінеарності**, яке насправді на практиці є достатньо рідкісною подією. Набагато більш поширеним є випадок, коли між факторними ознаками існує не функціональний, а досить щільний кореляційний зв'язок, що називають **недосконалою мультиколінеарністю**[2].

Наявність мультиколінеарності не дозволяє здійснити якісну побудову економетричної моделі, оскільки призводить до зміщення оцінок коефіцієнтів регресії між факторними ознаками, унеможливлюючи визначення чіткого впливу різних факторних на результативну ознаку.

До ***основних причин виникнення мультиколінеарності*** належать:

1. *Одночасне змінення в одному напрямку різних показників.* Воно відбувається, коли на результативну ознаку впливають одночасно однакові (колінеарні) факторні ознаки. Така ситуація пояснюється тим, що під час економічного зростання основні економічні показники здебільшого мають тенденцію також до зростання (із деяким лагом), наприклад, такі макроекономічні показники, як дохід, споживання, накопичення, інвестиції, ціни, зайнятість зростають у періоди економічного піднесення та спадають під час стагнації. Таким чином, якщо дві колінеарні факторні ознаки змінюються в одному напрямку, то дуже складно оцінити частковий вплив кожної з них на результативну ознаку.

2. *Використання лагових значень однієї і тієї самої ознаки під час побудови економетричних моделей.* Наприклад, у споживчих функціях витрати на споживання у попередньому періоді застосовують поряд із величиною споживчих витрат поточного періоду. Природною є неможливість

---

[2] Засновником терміна "мультиколінеарність" вважають Рагнара Фріша. Р. Фріш (1885 — 1973 рр.), норвезький вчений-економетрист, лауреат (разом із Яном Тінбергеном) Нобелівської премії з економіки (1969 р.), один із засновників та президент (1949) Економетричного товариства, член Королівського статистичного товариства в Лондоні та Американської економічної асоціації, засновник власної наукової школи, редактор журналу "Економетрика" у 1933 — 1955 рр.

уникнення принаймні деякого рівня залежності між факторними ознаками, а відповідно і певної мультиколінеарності.

Нехай множинна лінійна кореляційно-регресійна модель набуває вигляду

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon. \qquad (5.24)$$

Функціональна лінійна залежність між факторними ознаками у випадку досконалої мультиколінеарності є такою:

$$x_2 = \gamma_0 + \gamma_1 x_1, \qquad (5.25)$$

де γ_0 і γ_1 – наперед відомі коефіцієнти.

Підставивши вираз (5.25) у (5.24), звівши подібні, отримаємо:

$$y = \beta_0 + \beta_1 x_1 + \beta_2(\gamma_0 + \gamma_1 x_1) + \varepsilon = (\beta_0 + \beta_2\gamma_0) + (\beta_1 + \beta_2\gamma_1)x_1 + \varepsilon. \quad (5.26)$$

З урахуванням позначень: $a_0 = \beta_0 + \beta_2\gamma_0$ і $a_1 = \beta_1 + \beta_2\gamma_1$, отримаємо парну лінійну кореляційно-регресійну модель вигляду (5.27):

$$y = a_0 + a_1 x_1 + \varepsilon. \qquad (5.27)$$

Використовуючи МНК, можна отримати оцінки параметрів α_0 і α_1, враховуючи (5.26) та (5.27), на основі такої системи рівнянь, що містить невідомі b_0, b_1, b_2:

$$\begin{cases} b_0 + b_2\gamma_0 = a_0; \\ b_1 + b_2\gamma_1 = a_1. \end{cases} \qquad (5.28)$$

За умови трьох невідомих і наявності лише двох рівнянь система має безліч розв'язків, що і пояснює той факт, що досконала мультиколінеарність не дозволяє однозначно оцінити параметри множинної лінійної кореляційно-регресійної моделі (5.24). А це, у свою чергу, унеможливлює обґрунтовані статистичні висновки щодо впливу факторних ознак на результуючу змінну, тобто вони є ненадійними та неоднозначними.

За умов мультиколінеарності пошук оцінок невідомих параметрів множинної лінійної регресійної моделі не можна здійснити засобами класичного МНК, оскільки оператор $B = (X^T X)^{-1} X^T Y$ потребує визначення оберненої матриці $(X^T X)^{-1}$, що є неможливим через виродженість матриці $X^T X$ (один із стовпців (рядків) є лінійною комбінацією інших).

За умови щільної кореляції між факторними ознаками наявною є недосконала мультиколінеарність; при її наближенні до ± 1, вона стає майже досконалою.

Мультиколінеарність не знижує якості прогнозних значень економічних показників, оскільки зростання кількості досліджуваних факторів підвищує значення коефіцієнта детермінації, а відповідно і точність прогнозу. Проте на визначення оцінок дійсних значень параметрів регресії мультиколінеарність впливає негативно, оскільки є причиною виникнення великих стандартних похибок в них. Якщо ж значення коефіцієнта детермінації є великим і оцінки параметрів регресії значущими (характеризуються високими *t*-статистиками), то мультиколінеарність не становить проблеми, що є доволі рідкісним явищем.

Слід зазначити, що, за умов виконання усіх відповідних припущень кореляційно-регресійного аналізу, МНК уможливлює отримання найкращих лінійних оцінок параметрів моделі. Це спостерігається навіть тоді, коли деякі коефіцієнти регресії не є статистично значущими.

Наслідки мультиколінеарності. До ***теоретичних наслідків мультиколінеарності*** належать такі:

1. Оцінки параметрів множинної лінійної регресійної моделі за методом найменших квадратів за наявності великої мультиколінеарності залишаються незміщеними;

2. Мультиколінеарність може не впливати на ефективність оцінок параметрів моделі, однак дисперсія оцінки може бути відносно великою порівняно із значенням параметра регресії;

3. Коефіцієнти регресії, що оцінені засобами МНК, їх стандартні похибки стають дуже чутливими до змін вибіркових даних;

4. Ускладнене оцінювання частки кожної з факторних ознак у поясненій дисперсії результуючої ознаки, яку розраховують за множинною лінійною кореляційно-регресійною моделлю за умов мультиколінеарності;

5. Уможливлюється отримання неправильного знака коефіцієнта.

Серед ***практичних наслідків мультиколінеарності*** слід відзначити такі:

1. Під час зростання абсолютної величини коефіцієнта кореляції дисперсії параметрів, обчислених на основі МНК, та модуль коефіцієнта коваріації цих параметрів необмежено зростають;

2. Збільшення коефіцієнта кореляції між факторними ознаками призводить до збільшення значень середньоквадратичних відхилень параметрів регресії, що призводить до збільшення інтервалу довіри для них;

3. Параметри моделі стають незначимими, оскільки за умов мультиколінеарності дисперсія $S^2{}_{b_j}$ нескінченно зростає, а відповідно і значення критерію Стьюдента $t_{b_j}^{ем} = \dfrac{|b_j|}{S_{b_j}}$ наближається до нуля (параметр статистично значущий, якщо емпіричне значення критерію перевищує критичне).

5.9 Тестування наявності мультиколінеарності

Існує ціла низка методів тестування мультиколінеарності, кожен з яких має свої переваги й недоліки. Основними серед них є: визначення коефіцієнта множинної детермінації і статистичної незначущості параметрів моделі; аналіз значень парних коефіцієнтів кореляції для факторних ознак; аналіз значень часткових коефіцієнтів кореляції; алгоритм Феррара – Глобера та ін.

Розглянемо їх більш ретельно на конкретному прикладі.

1. ***Оцінювання значення коефіцієнта множинної детермінації та статистичної незначущості параметрів моделі.*** Велике значення коефіцієнта множинної детермінації та статистична незначущість деяких коефіцієнтів множинної регресії є характерною ознакою мультиколінеарності.

Приклад 5.7. Перевірити наявність мультиколінеарності для множинної лінійної кореляційно-регресійної моделі, що описує залежність між результативною ознакою y_i та сімома незалежними ознаками, статистику за якими наведено у табл. 5.12.

Таблиця 5.12 – Дані для побудови багатофакторної регресійної моделі

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	99	437	27	90	433	100	18	243
2	81	260	12	72	252	87	10	167
3	49	150	11	42	156	49	6	134
4	30	129	59	155	838	222	50	607
5	45	139	72	208	786	295	66	566
6	59	202	8	56	194	69	8	136
7	53	152	11	54	171	55	9	206
8	153	593	28	82	408	118	21	286
9	73	268	16	52	217	66	11	110
10	128	423	35	75	359	99	24	219
11	56	197	14	44	210	54	7	111
12	156	519	29	102	375	129	23	210
13	140	518	36	105	535	144	24	400
14	79	264	20	53	327	68	11	130
15	170	571	51	103	608	135	28	368
16	139	524	29	67	317	92	21	190
17	56	189	19	43	163	57	26	126
18	63	265	11	53	210	65	11	133
19	39	139	6	38	186	49	4	96
20	205	755	51	128	653	165	42	483
21	52	182	9	47	265	55	7	140
22	63	224	18	57	225	71	23	131
23	66	232	24	56	231	67	11	139
24	33	129	7	35	129	42	4	110
25	62	274	11	50	172	63	8	126

Розв'язання.

1. Оцінимо параметри множинної лінійної кореляційно-регресійної моделі засобами МНК.

На базі оператора оцінювання

$$b = \left(X^T X\right)^{-1} \cdot X^T Y$$

отримаємо таку вибіркову кореляційно-регресійну модель:

$$\hat{y} = 3{,}7713 + 0{,}261759x_1 + 0{,}879957x_2 + 0{,}21011x_3 - 0{,}051369\,x_4 -$$
$$- \; 0{,}108728\,x_5 - 0{,}374566\,x_6 - 0{,}005419728\,x_7.$$

2. Визначимо коефіцієнт множинної детермінації D як:

$$D = \frac{b^T X^T Y - n\overline{y}^2}{Y^T Y - n\overline{y}^2} = \frac{55090}{56520} = 0{,}975.$$

Значення коефіцієнта множинної детермінації дорівнює D = 0,975, тобто 97,5 % дисперсії результуючої змінної пояснюють усіма факторними ознаками, які містить модель.

3. Значення t-статистики для кожного коефіцієнта множинної регресії обчислимо за формулою

$$t_{b_j} = \frac{\left|b_j\right|}{S_{b_j}}.$$

Для визначення t_{b_j} спочатку обчислимо стандартну похибку S_{b_j} параметрів вибіркової моделі. Для цього:

3.1. Обчислимо незміщену оцінку дисперсії залишків, скориставшись такою залежністю:

$$\sigma_\varepsilon^2 = \frac{Y^T \cdot Y - B^T \cdot X^T \cdot Y}{n - m}.$$

Тоді $\sigma_\varepsilon^2 = 79{,}281.$

3.2. Визначимо дисперсії $var(b_j)$ оцінок параметрів:

$$var(b_j) = \sigma_{b_j}^2 = \sigma_\varepsilon^2 \cdot c_{jj},$$

де c_{jj} – елемент головної діагоналі матриці $(X^T \cdot X)^{-1}$:

$$
\left(x^{T}x\right)^{-1}=\begin{pmatrix}
0.403 & -3.328\times10^{-4} & 9.63\times10^{-3} & -0.015 & -5.696\times10^{-4} & 7.48\times10^{-3} & -7.952\times10^{-4} & 6.476\times10^{-4} \\
-3.328\times10^{-4} & 2.094\times10^{-6} & -2.596\times10^{-5} & -1.441\times10^{-5} & -1.111\times10^{-6} & 1.491\times10^{-5} & 1.199\times10^{-5} & 4.723\times10^{-7} \\
9.63\times10^{-3} & -2.596\times10^{-5} & 3.209\times10^{-3} & 1.048\times10^{-4} & -1.651\times10^{-4} & -9.133\times10^{-5} & -1.913\times10^{-3} & 5.554\times10^{-5} \\
-0.015 & -1.441\times10^{-5} & 1.048\times10^{-4} & 2.235\times10^{-3} & -3.558\times10^{-6} & -1.529\times10^{-3} & 2.07\times10^{-4} & -2.036\times10^{-5} \\
-5.696\times10^{-4} & -1.111\times10^{-6} & -1.651\times10^{-4} & -3.558\times10^{-6} & 2.918\times10^{-5} & -2.095\times10^{-5} & 1.104\times10^{-4} & -2.092\times10^{-5} \\
7.48\times10^{-3} & 1.491\times10^{-5} & -9.133\times10^{-5} & -1.529\times10^{-3} & -2.095\times10^{-5} & 1.251\times10^{-3} & -4.991\times10^{-4} & -5.814\times10^{-7} \\
-7.952\times10^{-4} & 1.199\times10^{-5} & -1.913\times10^{-3} & 2.07\times10^{-4} & 1.104\times10^{-4} & -4.991\times10^{-4} & 2.812\times10^{-3} & -6.071\times10^{-5} \\
6.476\times10^{-4} & 4.723\times10^{-7} & 5.554\times10^{-5} & -2.036\times10^{-5} & -2.092\times10^{-5} & -5.814\times10^{-7} & -6.071\times10^{-5} & 3.42\times10^{-5}
\end{pmatrix}
$$

Тоді:

$$var(b_0) = 79{,}281 \cdot 0{,}403 = 31{,}950\ ;$$
$$var(b_1) = 79{,}281 \cdot 0{,}000002094 = 0{,}000166\ ;$$
$$var(b_2) = 79{,}281 \cdot 0{,}0032090 = 0{,}2544\ ;$$
$$var(b_3) = 79{,}281 \cdot 0{,}002235 = 0{,}177193\ ;$$
$$var(b_4) = 79{,}281 \cdot 0{,}00002918 = 0{,}00231\ ;$$
$$var(b_5) = 79{,}281 \cdot 0{,}001251 = 0{,}09918\ ;$$
$$var(b_6) = 79{,}281 \cdot 0{,}002812 = 0{,}22294\ ;$$
$$var(b_7) = 79{,}281 \cdot 0{,}0000342 = 0{,}00271\ .$$

Отже, стандартні похибки оцінок параметрів b_j визначаються як $S_{b_j} = \sqrt{var(b_j)}$ і становлять:

$$S_{b_0} = \sqrt{31{,}95} = 5{,}652;\ S_{b_1} = \sqrt{0{,}000166} = 0{,}0129;\ S_{b_2} = \sqrt{0{,}2544} = 0{,}504;$$
$$S_{b_3} = \sqrt{0{,}177193} = 0{,}4209;$$
$$S_{b_4} = \sqrt{0{,}00231} = 0{,}048;\ S_{b_5} = \sqrt{0{,}09918} = 0{,}315;$$
$$S_{b_6} = \sqrt{0{,}22294} = 0{,}472;\ S_{b_7} = \sqrt{0{,}00271} = 0{,}052.$$

3.3. Значення t_{b_j} дорівнює:

- для коефіцієнта множинної регресії b_0: $t_{b_0}^{ем} = 0{,}667$;
- для коефіцієнта множинної регресії b_1 при факторній ознаці x_1: $t_{b_1}^{ем} = 20{,}291$;
- для коефіцієнта множинної регресії b_2 при факторній ознаці x_2 – величині основних засобів $t_{b_2}^{ем} = 1{,}746$;
- для коефіцієнта множинної регресії b_3 при факторній ознаці x_3 – величині основних засобів $t_{b_3}^{ем} = 0{,}499$;

- для коефіцієнта множинної регресії b_4 при факторній ознаці x_4 – величині основних засобів $t_{b_4}^{ем} = 1{,}07$;

- для коефіцієнта множинної регресії b_5 при факторній ознаці x_5 – величині основних засобів $t_{b_5}^{ем} = 0{,}345$;

- для коефіцієнта множинної регресії b_6 при факторній ознаці x_6 – величині основних засобів $t_{b_6}^{ем} = 0{,}794$;

- для коефіцієнта множинної регресії b_7 при факторній ознаці x_7 – величині основних засобів $t_{b_7}^{ем} = 0{,}104$.

Табличне значення t-статистики Стьюдента при заданому рівні значущості $\alpha = 0{,}01$ та кількості ступенів вільності $\nu = 25 - 7 - 1 = 17$ становить $t_{\alpha}^{кр} = 2{,}567$.

Оскільки розраховані значення $t_{b_2}^{ем}$, $t_{b_3}^{ем}$, $t_{b_4}^{ем}$, $t_{b_5}^{ем}$, $t_{b_6}^{ем}$, $t_{b_7}^{ем}$ є меншими за табличне значення (коефіцієнти множинної регресії b_2 – b_7 є статистично незначущими) та коефіцієнт множинної детермінації високий, то з імовірністю $\rho = 1 - 0{,}01 = 0{,}99$ можна стверджувати, що наявна мультиколінеарність.

2. Аналіз значень парних коефіцієнтів кореляції для факторних ознак. Розглянемо інший достатньо відомий тест визначення мультиколінеарності – *аналіз парних коефіцієнтів кореляції*. Слід зауважити, що якщо значення хоча б одного з коефіцієнтів парної кореляції є більшим за 0,8, то мультиколінеарність є великою.

Кореляційна матриця набуває вигляду:

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{pmatrix}, \tag{5.29}$$

де r_{ij} – коефіцієнт кореляції між факторними ознаками x_i та x_j.

Суттєвим недоліком цього тесту є те, що високі значення парних коефіцієнтів кореляції є достатньою, але не необхідною умовою наявності мультиколінеарності. Це пояснюється тим, що інколи кореляції у багатовимірних зв'язках можуть неадекватно ідентифікувати характер зв'язку між двома факторами. Наприклад, між двома факторами може бути великий додатний коефіцієнт кореляції не тому, що один із них стимулює зміну іншого, а тому, що обидва ці фактори змінюються в одному напрямі під впливом інших факторів. У таких ситуаціях потрібно вимірювати справжню щільність лінійного зв'язку між двома факторними ознаками без урахування впливу на них інших факторів.

Навіть за умови відносно невеликих значень парних коефіцієнтів кореляції у більше, ніж двофакторних, моделях може також бути наявною мультиколінеарність. У такому випадку знаходять визначник кореляційної матриці, що набуває значення $|R| \in [0,1]$. При цьому:

1. Якщо $|R| = 0$, то наявна повна мультиколінеарність;
2. Якщо $|R| = 1$, то мультиколінеарність відсутня;
3. Чим ближче $|R|$ прямує до нуля, тим ймовірнішою є мультиколінеарність.

Приклад 5.8. Із використанням аналізу коефіцієнтів парної кореляції перевірити наявність мультиколінеарності у регресійній моделі прикладу 5.7.

Розв'язання.

Спочатку обчислимо кореляційну матрицю R. Для семифакторної моделі прикладу 5.7 вона набуває вигляду:

$$R = \begin{pmatrix} 1 & 0{,}401 & 0{,}286 & 0{,}398 & 0{,}261 & 0{,}257 & 0{,}333 \\ 0{,}401 & 1 & 0{,}934 & 0{,}952 & 0{,}938 & 0{,}943 & 0{,}925 \\ 0{,}286 & 0{,}934 & 1 & 0{,}939 & 0{,}994 & 0{,}932 & 0{,}938 \\ 0{,}398 & 0{,}952 & 0{,}939 & 1 & 0{,}938 & 0{,}887 & 0{,}965 \\ 0{,}261 & 0{,}938 & 0{,}994 & 0{,}938 & 1 & 0{,}942 & 0{,}938 \\ 0{,}257 & 0{,}943 & 0{,}932 & 0{,}887 & 0{,}942 & 1 & 0{,}895 \\ 0{,}333 & 0{,}925 & 0{,}938 & 0{,}965 & 0{,}938 & 0{,}895 & 1 \end{pmatrix}. \tag{5.30}$$

Оскільки у кореляційній матриці усі коефіцієнти, окрім значень парної кореляції для x_1, є більшими за 0,8, то можна стверджувати, що між іншими змінними наявна суттєва мультиколінеарність. Оскільки факторна ознака x_1 порівняно слабко корелює з іншими, то вона має залишитись у кореляційно-регресійній моделі.

Оскільки визначник кореляційної матриці майже дорівнює нулю ($|R| = 0{,}0000002608$), то можна стверджувати, що наявна мультиколінеарність.

3. Аналіз значень часткових коефіцієнтів кореляції. Оскільки у нашому прикладі кількість факторних ознак є великою, то для перевірки на мультиколінеарність доцільно застосовувати часткові коефіцієнти кореляції. Вони визначають зв'язок між результуючою змінною і деякою факторною ознакою за умови фіксованих значень інших факторних ознак. Частковий коефіцієнт кореляції для факторних ознак характеризує тісноту зв'язку між двома факторними ознаками за умови, що всі інші факторні ознаки не впливають на цей зв'язок.

Отже, для визначення часткового коефіцієнта кореляції між першою і другою факторними ознаками (тобто за умови сталості інших (3)–(7) факторних ознак) скористаємося залежністю:

$$R_{12.34567} = -\frac{R_{12}}{\sqrt{R_{11} \cdot R_{22}}}, \tag{5.31}$$

де R_{kj} – алгебраїчне доповнення елемента r_{kj} матриці R вигляду (5.29).

Аналогічно запишемо часткові коефіцієнти для кореляційних залежностей між 1 і 3-ім, 1 і 4-им, 1 і 5-им, 1 і 6-им, 1 і 7-им факторами:

$$R_{12.34567} = -\frac{R_{12}}{\sqrt{R_{11} \cdot R_{22}}};\ R_{14.23567} = -\frac{R_{14}}{\sqrt{R_{11} \cdot R_{44}}};\ R_{15.23467} = -\frac{R_{15}}{\sqrt{R_{11} \cdot R_{55}}};$$
$$R_{16.23457} = -\frac{R_{16}}{\sqrt{R_{11} \cdot R_{66}}};\ R_{17.23456} = -\frac{R_{17}}{\sqrt{R_{11} \cdot R_{77}}} \text{ і т. п.} \tag{5.32}$$

Отже, загальна залежність набуває вигляду:

$$R_{kj.p...m} = -\frac{R_{kj}}{\sqrt{R_{kk} \cdot R_{jj}}}. \tag{5.33}$$

Приклад 5.9. За допомогою аналізу часткових коефіцієнтів кореляції перевірити наявність мультиколінеарності у кореляційно-регресійній моделі прикладу 5.7.

Розв'язання.

Скористаємося залежностями (5.31) – (5.33) для визначення часткових коефіцієнтів кореляції.

Розглянемо приклад визначення часткового коефіцієнта кореляції R_{12}.

Для цього спочатку з'ясуємо відповідні алгебраїчні доповнення (R_{11}, R_{12}, R_{22}):

а) алгебраїчне доповнення елемента r_{kj} визначається як:

$$R_{kj} = (-1)^{(k+j)} M_{kj},$$

де M_{kj} – мінор елемента r_{kj}.

Тоді для елемента r_{12} *алгебраїчне* доповнення R_{12} визначається так:

$$R_{12} = (-1)^{(1+2)} M_{12},$$

$$R_{12} = (-1)^{(1+2)} \cdot \begin{vmatrix} 1 & 0{,}286 & 0{,}398 & 0{,}261 & 0{,}257 & 0{,}333 \\ 0{,}286 & 1 & 0{,}939 & 0{,}994 & 0{,}932 & 0{,}938 \\ 0{,}398 & 0{,}939 & 1 & 0{,}938 & 0{,}887 & 0{,}965 \\ 0{,}261 & 0{,}994 & 0{,}938 & 1 & 0{,}942 & 0{,}938 \\ 0{,}257 & 0{,}932 & 0{,}887 & 0{,}942 & 1 & 0{,}895 \\ 0{,}333 & 0{,}938 & 0{,}965 & 0{,}938 & 0{,}895 & 1 \end{vmatrix} = 0{,}0000005156;$$

$$R_{11} = (-1)^{(1+1)} \cdot \begin{vmatrix} 1 & 0{,}934 & 0{,}952 & 0{,}938 & 0{,}943 & 0{,}925 \\ 0{,}934 & 1 & 0{,}939 & 0{,}994 & 0{,}932 & 0{,}938 \\ 0{,}952 & 0{,}939 & 1 & 0{,}938 & 0{,}887 & 0{,}965 \\ 0{,}938 & 0{,}994 & 0{,}938 & 1 & 0{,}942 & 0{,}938 \\ 0{,}943 & 0{,}932 & 0{,}887 & 0{,}942 & 1 & 0{,}895 \\ 0{,}925 & 0{,}938 & 0{,}965 & 0{,}938 & 0{,}895 & 1 \end{vmatrix} = 0{,}0000004158;$$

$$R_{22} = (-1)^{(2+2)} \cdot \begin{vmatrix} 1 & 0{,}286 & 0{,}398 & 0{,}261 & 0{,}257 & 0{,}333 \\ 0{,}286 & 1 & 0{,}939 & 0{,}994 & 0{,}932 & 0{,}938 \\ 0{,}398 & 0{,}939 & 1 & 0{,}938 & 0{,}887 & 0{,}965 \\ 0{,}261 & 0{,}994 & 0{,}938 & 1 & 0{,}942 & 0{,}938 \\ 0{,}257 & 0{,}932 & 0{,}887 & 0{,}942 & 1 & 0{,}895 \\ 0{,}333 & 0{,}938 & 0{,}965 & 0{,}938 & 0{,}895 & 1 \end{vmatrix} = 0{,}000006347\,;$$

б) на базі розрахованих R_{11}, R_{12}, R_{22} визначимо на основі залежності (5.31) частковий коефіцієнт кореляції $R_{12.34567}$:

$$R_{12.34567} = -\frac{0{,}0000005156}{\sqrt{0{,}0000004158 \cdot 0{,}000006347}} = -0{,}317\,.$$

Аналогічно було отримано решту часткових коефіцієнтів кореляції, які занесені до матриці $R^{част}$.

Оскільки деякі з часткових коефіцієнтів кореляції (особливо ті, що виділені жирним шрифтом) мають досить велике значення, то можна стверджувати, що наявна мультиколінеарність.

$$R^{част} = \begin{array}{c|ccccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ \hline x_1 & 1 & & & & & & \\ x_2 & -0{,}317 & 1 & & & & & \\ x_3 & 0{,}215 & 0{,}061 & 1 & & & & \\ x_4 & -0{,}135 & \mathbf{0{,}548} & -0{,}036 & 1 & & & \\ x_5 & -0{,}295 & -0{,}063 & \mathbf{0{,}916} & -0{,}09 & 1 & & \\ x_6 & -0{,}148 & \mathbf{0{,}6} & -0{,}064 & -0{,}368 & -0{,}236 & 1 & \\ x_7 & -0{,}05 & 0{,}173 & 0{,}06 & -\mathbf{0{,}659} & 0{,}013 & -0{,}19 & 1 \end{array}$$

4. Алгоритм Феррара-Глобера.

Алгоритм Феррара-Глобера базується на трьох статистичних критеріях, що визначають наявність мультиколінеарності:

1) критерій χ^2, на основі якого перевіряють мультиколінеарність усіх факторів моделі;

2) F-критерій, за його допомогою перевіряють гіпотезу H_0:

коефіцієнт детермінації R_j^2 *дорівнює нулю:* $R_j^2 = 0$

та гіпотезу H_1:

коефіцієнт детермінації R_j^2 *не дорівнює нулю:* $R_j^2 \neq 0$,

де R_j^2 – коефіцієнт детермінації в кореляційно-регресійній моделі, яка пов’язує факторну ознаку x_j з усіма іншими факторними ознаками.

Із використанням F-тесту вивчають кореляцію кожної факторної ознаки з усіма іншими;

3) T-критерій, що застосовують для перевірки H_0 гіпотези:

частковий коефіцієнт кореляції дорівнює нулю: $r_{lj.12...k} = 0$

та H_1 гіпотези:

частковий коефіцієнт кореляції не дорівнює нулю: $r_{lj.12...k} \neq 0$,

де $r_{lj.12...k}$ – часткові коефіцієнти кореляції, які описують щільність зв’язку між факторними ознаками x_l та x_j за умови, що решта факторних ознак не впливає на цей зв'язок.

На основі t-тесту визначають наявність лінійної кореляційної залежності кожної пари факторних ознак.

Шляхом порівняння отриманих значень цих критеріїв та їх критичних значень дозволяє зробити висновки щодо мультиколінеарності.

Застосування алгоритму Феррара-Глобера полягає у виконанні таких кроків.

Крок 1. Нормалізувати факторні ознаки x_1, x_2, …, x_k на основі такого перетворення

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{n\sigma_j^2}}, \tag{5.34}$$

де n – кількість спостережень ($i = \overline{1,n}$); k – кількість факторних ознак у моделі ($j = \overline{1,k}$); $\bar{x}_j$ – середнє значення *j*-ї факторної ознаки; σ_j^2 – дисперсія *j*-ї факторної ознаки.

Для нормалізованих значень факторних ознак використовуються умови:

$$E(x_j^*) = 0;\ \sigma_{x_j^*}^2 = 1.$$

Крок 2. Обчислити кореляційну матрицю R (матриця моментів нормалізованої системи нормальних рівнянь):

$$R = X^{*T} X^* = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{pmatrix}, \qquad (5.35)$$

де X^* матриця нормалізованих значень факторних ознак.

Елементами $r_{lj} = r_{x_l x_j}$ матриці *R* є парні коефіцієнти кореляції, які характеризують щільність зв’язку між *l*-м та *j*-м факторами. Проте на основі кореляційної матриці *R* не можна стверджувати, що отриманий зв'язок є явищем мультиколінеарності.

Крок 3. Обчислити значення χ^2-критерію:

$$\chi_{ем}^2 = -\left(n - 1 - \frac{1}{6}(2k+5)\right) \cdot \ln|R|,$$

де $|R|$ – визначник кореляційної матриці *R*.

Знайдемо табличне значення $\chi_{кр}^2$ при $\frac{1}{2}k(k-1)$ ступенях вільності і рівні значущості α. Якщо $\chi_{ем}^2 > \chi_{кр}^2$, то з імовірністю *p* = 1-α можна стверджувати, що в масиві факторних ознак наявна мільтиколінеарність.

Якщо $\chi_{ем}^2 \le \chi_{кр}^2$, то з імовірністю *p* = 1-α можемо зробити висновок щодо відсутності мультиколінеарності.

Крок 4. Визначити матрицю помилок *C*:

$$C = R = \left(X^{*T} X\right)^{-1} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} \end{pmatrix}. \qquad (5.36)$$

Крок 5. Розрахувати значення F-критерію:

$$F_j^{ем} = \frac{(c_{jj} - 1)(n - k)}{k - 1}, \tag{5.37}$$

де c_{jj} – діагональні елементи матриці C.

При заданих ступенях вільності n-k і k-1 та рівні значущості α знайдемо табличне значення критерію $F_\alpha^{кр}$ і порівняємо розраховані значення $F_j^{ем}$ із табличним $F_\alpha^{кр}$.

Якщо $F_j^{ем} > F_\alpha^{кр}$, то з імовірністю $p = 1\text{-}\alpha$ гіпотезу H_0 відкидають, а це означає, що j-та факторна ознака колінеарна з усіма іншими і потрібно вирішити питання про її вилучення з переліку змінних моделей.

За умови, що $F_j^{ем} < F_\alpha^{кр}$, з імовірністю $p=1\text{-}\alpha$ гіпотезу H_0 приймають, тобто факторна ознака x_j не є колінеарною з усіма іншими.

На основі діагональних елементів матриці C розраховують коефіцієнти детермінації для кожної факторної ознаки:

$$R_j^2 = 1 - \frac{1}{c_{jj}}. \tag{5.38}$$

Коефіцієнт детермінації R_j^2 характеризує вплив усіх інших факторних ознак на факторну змінну x_j.

Крок 6. Обчислити часткові коефіцієнти кореляції, що характеризують щільність зв'язку між двома факторними ознаками за умови, що всі інші факторні ознаки не впливають на цей зв'язок (тестування наявності парної колінеарності) таким чином:

$$r_{lj.12...k} = \frac{-c_{lj}}{\sqrt{c_{ll}c_{jj}}}, \tag{5.39}$$

де c_{lj} – елемент матриці C, який розміщений на перетині l*-го* рядка та j-го стовпця ($l = \overline{1,k}$; $j = \overline{1,k}$); c_{ll}, c_{jj} – діагональні елементи матриці C.

Якщо порівняти деякі кількісні значення часткових і парних коефіцієнтів кореляції, то можна побачити, що перші є значно меншими за других. Отже на підставі лише часткових коефіцієнтів кореляції висновок про парну колінеарність зробити неможливо. Для цього потрібно виконати ще сьомий крок.

Крок 7. Розрахунок значень t-критерію:

$$t_{lj}^{ем} = \left| r_{lj.12...k} \right| \cdot \frac{\sqrt{n-k}}{\sqrt{1 - r_{lj.12...k}}}. \tag{5.40}$$

Розраховані значення критерію $t_{lj}^{ем}$ порівнюємо з табличним значенням $t_{\alpha}^{кр}$ при n-k ступенях вільності і рівні значущості α.

Якщо $t_{lj}^{ем} > t_{\alpha}^{кр}$, то з імовірністю $p = 1$-α гіпотезу H_0 вслід відкинути, тобто між факторами x_l і x_j наявна колінеарність.

Якщо $t_{lj}^{ем} \leq t_{\alpha}^{кр}$, то з імовірністю $p = 1$-α гіпотезу H_0 приймаємо, тобто фактори x_l і x_j є неколінеарними.

Аналізуючи значення *F і T*-критеріїв, можна зробити висновок щодо того, який з факторів слід вилучити з кореляційно-регресійної моделі.

Якщо алгоритм Феррара – Глобера не допоміг визначити, яку з факторних ознак потрібно вилучити з переліку змінних моделі, то оцінювати параметри моделі методом найменших квадратів не слід. У такому разі використовують інші методи, наприклад, метод головних компонент або одну з його модифікацій.

Приклад 5.10. Перевірити за допомогою алгоритму Феррара – Глобера наявність мультиколінеарності у кореляційно-регресійній моделі з прикладу 5.7.

Крок 1. Здійснимо нормалізацію факторних ознак x_1, x_2, ..., x_7 за допомогою перетворення (5.34) (табл. 5.13).

Таблиця 5.13 – Нормовані значення факторних ознак

Номер спостереження	x_1^*	x_2^*	x_3^*	x_4^*	x_5^*	x_6^*	x_7^*
1	0.14626412	0.028223	0.076513	0.097728	0.01148	-0.0124	0.02813
2	-0.05662576	-0.14528	-0.01338	-0.08615	-0.03294	-0.12021	-0.07708
3	-0.18271552	-0.15685	-0.16321	-0.18367	-0.16278	-0.17412	-0.12276
4	-0.2067872	0.398366	0.401142	0.509163	0.428327	0.41886	0.532035
5	-0.1953245	0.548736	0.665839	0.456337	0.677752	0.63449	0.475277
6	-0.12310946	-0.19155	-0.09329	-0.14507	-0.09444	-0.14717	-0.12
7	-0.18042298	-0.15685	-0.10328	-0.16843	-0.14227	-0.13369	-0.02309
8	0.32508232	0.03979	0.036558	0.072331	0.072982	0.028032	0.087657
9	-0.0474556	-0.09901	-0.11327	-0.1217	-0.10469	-0.10674	-0.15599
10	0.13021633	0.120759	0.001598	0.022553	0.008064	0.068462	-0.00509
11	-0.12884081	-0.12215	-0.15322	-0.12881	-0.14569	-0.16064	-0.1546
12	0.2402583	0.051357	0.136444	0.038807	0.110567	0.054986	-0.01755
13	0.23911203	0.132326	0.151427	0.201349	0.161819	0.068462	0.245474
14	-0.05204068	-0.05275	-0.10828	-0.00996	-0.09786	-0.10674	-0.1283
15	0.29986437	0.30583	0.141438	0.275509	0.131068	0.12237	0.201174
16	0.24598966	0.051357	-0.03836	-0.02011	-0.01585	0.028032	-0.04524
17	-0.13801097	-0.06431	-0.15822	-0.17656	-0.13544	0.095416	-0.13384
18	-0.05089441	-0.15685	-0.10828	-0.12881	-0.10811	-0.10674	-0.12415

Продовження таблиці 5.13

Номер спостереження	x_1^*	x_2^*	x_3^*	x_4^*	x_5^*	x_6^*	x_7^*
19	-0.1953245	-0.21468	-0.18319	-0.1532	-0.16278	-0.20107	-0.17537
20	0.51077815	0.30583	0.266296	0.321224	0.233571	0.311046	0.360375
21	-0.14603487	-0.17998	-0.13824	-0.07294	-0.14227	-0.16064	-0.11446
22	-0.0978915	-0.07588	-0.0883	-0.11358	-0.08761	0.054986	-0.12692
23	-0.08872134	-0.00648	-0.09329	-0.10748	-0.10127	-0.10674	-0.11584
24	-0.2067872	-0.20312	-0.19817	-0.2111	-0.18669	-0.20107	-0.15599
25	-0.04057798	-0.15685	-0.12326	-0.16742	-0.11494	-0.14717	-0.13384

Крок 2. Обчислимо кореляційну матрицю *R*:

$$R = X_i^{*T} \cdot X_j^* = \begin{pmatrix} 1 & 0,401 & 0,286 & 0,398 & 0,261 & 0,257 & 0,333 \\ 0,401 & 1 & 0,934 & 0,952 & 0,938 & 0,943 & 0,925 \\ 0,286 & 0,934 & 1 & 0,939 & 0,994 & 0,932 & 0,938 \\ 0,398 & 0,952 & 0,939 & 1 & 0,938 & 0,887 & 0,965 \\ 0,261 & 0,938 & 0,994 & 0,938 & 1 & 0,942 & 0,938 \\ 0,257 & 0,943 & 0,932 & 0,887 & 0,942 & 1 & 0,895 \\ 0,333 & 0,925 & 0,938 & 0,965 & 0,938 & 0,895 & 1 \end{pmatrix}.$$

На підставі кореляційної матриці можна стверджувати, що між усіма факторними ознаками (окрім першої) наявний щільний зв’язок (усі парні коефіцієнти $r_{lj} > 0,7$ $(l, j \neq 1)$.

Крок 3. Обчислимо визначник кореляційної матриці:

$$|R| = 0,0000002608.$$

Знайдемо значення χ^2-критерію:

$$\chi^2_{ем} = -\left(n - 1 - \frac{1}{6}(2k+5)\right) \cdot \ln|R|,$$

$$\chi^2_{ем} = -\left(25 - 1 - \frac{1}{6}(2 \cdot 7 + 5)\right) \cdot \ln|0,0000002608| = 315,823.$$

Табличне значення критерію при $\frac{1}{2}k(k-1) = \frac{1}{2}7(7-1) = 21$ ступенях вільності і рівні значущості α = 0,01 дорівнює $\chi^2_{кр} = 38,932$.

Оскільки $\chi^2_{ем} > \chi^2_{кр}$ (315,823 > 38,932), то з імовірністю *p* = 1-0,01 = 0,99 можна стверджувати, що в масиві факторних ознак наявна мультиколінеарність.

Крок 4. Обчислимо матрицю помилок C:

$$C = R^{-1} = (X^{*T} X^{*})^{-1}.$$

$$C = \begin{pmatrix} 1{,}594 & & & & & & \\ -1{,}977 & 24{,}335 & & & & & \\ -2{,}605 & 2{,}865 & 92{,}082 & & & & \\ -0{,}914 & -14{,}441 & -1{,}833 & 28{,}531 & & & \\ 3{,}894 & -3{,}245 & -91{,}994 & -5{,}047 & 109{,}517 & & \\ 0{,}785 & -12{,}455 & 2{,}602 & 8{,}274 & -10{,}378 & 15{,}56 & \\ 0{,}268 & 3{,}588 & -2{,}426 & -14{,}822 & -0{,}574 & -3{,}358 & 17{,}714 \end{pmatrix}$$

Крок 5. Розрахуємо значення F-критерію для кожної факторної ознаки. Наприклад, для факторної ознаки x_1 значення F-критерію дорівнює

$$F_1^{ем} = \frac{(c_{11} - 1)(25 - 7)}{7 - 1} = 3 \cdot (1{,}594 - 1) = 1{,}782.$$

Аналогічно розрахуємо значення F-критеріїв і для інших факторних ознак:

$$F_2^{ем} = 70{,}005; \;\; F_3^{ем} = 273{,}246; \;\; F_4^{ем} = 82{,}593; \;\; F_5^{ем} = 325{,}551;$$

$$F_6^{ем} = 43{,}68; \;\; F_7^{ем} = 50{,}142.$$

Табличне значення критерію при 25-7 = 18 і 7-1= 6 ступенях вільності та рівні значущості α = 0,01 становить $F_{\alpha}^{кр} = 4{,}01$.

Оскільки всі розраховані значення $F_j^{ем}$ окрім першого ($j = 2, \ldots, 7$) є більшими за табличне значення $(F_j^{ем} > F_{\alpha}^{кр})$, то з імовірністю p = 1 - 0,01 = = 0,99 гіпотезу H_0 для кожної факторної ознаки (окрім першої) відкидаємо, тобто кожна з них є колінеарною з усіма іншими.

Розрахуємо коефіцієнти детермінації для кожної множинної моделі, яка описує залежність однієї факторної ознаки від усіх інших:

$$R_1^2 = 1 - \frac{1}{c_{11}} = 1 - \frac{1}{1{,}594} = 0{,}373; \;\; R_2^2 = 1 - \frac{1}{24{,}335} = 0{,}959;$$

$$R_3^2 = 1 - \frac{1}{92{,}082} = 0{,}989; \;\; R_4^2 = 1 - \frac{1}{28{,}531} = 0{,}965;$$

$$R_5^2 = 1 - \frac{1}{109{,}517} = 0{,}991;\ R_6^2 = 1 - \frac{1}{15{,}56} = 0{,}936;$$

$$R_7^2 = 1 - \frac{1}{17{,}714} = 0{,}944.$$

Коефіцієнт детермінації, наприклад $R_2^2 = 0{,}959$, показує, що зміна факторних ознак $x_1, x_3, \ldots, x_7$ пояснює 95,9% дисперсії факторної ознаки x_2.

Крок 6. Обчислимо значення часткових коефіцієнтів кореляції. Наприклад, значення часткового коефіцієнта кореляції між x_1 та x_2 є

$$r_{12.34567} = \frac{1{,}977}{\sqrt{1{,}594 \cdot 24{,}335}} = 0{,}317.$$

Значення часткових коефіцієнтів кореляції між усіма факторними ознаками зобразимо як матрицю $R^{част} = \{r_{lj.12..k}\}$.

$$R^{част} = \left(\begin{array}{c|ccccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ \hline x_1 & 1 & & & & & & \\ x_2 & 0{,}317 & 1 & & & & & \\ x_3 & 0{,}215 & -0{,}061 & 1 & & & & \\ x_4 & 0{,}136 & 0{,}548 & 0{,}036 & 1 & & & \\ x_5 & -0{,}295 & 0{,}063 & 0{,}916 & 0{,}09 & 1 & & \\ x_6 & -0{,}158 & 0{,}64 & -0{,}069 & -0{,}393 & 0{,}251 & 1 & \\ x_7 & -0{,}05 & -0{,}173 & 0{,}06 & 0{,}659 & 0{,}013 & 0{,}202 & 1 \end{array}\right).$$

Крок 7. Розрахуємо значення t-критерію. Наприклад, значення t-статистики для перевіряння значущості часткового коефіцієнта кореляції між x_1 та x_2 дорівнює

$$t_{12}^{ем} = |0{,}317| \cdot \frac{\sqrt{25-7}}{\sqrt{1-0{,}317^2}} = 1{,}418.$$

Значення t-критерію для перевіряння значущості часткових коефіцієнтів кореляції між усіма факторними ознаками зобразимо як матрицю

$$T^{ем}=\left\{t_{lj}^{ем}\right\}=\left(\begin{array}{c|ccccccc}
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\
\hline
x_1 & - & & & & & & \\
x_2 & 1,418 & - & & & & & \\
x_3 & 0,934 & 0,259 & - & & & & \\
x_4 & 0,582 & 2,779 & 0,153 & - & & & \\
x_5 & 1,31 & 0,268 & 9,687 & 0,383 & - & & \\
x_6 & 0,679 & 3,534 & 0,293 & 1,813 & 1,1 & - & \\
x_7 & 0,212 & 0,745 & 0,255 & 3,717 & 0,055 & 0,875 & -
\end{array}\right).$$

Табличне значення критерію при 25 - 7 - 1 = 17 ступенях вільності та рівні значущості $\alpha = 0{,}01$ становить $t_{\alpha}^{кр} = 2{,}567$.

Оскільки, для нашого прикладу, $t_{12}^{ем} < t_{\alpha}^{кр}$ (1,418 < 2,657), то з імовірністю $p = 1 - 0{,}01 = 0{,}99$ гіпотезу H_0 приймаємо, тобто x_1 та x_2 неколінеарні.

За *t*-критерієм можна виділити 6 факторних ознак, які є попарно колінеарними, а саме:

- x_2 та x_4: $t_{24}^{ем} = 2{,}779$;
- x_2 та x_6: $t_{26}^{ем} = 3{,}534$;
- x_3 та x_5: $t_{35}^{ем} = 9{,}687$;
- x_3 та x_6: $t_{36}^{ем} = 2{,}2143$;
- x_4 та x_7: $t_{47}^{ем} = 3{,}717$.

Таким чином, серед семи факторних ознак, які містить кореляційно-регресійна модель, наявна мультиколінеарність, отже, слід повторно специфікувати модель, вилучивши певні факторні ознаки, і знову перевірити наявність мультиколінеарності.

5.10 Методи усунення мультиколінеарності

Вилучення однієї або кількох факторних ознак. Найбільш поширеним способом усунення мультиколінеарності є вилучення з моделі однієї або кількох корельованих факторів, проте в деяких випадках це може призвести до помилок специфікації. Наприклад, під час дослідження попиту на деякий товар факторами можуть бути ціна цього товару і ціни його замінників, що часто пов’язані між собою, тому вилучення з моделі ціни замінників спричинює помилки специфікації, що може призвести до формулювання необґрунтованих висновків на підставі моделі. Наслідки від цього можуть бути гіршими, ніж сама проблема мультиколінеарності, оскільки за умов мультиколінеарності оцінки параметрів залишаються найкращими лінійними оцінками, тоді як вилучення змінної може призвести до зміщених оцінок. Це пояснює досить обережне і зважене ставлення до вилучення факторних змінних, яке виправдовується лише за умови усунення при цьому суттєвих проблем, що може викликати мультиколінеарність.

Збільшення кількості спостережень або побудова нової вибірки. Оскільки мультиколінеарність безпосередньо залежить від вибірки, то можливо, що при іншій вибірці вона буде відсутньою або меншою.

Існують випадки, в яких для зменшення мультиколінеарності достатньо збільшити обсяг вибірки, зокрема, під час використання щорічних даних можна розширити їх за рахунок поквартальних. При цьому зростання вибірки даних зменшує дисперсії коефіцієнтів регресії і збільшує їхню статистичну значущість.

Проте слід зауважити, що отримання нової вибірки або розширення старої не завжди можливе і може бути пов'язане із суттєвими витратами. Окрім цього, зростання розміру вибірки може посилити автокореляцію.

Таким чином, вищевикладені недоліки обмежують можливість використання розглянутого методу.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Поясніть сутність явища автокореляції.
2. Чим характерна автокореляція першого порядку?
3. У чому різниця між автокореляцією першого та другого порядку?
4. Поясніть відмінність між автокореляцією та серійною кореляцією.
5. Що є основними причинами, які уможливлюють появу автокореляції?
6. Назвіть основні наслідки автокореляції.
7. Які Ви знаєте методи тестування автокореляції?
8. Що передбачає графічний метод тестування автокореляції?
9. Сформулюйте критерій Дарбіна-Уотсона для тестування автокореляції першого порядку.
10. Поясніть d-статистику – коефіцієнт Дарбіна-Уотсона (DW). Що є необхідною умовою незалежності випадкових відхилень у ньому?
11. Яких умов слід дотримуватися при використанні критерію Дарбіна-Уотсона?
12. Що таке гомоскедастичність і чому її дотримання є обов'язковою умовою побудови якісної кореляційно-регресійної моделі?
13. У чому полягає явище гетероскедастичності?
14. Назвіть причини виникнення гетероскедастичності. Наведіть приклади.
15. Сформулюйте основні наслідки гетероскедастичності.
16. Які підходи використовують для тестування гетероскедастичності?

17. У чому полягають переваги графічного підходу до тестування гетероскедастичності?
18. Назвіть низку аналітичних методів тестування гетероскедастичності.
19. Визначте методи, що здійснюють тестування гетероскедастичності у довільних неперервних функціях.
20. Сформулюйте методи, що тестують гетероскедастичність у монотонних функціях.
21. Перелічіть методи, що визначають гетероскедастичність у функціях, для яких задано форму залежності та кількість параметрів.
22. У чому полягає суть методу Бартлета?
23. З якою метою та яким чином застосовують тест рангової кореляції Спірмена?
24. Назвіть основні етапи застосування тесту Гольдфельда-Квандта.
25. Визначте механізм тестування гетероскедастичності за допомогою тесту Парка.
26. Опишіть основні випадки вилучення гетероскедастичності.
27. У чому полягає явище мультиколінеарності?
28. Зазначте відмінності між досконалою та недосконалою мультиколінеарністю.
29. Сформулюйте основні причини виникнення мультиколінеарності.
30. Що є теоретичними наслідками мультиколінеарності?
31. Які Вам відомі практичні наслідки мультиколінеарності?
32. Назвіть основні методи тестування мультиколінеарності.
33. Визначте основні етапи тестування мультиколінеарності на основі коефіцієнта множинної детермінації.
34. У чому полягає тест визначення мультиколінеарності на базі аналізу парних коефіцієнтів кореляції?
35. Як відбувається перевірка мультиколінеарності із застосуванням часткових коефіцієнтів кореляції?
36. Сформулюйте основні етапи реалізації алгоритму Феррара-Глобера.
37. Які Вам відомі методи усунення мультиколінераності?
38. Назвіть недоліки існуючих підходів до усунення мультиколінеарності.

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Завдання 5.1. Перевірити кореляційно-регресійну модель, статистичні дані для якої зазначено у табл. 5.14, на наявність автокореляції за допомогою критерію Дарбіна-Уотсона.

Таблиця 5.14 – Статистичні дані для перевірки моделі на автокореляцію

	Варіант 1		Варіант 2		Варіант 3		Варіант 4	
№	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t
1	15	29	16.500	31.900	18.000	34.800	19.500	37.700
2	18	42	19.800	46.200	21.600	50.400	23.400	54.600
3	18	44	19.800	48.400	21.600	52.800	23.400	57.200
4	27	63	29.700	69.300	32.400	75.600	35.100	81.900
5	40	109	44.000	119.900	48.000	130.800	52.000	141.700
6	36	102	39.600	112.200	43.200	122.400	46.800	132.600
7	42	112	46.200	123.200	50.400	134.400	54.600	145.600
8	38	104	41.800	114.400	45.600	124.800	49.400	135.200
9	31	84	34.100	92.400	37.200	100.800	40.300	109.200
10	40	110	44.000	121.000	48.000	132.000	52.000	143.000
11	41	111	45.100	122.100	49.200	133.200	53.300	144.300
12	40	111	44.000	122.100	48.000	133.200	52.000	144.300
13	38.12	102.4	41.932	112.640	45.744	122.880	49.556	133.120
14	42.16	112.8	46.376	124.080	50.592	135.360	54.808	146.640
15	62.4	148.8	68.640	163.680	74.880	178.560	81.120	193.440
16	69.02	160	75.922	176.000	82.824	192.000	89.726	208.000
17	84.46	183.2	92.906	201.520	101.352	219.840	109.798	238.160
18	98.9	200	108.790	220.000	118.680	240.000	128.570	260.000
19	123.04	228.8	135.344	251.680	147.648	274.560	159.952	297.440
20	124.4	229.6	136.840	252.560	149.280	275.520	161.720	298.480

	Варіант 5		Варіант 6		Варіант 7		Варіант 8	
№	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	43,500	22,500	46,400	24,000	49,300	25,500	52,200	27,000
2	63,000	27,000	67,200	28,800	71,400	30,600	75,600	32,400
3	66,000	27,000	70,400	28,800	74,800	30,600	79,200	32,400
4	94,500	40,500	100,800	43,200	107,100	45,900	113,400	48,600
5	163,500	60,000	174,400	64,000	185,300	68,000	196,200	72,000
6	153,000	54,000	163,200	57,600	173,400	61,200	183,600	64,800
7	168,000	63,000	179,200	67,200	190,400	71,400	201,600	75,600
8	156,000	57,000	166,400	60,800	176,800	64,600	187,200	68,400
9	126,000	46,500	134,400	49,600	142,800	52,700	151,200	55,800
10	165,000	60,000	176,000	64,000	187,000	68,000	198,000	72,000
11	166,500	61,500	177,600	65,600	188,700	69,700	199,800	73,800
12	166,500	60,000	177,600	64,000	188,700	68,000	199,800	72,000
13	153,600	57,180	163,840	60,992	174,080	64,804	184,320	68,616
14	169,200	63,240	180,480	67,456	191,760	71,672	203,040	75,888
15	223,200	93,600	238,080	99,840	252,960	106,080	267,840	112,320
16	240,000	103,530	256,000	110,432	272,000	117,334	288,000	124,236
17	274,800	126,690	293,120	135,136	311,440	143,582	329,760	152,028
18	300,000	148,350	320,000	158,240	340,000	168,130	360,000	178,020
19	343,200	184,560	366,080	196,864	388,960	209,168	411,840	221,472
20	344,400	186,600	367,360	199,040	390,320	211,480	413,280	223,920

	Варіант 9		Варіант 10		Варіант 11		Варіант 12	
№	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	55.100	28.500	58.000	30.000	60.900	31.500	63.800	33.000
2	79.800	34.200	84.000	36.000	88.200	37.800	92.400	39.600
3	83.600	34.200	88.000	36.000	92.400	37.800	96.800	39.600
4	119.700	51.300	126.000	54.000	132.300	56.700	138.600	59.400
5	207.100	76.000	218.000	80.000	228.900	84.000	239.800	88.000
6	193.800	68.400	204.000	72.000	214.200	75.600	224.400	79.200
7	212.800	79.800	224.000	84.000	235.200	88.200	246.400	92.400
8	197.600	72.200	208.000	76.000	218.400	79.800	228.800	83.600
9	159.600	58.900	168.000	62.000	176.400	65.100	184.800	68.200
10	209.000	76.000	220.000	80.000	231.000	84.000	242.000	88.000
11	210.900	77.900	222.000	82.000	233.100	86.100	244.200	90.200
12	210.900	76.000	222.000	80.000	233.100	84.000	244.200	88.000
13	194.560	72.428	204.800	76.240	215.040	80.052	225.280	83.864
14	214.320	80.104	225.600	84.320	236.880	88.536	248.160	92.752
15	282.720	118.560	297.600	124.800	312.480	131.040	327.360	137.280
16	304.000	131.138	320.000	138.040	336.000	144.942	352.000	151.844
17	348.080	160.474	366.400	168.920	384.720	177.366	403.040	185.812
18	380.000	187.910	400.000	197.800	420.000	207.690	440.000	217.580
19	434.720	233.776	457.600	246.080	480.480	258.384	503.360	270.688
20	436.240	236.360	459.200	248.800	482.160	261.240	505.120	273.680

	Варіант 13		Варіант 14		Варіант 15		Варіант 16	
№	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	66.700	34.500	69.600	36.000	72.500	37.500	75.400	39.000
2	96.600	41.400	100.800	43.200	105.000	45.000	109.200	46.800
3	101.200	41.400	105.600	43.200	110.000	45.000	114.400	46.800
4	144.900	62.100	151.200	64.800	157.500	67.500	163.800	70.200
5	250.700	92.000	261.600	96.000	272.500	100.000	283.400	104.000
6	234.600	82.800	244.800	86.400	255.000	90.000	265.200	93.600
7	257.600	96.600	268.800	100.800	280.000	105.000	291.200	109.200
8	239.200	87.400	249.600	91.200	260.000	95.000	270.400	98.800
9	193.200	71.300	201.600	74.400	210.000	77.500	218.400	80.600
10	253.000	92.000	264.000	96.000	275.000	100.000	286.000	104.000
11	255.300	94.300	266.400	98.400	277.500	102.500	288.600	106.600
12	255.300	92.000	266.400	96.000	277.500	100.000	288.600	104.000
13	235.520	87.676	245.760	91.488	256.000	95.300	266.240	99.112
14	259.440	96.968	270.720	101.184	282.000	105.400	293.280	109.616
15	342.240	143.520	357.120	149.760	372.000	156.000	386.880	162.240
16	368.000	158.746	384.000	165.648	400.000	172.550	416.000	179.452
17	421.360	194.258	439.680	202.704	458.000	211.150	476.320	219.596
18	460.000	227.470	480.000	237.360	500.000	247.250	520.000	257.140
19	526.240	282.992	549.120	295.296	572.000	307.600	594.880	319.904
20	528.080	286.120	551.040	298.560	574.000	311.000	596.960	323.440

	Варіант 17		Варіант 18		Варіант 19		Варіант 20	
№	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	78.300	40.500	81.200	42.000	84.100	43.500	87.000	45.000
2	113.400	48.600	117.600	50.400	121.800	52.200	126.000	54.000
3	118.800	48.600	123.200	50.400	127.600	52.200	132.000	54.000
4	170.100	72.900	176.400	75.600	182.700	78.300	189.000	81.000
5	294.300	108.000	305.200	112.000	316.100	116.000	327.000	120.000
6	275.400	97.200	285.600	100.800	295.800	104.400	306.000	108.000
7	302.400	113.400	313.600	117.600	324.800	121.800	336.000	126.000
8	280.800	102.600	291.200	106.400	301.600	110.200	312.000	114.000
9	226.800	83.700	235.200	86.800	243.600	89.900	252.000	93.000
10	297.000	108.000	308.000	112.000	319.000	116.000	330.000	120.000
11	299.700	110.700	310.800	114.800	321.900	118.900	333.000	123.000
12	299.700	108.000	310.800	112.000	321.900	116.000	333.000	120.000
13	276.480	102.924	286.720	106.736	296.960	110.548	307.200	114.360
14	304.560	113.832	315.840	118.048	327.120	122.264	338.400	126.480
15	401.760	168.480	416.640	174.720	431.520	180.960	446.400	187.200
16	432.000	186.354	448.000	193.256	464.000	200.158	480.000	207.060
17	494.640	228.042	512.960	236.488	531.280	244.934	549.600	253.380
18	540.000	267.030	560.000	276.920	580.000	286.810	600.000	296.700
19	617.760	332.208	640.640	344.512	663.520	356.816	686.400	369.120
20	619.920	335.880	642.880	348.320	665.840	360.760	688.800	373.200

	Варіант 21		Варіант 22		Варіант 23		Варіант 24		Варіант 25	
№	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	89.90	46.50	92.80	48.00	95.70	49.50	98.60	51.00	101.50	52.50
2	130.20	55.80	134.40	57.60	138.60	59.40	142.80	61.20	147.00	63.00
3	136.40	55.80	140.80	57.60	145.20	59.40	149.60	61.20	154.00	63.00
4	195.30	83.70	201.60	86.40	207.90	89.10	214.20	91.80	220.50	94.50
5	337.90	124.00	348.80	128.00	359.70	132.00	370.60	136.00	381.50	140.00
6	316.20	111.60	326.40	115.20	336.60	118.80	346.80	122.40	357.00	126.00
7	347.20	130.20	358.40	134.40	369.60	138.60	380.80	142.80	392.00	147.00
8	322.40	117.80	332.80	121.60	343.20	125.40	353.60	129.20	364.00	133.00
9	260.40	96.10	268.80	99.20	277.20	102.30	285.60	105.40	294.00	108.50
10	341.00	124.00	352.00	128.00	363.00	132.00	374.00	136.00	385.00	140.00
11	344.10	127.10	355.20	131.20	366.30	135.30	377.40	139.40	388.50	143.50
12	344.10	124.00	355.20	128.00	366.30	132.00	377.40	136.00	388.50	140.00
13	317.44	118.17	327.68	121.98	337.92	125.80	348.16	129.61	358.40	133.42
14	349.68	130.70	360.96	134.91	372.24	139.13	383.52	143.34	394.80	147.56
15	461.28	193.44	476.16	199.68	491.04	205.92	505.92	212.16	520.80	218.40
16	496.00	213.96	512.00	220.86	528.00	227.77	544.00	234.67	560.00	241.57
17	567.92	261.83	586.24	270.27	604.56	278.72	622.88	287.16	641.20	295.61
18	620.00	306.59	640.00	316.48	660.00	326.37	680.00	336.26	700.00	346.15
19	709.28	381.42	732.16	393.73	755.04	406.03	777.92	418.34	800.80	430.64
20	711.76	385.64	734.72	398.08	757.68	410.52	780.64	422.96	803.60	435.40

Завдання 5.2. За допомогою критерію Бартлета здійснити тестування гетероскедастичності для кореляційно-регресійної моделі, побудованої на основі статистичних даних, що надані у завданні 5.1 (табл. 5.14).

Завдання 5.3. Протестувати наявність гетероскедастичності на основі тесту рангової кореляції Спірмена для статистичних даних, що надані у завданні 5.1 (табл. 5.14).

Завдання 5.4. На основі параметричного тесту Гольдфельда-Квандта перевірити наявність гетероскедастичності для кореляційно-регресійної моделі, побудованої на основі статистичних даних, що надані у завданні 5.1 (табл. 5.14).

Завдання 5.5. На основі тесту Парка перевірити наявність гетероскедастичності для статистичних даних, що надані у табл. 5.14.

У разі виникнення гетероскедастичності, здійснити її вилучення.

Завдання 5.6. Перевірити наявність мультиколінеарності для множинної лінійної кореляційно-регресійної моделі на основі коефіцієнта множинної детермінації та статистичної значущості параметрів моделі, дані для побудови якої наведено у табл. 5.15.

Таблиця 5.15 – Дані для перевірки моделі на мультиколінеарність

Варіант 1

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	109	480.7	29.7	99	476.3	110	19.8	267.3
2	89	286	13.2	79.2	277.2	95.7	11	183.7
3	54	165	12.1	46.2	171.6	53.9	6.6	147.4
4	33	141.9	64.9	170.5	921.8	244.2	55	667.7
5	50	152.9	79.2	228.8	864.6	324.5	72.6	622.6
6	65	222.2	8.8	61.6	213.4	75.9	8.8	149.6
7	58	167.2	12.1	59.4	188.1	60.5	9.9	226.6
8	168	652.3	30.8	90.2	448.8	129.8	23.1	314.6
9	80	294.8	17.6	57.2	238.7	72.6	12.1	121
10	141	465.3	38.5	82.5	394.9	108.9	26.4	240.9
11	62	216.7	15.4	48.4	231	59.4	7.7	122.1
12	172	570.9	31.9	112.2	412.5	141.9	25.3	231
13	154	569.8	39.6	115.5	588.5	158.4	26.4	440
14	87	290.4	22	58.3	359.7	74.8	12.1	143
15	187	628.1	56.1	113.3	668.8	148.5	30.8	404.8
16	153	576.4	31.9	73.7	348.7	101.2	23.1	209
17	62	207.9	20.9	47.3	179.3	62.7	28.6	138.6
18	69	291.5	12.1	58.3	231	71.5	12.1	146.3
19	43	152.9	6.6	41.8	204.6	53.9	4.4	105.6
20	226	830.5	56.1	140.8	718.3	181.5	46.2	531.3
21	57	200.2	9.9	51.7	291.5	60.5	7.7	154
22	69	246.4	19.8	62.7	247.5	78.1	25.3	144.1
23	73	255.2	26.4	61.6	254.1	73.7	12.1	152.9
24	36	141.9	7.7	38.5	141.9	46.2	4.4	121
25	68	301.4	12.1	55	189.2	69.3	8.8	138.6

Варіант 2

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	119	524,4	32,4	108	519,6	120	21,6	291,6
2	97	312	14,4	86,4	302,4	104,4	12	200,4
3	59	180	13,2	50,4	187,2	58,8	7,2	160,8
4	36	154,8	70,8	186	1005,6	266,4	60	728,4
5	54	166,8	86,4	249,6	943,2	354	79,2	679,2
6	71	242,4	9,6	67,2	232,8	82,8	9,6	163,2
7	64	182,4	13,2	64,8	205,2	66	10,8	247,2
8	184	711,6	33,6	98,4	489,6	141,6	25,2	343,2
9	88	321,6	19,2	62,4	260,4	79,2	13,2	132
10	154	507,6	42	90	430,8	118,8	28,8	262,8
11	67	236,4	16,8	52,8	252	64,8	8,4	133,2
12	187	622,8	34,8	122,4	450	154,8	27,6	252
13	168	621,6	43,2	126	642	172,8	28,8	480
14	95	316,8	24	63,6	392,4	81,6	13,2	156
15	204	685,2	61,2	123,6	729,6	162	33,6	441,6
16	167	628,8	34,8	80,4	380,4	110,4	25,2	228
17	67	226,8	22,8	51,6	195,6	68,4	31,2	151,2
18	76	318	13,2	63,6	252	78	13,2	159,6
19	47	166,8	7,2	45,6	223,2	58,8	4,8	115,2
20	246	906	61,2	153,6	783,6	198	50,4	579,6
21	62	218,4	10,8	56,4	318	66	8,4	168
22	76	268,8	21,6	68,4	270	85,2	27,6	157,2
23	79	278,4	28,8	67,2	277,2	80,4	13,2	166,8
24	40	154,8	8,4	42	154,8	50,4	4,8	132
25	74	328,8	13,2	60	206,4	75,6	9,6	151,2

Варіант 3

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	129	568.1	35.1	117	562.9	130	23.4	315.9
2	105	338	15.6	93.6	327.6	113.1	13	217.1
3	64	195	14.3	54.6	202.8	63.7	7.8	174.2
4	39	167.7	76.7	201.5	1089.4	288.6	65	789.1
5	59	180.7	93.6	270.4	1021.8	383.5	85.8	735.8
6	77	262.6	10.4	72.8	252.2	89.7	10.4	176.8
7	69	197.6	14.3	70.2	222.3	71.5	11.7	267.8
8	199	770.9	36.4	106.6	530.4	153.4	27.3	371.8
9	95	348.4	20.8	67.6	282.1	85.8	14.3	143
10	166	549.9	45.5	97.5	466.7	128.7	31.2	284.7
11	73	256.1	18.2	57.2	273	70.2	9.1	144.3
12	203	674.7	37.7	132.6	487.5	167.7	29.9	273
13	182	673.4	46.8	136.5	695.5	187.2	31.2	520
14	103	343.2	26	68.9	425.1	88.4	14.3	169
15	221	742.3	66.3	133.9	790.4	175.5	36.4	478.4

Продовження таблиці 5.15, варіант 3

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
16	181	681.2	37.7	87.1	412.1	119.6	27.3	247
17	73	245.7	24.7	55.9	211.9	74.1	33.8	163.8
18	82	344.5	14.3	68.9	273	84.5	14.3	172.9
19	51	180.7	7.8	49.4	241.8	63.7	5.2	124.8
20	267	981.5	66.3	166.4	848.9	214.5	54.6	627.9
21	68	236.6	11.7	61.1	344.5	71.5	9.1	182
22	82	291.2	23.4	74.1	292.5	92.3	29.9	170.3
23	86	301.6	31.2	72.8	300.3	87.1	14.3	180.7
24	43	167.7	9.1	45.5	167.7	54.6	5.2	143
25	81	356.2	14.3	65	223.6	81.9	10.4	163.8

Варіант 4

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	139	611,8	37,8	126	606,2	140	25,2	340,2
2	113	364	16,8	100,8	352,8	121,8	14	233,8
3	69	210	15,4	58,8	218,4	68,6	8,4	187,6
4	42	180,6	82,6	217	1173,2	310,8	70	849,8
5	63	194,6	100,8	291,2	1100,4	413	92,4	792,4
6	83	282,8	11,2	78,4	271,6	96,6	11,2	190,4
7	74	212,8	15,4	75,6	239,4	77	12,6	288,4
8	214	830,2	39,2	114,8	571,2	165,2	29,4	400,4
9	102	375,2	22,4	72,8	303,8	92,4	15,4	154
10	179	592,2	49	105	502,6	138,6	33,6	306,6
11	78	275,8	19,6	61,6	294	75,6	9,8	155,4
12	218	726,6	40,6	142,8	525	180,6	32,2	294
13	196	725,2	50,4	147	749	201,6	33,6	560
14	111	369,6	28	74,2	457,8	95,2	15,4	182
15	238	799,4	71,4	144,2	851,2	189	39,2	515,2
16	195	733,6	40,6	93,8	443,8	128,8	29,4	266
17	78	264,6	26,6	60,2	228,2	79,8	36,4	176,4
18	88	371	15,4	74,2	294	91	15,4	186,2
19	55	194,6	8,4	53,2	260,4	68,6	5,6	134,4
20	287	1057	71,4	179,2	914,2	231	58,8	676,2
21	73	254,8	12,6	65,8	371	77	9,8	196
22	88	313,6	25,2	79,8	315	99,4	32,2	183,4
23	92	324,8	33,6	78,4	323,4	93,8	15,4	194,6
24	46	180,6	9,8	49	180,6	58,8	5,6	154
25	87	383,6	15,4	70	240,8	88,2	11,2	176,4

Варіант 5

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	149	655.5	40.5	135	649.5	150	27	364.5
2	122	390	18	108	378	130.5	15	250.5
3	74	225	16.5	63	234	73.5	9	201
4	45	193.5	88.5	232.5	1257	333	75	910.5
5	68	208.5	108	312	1179	442.5	99	849
6	89	303	12	84	291	103.5	12	204
7	80	228	16.5	81	256.5	82.5	13.5	309
8	230	889.5	42	123	612	177	31.5	429
9	110	402	24	78	325.5	99	16.5	165
10	192	634.5	52.5	112.5	538.5	148.5	36	328.5
11	84	295.5	21	66	315	81	10.5	166.5
12	234	778.5	43.5	153	562.5	193.5	34.5	315
13	210	777	54	157.5	802.5	216	36	600
14	119	396	30	79.5	490.5	102	16.5	195
15	255	856.5	76.5	154.5	912	202.5	42	552
16	209	786	43.5	100.5	475.5	138	31.5	285
17	84	283.5	28.5	64.5	244.5	85.5	39	189
18	95	397.5	16.5	79.5	315	97.5	16.5	199.5
19	59	208.5	9	57	279	73.5	6	144
20	308	1132.5	76.5	192	979.5	247.5	63	724.5
21	78	273	13.5	70.5	397.5	82.5	10.5	210
22	95	336	27	85.5	337.5	106.5	34.5	196.5
23	99	348	36	84	346.5	100.5	16.5	208.5
24	50	193.5	10.5	52.5	193.5	63	6	165
25	93	411	16.5	75	258	94.5	12	189

Варіант 6

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	158	699,2	43,2	144	692,8	160	28,8	388,8
2	130	416	19,2	115,2	403,2	139,2	16	267,2
3	78	240	17,6	67,2	249,6	78,4	9,6	214,4
4	48	206,4	94,4	248	1340,8	355,2	80	971,2
5	72	222,4	115,2	332,8	1257,6	472	105,6	905,6
6	94	323,2	12,8	89,6	310,4	110,4	12,8	217,6
7	85	243,2	17,6	86,4	273,6	88	14,4	329,6
8	245	948,8	44,8	131,2	652,8	188,8	33,6	457,6
9	117	428,8	25,6	83,2	347,2	105,6	17,6	176
10	205	676,8	56	120	574,4	158,4	38,4	350,4
11	90	315,2	22,4	70,4	336	86,4	11,2	177,6
12	250	830,4	46,4	163,2	600	206,4	36,8	336
13	224	828,8	57,6	168	856	230,4	38,4	640
14	126	422,4	32	84,8	523,2	108,8	17,6	208
15	272	913,6	81,6	164,8	972,8	216	44,8	588,8
16	222	838,4	46,4	107,2	507,2	147,2	33,6	304

Продовження таблиці 5.15, варіант 6

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
17	90	302,4	30,4	68,8	260,8	91,2	41,6	201,6
18	101	424	17,6	84,8	336	104	17,6	212,8
19	62	222,4	9,6	60,8	297,6	78,4	6,4	153,6
20	328	1208	81,6	204,8	1044,8	264	67,2	772,8
21	83	291,2	14,4	75,2	424	88	11,2	224
22	101	358,4	28,8	91,2	360	113,6	36,8	209,6
23	106	371,2	38,4	89,6	369,6	107,2	17,6	222,4
24	53	206,4	11,2	56	206,4	67,2	6,4	176
25	99	438,4	17,6	80	275,2	100,8	12,8	201,6

Варіант 7

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	168	742,9	45,9	153	736,1	170	30,6	413,1
2	138	442	20,4	122,4	428,4	147,9	17	283,9
3	83	255	18,7	71,4	265,2	83,3	10,2	227,8
4	51	219,3	100,3	263,5	1424,6	377,4	85	1031,9
5	77	236,3	122,4	353,6	1336,2	501,5	112,2	962,2
6	100	343,4	13,6	95,2	329,8	117,3	13,6	231,2
7	90	258,4	18,7	91,8	290,7	93,5	15,3	350,2
8	260	1008,1	47,6	139,4	693,6	200,6	35,7	486,2
9	124	455,6	27,2	88,4	368,9	112,2	18,7	187
10	218	719,1	59,5	127,5	610,3	168,3	40,8	372,3
11	95	334,9	23,8	74,8	357	91,8	11,9	188,7
12	265	882,3	49,3	173,4	637,5	219,3	39,1	357
13	238	880,6	61,2	178,5	909,5	244,8	40,8	680
14	134	448,8	34	90,1	555,9	115,6	18,7	221
15	289	970,7	86,7	175,1	1033,6	229,5	47,6	625,6
16	236	890,8	49,3	113,9	538,9	156,4	35,7	323
17	95	321,3	32,3	73,1	277,1	96,9	44,2	214,2
18	107	450,5	18,7	90,1	357	110,5	18,7	226,1
19	66	236,3	10,2	64,6	316,2	83,3	6,8	163,2
20	349	1283,5	86,7	217,6	1110,1	280,5	71,4	821,1
21	88	309,4	15,3	79,9	450,5	93,5	11,9	238
22	107	380,8	30,6	96,9	382,5	120,7	39,1	222,7
23	112	394,4	40,8	95,2	392,7	113,9	18,7	236,3
24	56	219,3	11,9	59,5	219,3	71,4	6,8	187
25	105	465,8	18,7	85	292,4	107,1	13,6	214,2

Варіант 8

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	178	786.6	48.6	162	779.4	180	32.4	437.4
2	146	468	21.6	129.6	453.6	156.6	18	300.6
3	88	270	19.8	75.6	280.8	88.2	10.8	241.2
4	54	232.2	106.2	279	1508.4	399.6	90	1092.6
5	81	250.2	129.6	374.4	1414.8	531	118.8	1018.8
6	106	363.6	14.4	100.8	349.2	124.2	14.4	244.8
7	95	273.6	19.8	97.2	307.8	99	16.2	370.8
8	275	1067.4	50.4	147.6	734.4	212.4	37.8	514.8
9	131	482.4	28.8	93.6	390.6	118.8	19.8	198
10	230	761.4	63	135	646.2	178.2	43.2	394.2
11	101	354.6	25.2	79.2	378	97.2	12.6	199.8
12	281	934.2	52.2	183.6	675	232.2	41.4	378
13	252	932.4	64.8	189	963	259.2	43.2	720
14	142	475.2	36	95.4	588.6	122.4	19.8	234
15	306	1027.8	91.8	185.4	1094.4	243	50.4	662.4
16	250	943.2	52.2	120.6	570.6	165.6	37.8	342
17	101	340.2	34.2	77.4	293.4	102.6	46.8	226.8
18	113	477	19.8	95.4	378	117	19.8	239.4
19	70	250.2	10.8	68.4	334.8	88.2	7.2	172.8
20	369	1359	91.8	230.4	1175.4	297	75.6	869.4
21	94	327.6	16.2	84.6	477	99	12.6	252
22	113	403.2	32.4	102.6	405	127.8	41.4	235.8
23	119	417.6	43.2	100.8	415.8	120.6	19.8	250.2
24	59	232.2	12.6	63	232.2	75.6	7.2	198
25	112	493.2	19.8	90	309.6	113.4	14.4	226.8

Варіант 9

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	188	830,3	51,3	171	822,7	190	34,2	461,7
2	154	494	22,8	136,8	478,8	165,3	19	317,3
3	93	285	20,9	79,8	296,4	93,1	11,4	254,6
4	57	245,1	112,1	294,5	1592,2	421,8	95	1153,3
5	86	264,1	136,8	395,2	1493,4	560,5	125,4	1075,4
6	112	383,8	15,2	106,4	368,6	131,1	15,2	258,4
7	101	288,8	20,9	102,6	324,9	104,5	17,1	391,4
8	291	1126,7	53,2	155,8	775,2	224,2	39,9	543,4
9	139	509,2	30,4	98,8	412,3	125,4	20,9	209
10	243	803,7	66,5	142,5	682,1	188,1	45,6	416,1
11	106	374,3	26,6	83,6	399	102,6	13,3	210,9
12	296	986,1	55,1	193,8	712,5	245,1	43,7	399
13	266	984,2	68,4	199,5	1016,5	273,6	45,6	760
14	150	501,6	38	100,7	621,3	129,2	20,9	247
15	323	1084,9	96,9	195,7	1155,2	256,5	53,2	699,2
16	264	995,6	55,1	127,3	602,3	174,8	39,9	361

Продовження таблиці 5.15, варіант 9

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
17	106	359,1	36,1	81,7	309,7	108,3	49,4	239,4
18	120	503,5	20,9	100,7	399	123,5	20,9	252,7
19	74	264,1	11,4	72,2	353,4	93,1	7,6	182,4
20	390	1434,5	96,9	243,2	1240,7	313,5	79,8	917,7
21	99	345,8	17,1	89,3	503,5	104,5	13,3	266
22	120	425,6	34,2	108,3	427,5	134,9	43,7	248,9
23	125	440,8	45,6	106,4	438,9	127,3	20,9	264,1
24	63	245,1	13,3	66,5	245,1	79,8	7,6	209
25	118	520,6	20,9	95	326,8	119,7	15,2	239,4

Варіант 10

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	990	4370	270	900	4330	1000	180	2430
2	810	2600	120	720	2520	870	100	1670
3	490	1500	110	420	1560	490	60	1340
4	300	1290	590	1550	8380	2220	500	6070
5	450	1390	720	2080	7860	2950	660	5660
6	590	2020	80	560	1940	690	80	1360
7	530	1520	110	540	1710	550	90	2060
8	1530	5930	280	820	4080	1180	210	2860
9	730	2680	160	520	2170	660	110	1100
10	1280	4230	350	750	3590	990	240	2190
11	560	1970	140	440	2100	540	70	1110
12	1560	5190	290	1020	3750	1290	230	2100
13	1400	5180	360	1050	5350	1440	240	4000
14	790	2640	200	530	3270	680	110	1300
15	1700	5710	510	1030	6080	1350	280	3680
16	1390	5240	290	670	3170	920	210	1900
17	560	1890	190	430	1630	570	260	1260
18	630	2650	110	530	2100	650	110	1330
19	390	1390	60	380	1860	490	40	960
20	2050	7550	510	1280	6530	1650	420	4830
21	520	1820	90	470	2650	550	70	1400
22	630	2240	180	570	2250	710	230	1310
23	660	2320	240	560	2310	670	110	1390
24	330	1290	70	350	1290	420	40	1100
25	620	2740	110	500	1720	630	80	1260

Варіант 11

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	9,9	43,7	2,7	9	43,3	10	1,8	24,3
2	8,1	26	1,2	7,2	25,2	8,7	1	16,7
3	4,9	15	1,1	4,2	15,6	4,9	0,6	13,4
4	3	12,9	5,9	15,5	83,8	22,2	5	60,7
5	4,5	13,9	7,2	20,8	78,6	29,5	6,6	56,6
6	5,9	20,2	0,8	5,6	19,4	6,9	0,8	13,6
7	5,3	15,2	1,1	5,4	17,1	5,5	0,9	20,6
8	15,3	59,3	2,8	8,2	40,8	11,8	2,1	28,6
9	7,3	26,8	1,6	5,2	21,7	6,6	1,1	11
10	12,8	42,3	3,5	7,5	35,9	9,9	2,4	21,9
11	5,6	19,7	1,4	4,4	21	5,4	0,7	11,1
12	15,6	51,9	2,9	10,2	37,5	12,9	2,3	21
13	14	51,8	3,6	10,5	53,5	14,4	2,4	40
14	7,9	26,4	2	5,3	32,7	6,8	1,1	13
15	17	57,1	5,1	10,3	60,8	13,5	2,8	36,8
16	13,9	52,4	2,9	6,7	31,7	9,2	2,1	19
17	5,6	18,9	1,9	4,3	16,3	5,7	2,6	12,6
18	6,3	26,5	1,1	5,3	21	6,5	1,1	13,3
19	3,9	13,9	0,6	3,8	18,6	4,9	0,4	9,6
20	20,5	75,5	5,1	12,8	65,3	16,5	4,2	48,3
21	5,2	18,2	0,9	4,7	26,5	5,5	0,7	14
22	6,3	22,4	1,8	5,7	22,5	7,1	2,3	13,1
23	6,6	23,2	2,4	5,6	23,1	6,7	1,1	13,9
24	3,3	12,9	0,7	3,5	12,9	4,2	0,4	11
25	6,2	27,4	1,1	5	17,2	6,3	0,8	12,6

Варіант 12

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	19,8	87,4	5,4	18	86,6	20	3,6	48,6
2	16,2	52	2,4	14,4	50,4	17,4	2	33,4
3	9,8	30	2,2	8,4	31,2	9,8	1,2	26,8
4	6	25,8	11,8	31	167,6	44,4	10	121,4
5	9	27,8	14,4	41,6	157,2	59	13,2	113,2
6	11,8	40,4	1,6	11,2	38,8	13,8	1,6	27,2
7	10,6	30,4	2,2	10,8	34,2	11	1,8	41,2
8	30,6	118,6	5,6	16,4	81,6	23,6	4,2	57,2
9	14,6	53,6	3,2	10,4	43,4	13,2	2,2	22
10	25,6	84,6	7	15	71,8	19,8	4,8	43,8
11	11,2	39,4	2,8	8,8	42	10,8	1,4	22,2

Продовження таблиці 5.15, варіант 12

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
12	31,2	103,8	5,8	20,4	75	25,8	4,6	42
13	28	103,6	7,2	21	107	28,8	4,8	80
14	15,8	52,8	4	10,6	65,4	13,6	2,2	26
15	34	114,2	10,2	20,6	121,6	27	5,6	73,6
16	27,8	104,8	5,8	13,4	63,4	18,4	4,2	38
17	11,2	37,8	3,8	8,6	32,6	11,4	5,2	25,2
18	12,6	53	2,2	10,6	42	13	2,2	26,6
19	7,8	27,8	1,2	7,6	37,2	9,8	0,8	19,2
20	41	151	10,2	25,6	130,6	33	8,4	96,6
21	10,4	36,4	1,8	9,4	53	11	1,4	28
22	12,6	44,8	3,6	11,4	45	14,2	4,6	26,2
23	13,2	46,4	4,8	11,2	46,2	13,4	2,2	27,8
24	6,6	25,8	1,4	7	25,8	8,4	0,8	22
25	12,4	54,8	2,2	10	34,4	12,6	1,6	25,2

Варіант 13

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	29.7	131.1	8.1	27	129.9	30	5.4	72.9
2	24.3	78	3.6	21.6	75.6	26.1	3	50.1
3	14.7	45	3.3	12.6	46.8	14.7	1.8	40.2
4	9	38.7	17.7	46.5	251.4	66.6	15	182.1
5	13.5	41.7	21.6	62.4	235.8	88.5	19.8	169.8
6	17.7	60.6	2.4	16.8	58.2	20.7	2.4	40.8
7	15.9	45.6	3.3	16.2	51.3	16.5	2.7	61.8
8	45.9	177.9	8.4	24.6	122.4	35.4	6.3	85.8
9	21.9	80.4	4.8	15.6	65.1	19.8	3.3	33
10	38.4	126.9	10.5	22.5	107.7	29.7	7.2	65.7
11	16.8	59.1	4.2	13.2	63	16.2	2.1	33.3
12	46.8	155.7	8.7	30.6	112.5	38.7	6.9	63
13	42	155.4	10.8	31.5	160.5	43.2	7.2	120
14	23.7	79.2	6	15.9	98.1	20.4	3.3	39
15	51	171.3	15.3	30.9	182.4	40.5	8.4	110.4
16	41.7	157.2	8.7	20.1	95.1	27.6	6.3	57
17	16.8	56.7	5.7	12.9	48.9	17.1	7.8	37.8
18	18.9	79.5	3.3	15.9	63	19.5	3.3	39.9
19	11.7	41.7	1.8	11.4	55.8	14.7	1.2	28.8
20	61.5	226.5	15.3	38.4	195.9	49.5	12.6	144.9
21	15.6	54.6	2.7	14.1	79.5	16.5	2.1	42
22	18.9	67.2	5.4	17.1	67.5	21.3	6.9	39.3
23	19.8	69.6	7.2	16.8	69.3	20.1	3.3	41.7
24	9.9	38.7	2.1	10.5	38.7	12.6	1.2	33
25	18.6	82.2	3.3	15	51.6	18.9	2.4	37.8

Варіант 14

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	39,6	174,8	10,8	36	173,2	40	7,2	97,2
2	32,4	104	4,8	28,8	100,8	34,8	4	66,8
3	19,6	60	4,4	16,8	62,4	19,6	2,4	53,6
4	12	51,6	23,6	62	335,2	88,8	20	242,8
5	18	55,6	28,8	83,2	314,4	118	26,4	226,4
6	23,6	80,8	3,2	22,4	77,6	27,6	3,2	54,4
7	21,2	60,8	4,4	21,6	68,4	22	3,6	82,4
8	61,2	237,2	11,2	32,8	163,2	47,2	8,4	114,4
9	29,2	107,2	6,4	20,8	86,8	26,4	4,4	44
10	51,2	169,2	14	30	143,6	39,6	9,6	87,6
11	22,4	78,8	5,6	17,6	84	21,6	2,8	44,4
12	62,4	207,6	11,6	40,8	150	51,6	9,2	84
13	56	207,2	14,4	42	214	57,6	9,6	160
14	31,6	105,6	8	21,2	130,8	27,2	4,4	52
15	68	228,4	20,4	41,2	243,2	54	11,2	147,2
16	55,6	209,6	11,6	26,8	126,8	36,8	8,4	76
17	22,4	75,6	7,6	17,2	65,2	22,8	10,4	50,4
18	25,2	106	4,4	21,2	84	26	4,4	53,2
19	15,6	55,6	2,4	15,2	74,4	19,6	1,6	38,4
20	82	302	20,4	51,2	261,2	66	16,8	193,2
21	20,8	72,8	3,6	18,8	106	22	2,8	56
22	25,2	89,6	7,2	22,8	90	28,4	9,2	52,4
23	26,4	92,8	9,6	22,4	92,4	26,8	4,4	55,6
24	13,2	51,6	2,8	14	51,6	16,8	1,6	44
25	24,8	109,6	4,4	20	68,8	25,2	3,2	50,4

Варіант 15

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	990	4370	270	900	4330	1000	180	2430
2	810	2600	120	720	2520	870	100	1670
3	490	1500	110	420	1560	490	60	1340
4	300	1290	590	1550	8380	2220	500	6070
5	450	1390	720	2080	7860	2950	660	5660
6	590	2020	80	560	1940	690	80	1360
7	530	1520	110	540	1710	550	90	2060
8	1530	5930	280	820	4080	1180	210	2860
9	730	2680	160	520	2170	660	110	1100
10	1280	4230	350	750	3590	990	240	2190

Продовження таблиці 5.15, варіант 15

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
11	560	1970	140	440	2100	540	70	1110
12	1560	5190	290	1020	3750	1290	230	2100
13	1400	5180	360	1050	5350	1440	240	4000
14	790	2640	200	530	3270	680	110	1300
15	1700	5710	510	1030	6080	1350	280	3680
16	1390	5240	290	670	3170	920	210	1900
17	560	1890	190	430	1630	570	260	1260
18	630	2650	110	530	2100	650	110	1330
19	390	1390	60	380	1860	490	40	960
20	2050	7550	510	1280	6530	1650	420	4830
21	520	1820	90	470	2650	550	70	1400
22	630	2240	180	570	2250	710	230	1310
23	660	2320	240	560	2310	670	110	1390
24	330	1290	70	350	1290	420	40	1100
25	620	2740	110	500	1720	630	80	1260

Варіант 16

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	59,4	262,2	16,2	54	259,8	60	10,8	145,8
2	48,6	156	7,2	43,2	151,2	52,2	6	100,2
3	29,4	90	6,6	25,2	93,6	29,4	3,6	80,4
4	18	77,4	35,4	93	502,8	133,2	30	364,2
5	27	83,4	43,2	124,8	471,6	177	39,6	339,6
6	35,4	121,2	4,8	33,6	116,4	41,4	4,8	81,6
7	31,8	91,2	6,6	32,4	102,6	33	5,4	123,6
8	91,8	355,8	16,8	49,2	244,8	70,8	12,6	171,6
9	43,8	160,8	9,6	31,2	130,2	39,6	6,6	66
10	76,8	253,8	21	45	215,4	59,4	14,4	131,4
11	33,6	118,2	8,4	26,4	126	32,4	4,2	66,6
12	93,6	311,4	17,4	61,2	225	77,4	13,8	126
13	84	310,8	21,6	63	321	86,4	14,4	240
14	47,4	158,4	12	31,8	196,2	40,8	6,6	78
15	102	342,6	30,6	61,8	364,8	81	16,8	220,8
16	83,4	314,4	17,4	40,2	190,2	55,2	12,6	114
17	33,6	113,4	11,4	25,8	97,8	34,2	15,6	75,6
18	37,8	159	6,6	31,8	126	39	6,6	79,8
19	23,4	83,4	3,6	22,8	111,6	29,4	2,4	57,6
20	123	453	30,6	76,8	391,8	99	25,2	289,8
21	31,2	109,2	5,4	28,2	159	33	4,2	84
22	37,8	134,4	10,8	34,2	135	42,6	13,8	78,6
23	39,6	139,2	14,4	33,6	138,6	40,2	6,6	83,4
24	19,8	77,4	4,2	21	77,4	25,2	2,4	66
25	37,2	164,4	6,6	30	103,2	37,8	4,8	75,6

Варіант 17

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	69,3	305,9	18,9	63	303,1	70	12,6	170,1
2	56,7	182	8,4	50,4	176,4	60,9	7	116,9
3	34,3	105	7,7	29,4	109,2	34,3	4,2	93,8
4	21	90,3	41,3	108,5	586,6	155,4	35	424,9
5	31,5	97,3	50,4	145,6	550,2	206,5	46,2	396,2
6	41,3	141,4	5,6	39,2	135,8	48,3	5,6	95,2
7	37,1	106,4	7,7	37,8	119,7	38,5	6,3	144,2
8	107	415,1	19,6	57,4	285,6	82,6	14,7	200,2
9	51,1	187,6	11,2	36,4	151,9	46,2	7,7	77
10	89,6	296,1	24,5	52,5	251,3	69,3	16,8	153,3
11	39,2	137,9	9,8	30,8	147	37,8	4,9	77,7
12	109	363,3	20,3	71,4	262,5	90,3	16,1	147
13	98	362,6	25,2	73,5	374,5	100,8	16,8	280
14	55,3	184,8	14	37,1	228,9	47,6	7,7	91
15	119	399,7	35,7	72,1	425,6	94,5	19,6	257,6
16	97,3	366,8	20,3	46,9	221,9	64,4	14,7	133
17	39,2	132,3	13,3	30,1	114,1	39,9	18,2	88,2
18	44,1	185,5	7,7	37,1	147	45,5	7,7	93,1
19	27,3	97,3	4,2	26,6	130,2	34,3	2,8	67,2
20	144	528,5	35,7	89,6	457,1	115,5	29,4	338,1
21	36,4	127,4	6,3	32,9	185,5	38,5	4,9	98
22	44,1	156,8	12,6	39,9	157,5	49,7	16,1	91,7
23	46,2	162,4	16,8	39,2	161,7	46,9	7,7	97,3
24	23,1	90,3	4,9	24,5	90,3	29,4	2,8	77
25	43,4	191,8	7,7	35	120,4	44,1	5,6	88,2

Варіант 18

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	79.2	349.6	21.6	72	346.4	80	14.4	194.4
2	64.8	208	9.6	57.6	201.6	69.6	8	133.6
3	39.2	120	8.8	33.6	124.8	39.2	4.8	107.2
4	24	103.2	47.2	124	670.4	177.6	40	485.6
5	36	111.2	57.6	166.4	628.8	236	52.8	452.8
6	47.2	161.6	6.4	44.8	155.2	55.2	6.4	108.8
7	42.4	121.6	8.8	43.2	136.8	44	7.2	164.8
8	122	474.4	22.4	65.6	326.4	94.4	16.8	228.8
9	58.4	214.4	12.8	41.6	173.6	52.8	8.8	88
10	102	338.4	28	60	287.2	79.2	19.2	175.2
11	44.8	157.6	11.2	35.2	168	43.2	5.6	88.8
12	125	415.2	23.2	81.6	300	103.2	18.4	168
13	112	414.4	28.8	84	428	115.2	19.2	320
14	63.2	211.2	16	42.4	261.6	54.4	8.8	104

15	136	456.8	40.8	82.4	486.4	108	22.4	294.4

Продовження таблиці 5.15, варіант 18

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
16	111	419.2	23.2	53.6	253.6	73.6	16.8	152
17	44.8	151.2	15.2	34.4	130.4	45.6	20.8	100.8
18	50.4	212	8.8	42.4	168	52	8.8	106.4
19	31.2	111.2	4.8	30.4	148.8	39.2	3.2	76.8
20	164	604	40.8	102.4	522.4	132	33.6	386.4
21	41.6	145.6	7.2	37.6	212	44	5.6	112
22	50.4	179.2	14.4	45.6	180	56.8	18.4	104.8
23	52.8	185.6	19.2	44.8	184.8	53.6	8.8	111.2
24	26.4	103.2	5.6	28	103.2	33.6	3.2	88
25	49.6	219.2	8.8	40	137.6	50.4	6.4	100.8

Варіант 19

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	89,1	393,3	24,3	81	389,7	90	16,2	218,7
2	72,9	234	10,8	64,8	226,8	78,3	9	150,3
3	44,1	135	9,9	37,8	140,4	44,1	5,4	120,6
4	27	116,1	53,1	139,5	754,2	199,8	45	546,3
5	40,5	125,1	64,8	187,2	707,4	265,5	59,4	509,4
6	53,1	181,8	7,2	50,4	174,6	62,1	7,2	122,4
7	47,7	136,8	9,9	48,6	153,9	49,5	8,1	185,4
8	138	533,7	25,2	73,8	367,2	106,2	18,9	257,4
9	65,7	241,2	14,4	46,8	195,3	59,4	9,9	99
10	115	380,7	31,5	67,5	323,1	89,1	21,6	197,1
11	50,4	177,3	12,6	39,6	189	48,6	6,3	99,9
12	140	467,1	26,1	91,8	337,5	116,1	20,7	189
13	126	466,2	32,4	94,5	481,5	129,6	21,6	360
14	71,1	237,6	18	47,7	294,3	61,2	9,9	117
15	153	513,9	45,9	92,7	547,2	121,5	25,2	331,2
16	125	471,6	26,1	60,3	285,3	82,8	18,9	171
17	50,4	170,1	17,1	38,7	146,7	51,3	23,4	113,4
18	56,7	238,5	9,9	47,7	189	58,5	9,9	119,7
19	35,1	125,1	5,4	34,2	167,4	44,1	3,6	86,4
20	185	679,5	45,9	115,2	587,7	148,5	37,8	434,7
21	46,8	163,8	8,1	42,3	238,5	49,5	6,3	126
22	56,7	201,6	16,2	51,3	202,5	63,9	20,7	117,9
23	59,4	208,8	21,6	50,4	207,9	60,3	9,9	125,1
24	29,7	116,1	6,3	31,5	116,1	37,8	3,6	99
25	55,8	246,6	9,9	45	154,8	56,7	7,2	113,4

Варіант 20

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	198	874	54	180	866	200	36	486
2	162	520	24	144	504	174	20	334
3	98	300	22	84	312	98	12	268
4	60	258	118	310	1676	444	100	1214
5	90	278	144	416	1572	590	132	1132
6	118	404	16	112	388	138	16	272
7	106	304	22	108	342	110	18	412
8	306	1186	56	164	816	236	42	572
9	146	536	32	104	434	132	22	220
10	256	846	70	150	718	198	48	438
11	112	394	28	88	420	108	14	222
12	312	1038	58	204	750	258	46	420
13	280	1036	72	210	1070	288	48	800
14	158	528	40	106	654	136	22	260
15	340	1142	102	206	1216	270	56	736
16	278	1048	58	134	634	184	42	380
17	112	378	38	86	326	114	52	252
18	126	530	22	106	420	130	22	266
19	78	278	12	76	372	98	8	192
20	410	1510	102	256	1306	330	84	966
21	104	364	18	94	530	110	14	280
22	126	448	36	114	450	142	46	262
23	132	464	48	112	462	134	22	278
24	66	258	14	70	258	84	8	220
25	124	548	22	100	344	126	16	252

Варіант 21

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	208	917,7	56,7	189	909,3	210	37,8	510,3
2	170	546	25,2	151,2	529,2	182,7	21	350,7
3	103	315	23,1	88,2	327,6	102,9	12,6	281,4
4	63	270,9	123,9	325,5	1759,8	466,2	105	1274,7
5	94,5	291,9	151,2	436,8	1650,6	619,5	138,6	1188,6
6	124	424,2	16,8	117,6	407,4	144,9	16,8	285,6
7	111	319,2	23,1	113,4	359,1	115,5	18,9	432,6
8	321	1245,3	58,8	172,2	856,8	247,8	44,1	600,6
9	153	562,8	33,6	109,2	455,7	138,6	23,1	231
10	269	888,3	73,5	157,5	753,9	207,9	50,4	459,9
11	118	413,7	29,4	92,4	441	113,4	14,7	233,1
12	328	1089,9	60,9	214,2	787,5	270,9	48,3	441
13	294	1087,8	75,6	220,5	1123,5	302,4	50,4	840
14	166	554,4	42	111,3	686,7	142,8	23,1	273
15	357	1199,1	107,1	216,3	1276,8	283,5	58,8	772,8
16	292	1100,4	60,9	140,7	665,7	193,2	44,1	399

Продовження таблиці 5.15, варіант 21

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
17	118	396,9	39,9	90,3	342,3	119,7	54,6	264,6
18	132	556,5	23,1	111,3	441	136,5	23,1	279,3
19	81,9	291,9	12,6	79,8	390,6	102,9	8,4	201,6
20	431	1585,5	107,1	268,8	1371,3	346,5	88,2	1014,3
21	109	382,2	18,9	98,7	556,5	115,5	14,7	294
22	132	470,4	37,8	119,7	472,5	149,1	48,3	275,1
23	139	487,2	50,4	117,6	485,1	140,7	23,1	291,9
24	69,3	270,9	14,7	73,5	270,9	88,2	8,4	231
25	130	575,4	23,1	105	361,2	132,3	16,8	264,6

Варіант 22

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	218	961,4	59,4	198	952,6	220	39,6	534,6
2	178	572	26,4	158,4	554,4	191,4	22	367,4
3	108	330	24,2	92,4	343,2	107,8	13,2	294,8
4	66	283,8	129,8	341	1843,6	488,4	110	1335,4
5	99	305,8	158,4	457,6	1729,2	649	145,2	1245,2
6	130	444,4	17,6	123,2	426,8	151,8	17,6	299,2
7	117	334,4	24,2	118,8	376,2	121	19,8	453,2
8	337	1304,6	61,6	180,4	897,6	259,6	46,2	629,2
9	161	589,6	35,2	114,4	477,4	145,2	24,2	242
10	282	930,6	77	165	789,8	217,8	52,8	481,8
11	123	433,4	30,8	96,8	462	118,8	15,4	244,2
12	343	1141,8	63,8	224,4	825	283,8	50,6	462
13	308	1139,6	79,2	231	1177	316,8	52,8	880
14	174	580,8	44	116,6	719,4	149,6	24,2	286
15	374	1256,2	112,2	226,6	1337,6	297	61,6	809,6
16	306	1152,8	63,8	147,4	697,4	202,4	46,2	418
17	123	415,8	41,8	94,6	358,6	125,4	57,2	277,2
18	139	583	24,2	116,6	462	143	24,2	292,6
19	85,8	305,8	13,2	83,6	409,2	107,8	8,8	211,2
20	451	1661	112,2	281,6	1436,6	363	92,4	1062,6
21	114	400,4	19,8	103,4	583	121	15,4	308
22	139	492,8	39,6	125,4	495	156,2	50,6	288,2
23	145	510,4	52,8	123,2	508,2	147,4	24,2	305,8
24	72,6	283,8	15,4	77	283,8	92,4	8,8	242
25	136	602,8	24,2	110	378,4	138,6	17,6	277,2

Варіант 23

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	228	1005.1	62.1	207	995.9	230	41.4	558.9
2	186	598	27.6	165.6	579.6	200.1	23	384.1
3	113	345	25.3	96.6	358.8	112.7	13.8	308.2
4	69	296.7	135.7	356.5	1927.4	510.6	115	1396.1
5	104	319.7	165.6	478.4	1807.8	678.5	151.8	1301.8
6	136	464.6	18.4	128.8	446.2	158.7	18.4	312.8
7	122	349.6	25.3	124.2	393.3	126.5	20.7	473.8
8	352	1363.9	64.4	188.6	938.4	271.4	48.3	657.8
9	168	616.4	36.8	119.6	499.1	151.8	25.3	253
10	294	972.9	80.5	172.5	825.7	227.7	55.2	503.7
11	129	453.1	32.2	101.2	483	124.2	16.1	255.3
12	359	1193.7	66.7	234.6	862.5	296.7	52.9	483
13	322	1191.4	82.8	241.5	1230.5	331.2	55.2	920
14	182	607.2	46	121.9	752.1	156.4	25.3	299
15	391	1313.3	117.3	236.9	1398.4	310.5	64.4	846.4
16	320	1205.2	66.7	154.1	729.1	211.6	48.3	437
17	129	434.7	43.7	98.9	374.9	131.1	59.8	289.8
18	145	609.5	25.3	121.9	483	149.5	25.3	305.9
19	89.7	319.7	13.8	87.4	427.8	112.7	9.2	220.8
20	472	1736.5	117.3	294.4	1501.9	379.5	96.6	1110.9
21	120	418.6	20.7	108.1	609.5	126.5	16.1	322
22	145	515.2	41.4	131.1	517.5	163.3	52.9	301.3
23	152	533.6	55.2	128.8	531.3	154.1	25.3	319.7
24	75.9	296.7	16.1	80.5	296.7	96.6	9.2	253
25	143	630.2	25.3	115	395.6	144.9	18.4	289.8

Варіант 24

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	238	1048,8	64,8	216	1039,2	240	43,2	583,2
2	194	624	28,8	172,8	604,8	208,8	24	400,8
3	118	360	26,4	100,8	374,4	117,6	14,4	321,6
4	72	309,6	141,6	372	2011,2	532,8	120	1456,8
5	108	333,6	172,8	499,2	1886,4	708	158,4	1358,4
6	142	484,8	19,2	134,4	465,6	165,6	19,2	326,4
7	127	364,8	26,4	129,6	410,4	132	21,6	494,4
8	367	1423,2	67,2	196,8	979,2	283,2	50,4	686,4
9	175	643,2	38,4	124,8	520,8	158,4	26,4	264
10	307	1015,2	84	180	861,6	237,6	57,6	525,6
11	134	472,8	33,6	105,6	504	129,6	16,8	266,4
12	374	1245,6	69,6	244,8	900	309,6	55,2	504
13	336	1243,2	86,4	252	1284	345,6	57,6	960
14	190	633,6	48	127,2	784,8	163,2	26,4	312
15	408	1370,4	122,4	247,2	1459,2	324	67,2	883,2
16	334	1257,6	69,6	160,8	760,8	220,8	50,4	456

Продовження таблиці 5.15, варіант 24

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
17	134	453,6	45,6	103,2	391,2	136,8	62,4	302,4
18	151	636	26,4	127,2	504	156	26,4	319,2
19	93,6	333,6	14,4	91,2	446,4	117,6	9,6	230,4
20	492	1812	122,4	307,2	1567,2	396	100,8	1159,2
21	125	436,8	21,6	112,8	636	132	16,8	336
22	151	537,6	43,2	136,8	540	170,4	55,2	314,4
23	158	556,8	57,6	134,4	554,4	160,8	26,4	333,6
24	79,2	309,6	16,8	84	309,6	100,8	9,6	264
25	149	657,6	26,4	120	412,8	151,2	19,2	302,4

Варіант 25

№	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	248	1092.5	67.5	225	1082.5	250	45	607.5
2	203	650	30	180	630	217.5	25	417.5
3	123	375	27.5	105	390	122.5	15	335
4	75	322.5	147.5	387.5	2095	555	125	1517.5
5	113	347.5	180	520	1965	737.5	165	1415
6	148	505	20	140	485	172.5	20	340
7	133	380	27.5	135	427.5	137.5	22.5	515
8	383	1482.5	70	205	1020	295	52.5	715
9	183	670	40	130	542.5	165	27.5	275
10	320	1057.5	87.5	187.5	897.5	247.5	60	547.5
11	140	492.5	35	110	525	135	17.5	277.5
12	390	1297.5	72.5	255	937.5	322.5	57.5	525
13	350	1295	90	262.5	1337.5	360	60	1000
14	198	660	50	132.5	817.5	170	27.5	325
15	425	1427.5	127.5	257.5	1520	337.5	70	920
16	348	1310	72.5	167.5	792.5	230	52.5	475
17	140	472.5	47.5	107.5	407.5	142.5	65	315
18	158	662.5	27.5	132.5	525	162.5	27.5	332.5
19	97.5	347.5	15	95	465	122.5	10	240
20	513	1887.5	127.5	320	1632.5	412.5	105	1207.5
21	130	455	22.5	117.5	662.5	137.5	17.5	350
22	158	560	45	142.5	562.5	177.5	57.5	327.5
23	165	580	60	140	577.5	167.5	27.5	347.5
24	82.5	322.5	17.5	87.5	322.5	105	10	275
25	155	685	27.5	125	430	157.5	20	315

Завдання 5.7. Із використанням аналізу коефіцієнтів парної кореляції перевірити наявність мультиколінеарності у регресійній моделі, дані для побудови якої надано у табл. 5.15.

Завдання 5.8. За допомогою аналізу часткових коефіцієнтів кореляції перевірити наявність мультиколінеарності у кореляційно-регресійній моделі, дані для побудови якої надано у табл. 5.15.

Завдання 5.9. Перевірити за допомогою алгоритму Феррара – Глобера наявність мультиколінеарності у кореляційно-регресійній моделі, дані для побудови якої надано у табл. 5.15.

На основі проведених розрахунків здійсніть ґрунтовні пропозиції щодо доцільності та способів її усунення.

ТЕСТИ

Оберіть правильну (правильні) відповіді

1. Обов’язковою умовою побудови якісної кореляційно-регресійної моделі є:

а) незалежність будь-якого ε_i значення випадкової величини від значення ε_j;

б) незалежність факторної ознаки від результативної;

в) залежність будь-якого ε_i значення випадкової величини від значення x_i;

г) жодної правильної відповіді.

2. Автокореляція виявляється в тому, що:

а) наявна мультиколінеарність значень факторних ознак;

б) умовна дисперсія випадкових величин не є сталою, тобто значення σ_ε^2 залежать від x;

в) явище кореляції значень результуючої змінної, що виникає унаслідок залежності значень випадкової величини ε_i;

г) жодної правильної відповіді.

3. Різниця між автокореляцією другого та першого порядку полягає у:

а) наявності запізнення в один лаг між значеннями випадкової величини;

б) наявності щільного зв’язку між факторною та результативною ознаками;

в) наявності слабкого зв’язку між факторною та результативною ознаками;

г) жодної правильної відповіді.

4. Серійною кореляцією називають:

а) залежність між значеннями випадкових відхилень двох різних вибірок;

б) залежність між факторною та результативною ознаками;

в) залежність між факторною ознакою та випадковою величиною;

г) жодної правильної відповіді.

5. Основні причини, що уможливлюють появу автокореляції, – це:

а) помилки специфікації;

б) інертність у зміні певних економічних показників;

в) реакція економічних показників на зміну економічних умов із часовим лагом;

г) жодної правильної відповіді.

6. Серед основних наслідків автокореляції виділяють такі:

а) оцінки параметрів, залишаючись лінійними та незміщеними, перестають бути ефективними;

б) дисперсії оцінок параметрів є зміщеними, що призводить до збільшення t-статистик;

в) дисперсії оцінок параметрів завжди залишаються незміщеними, що призводить до зменшення t-статистик;

г) оцінка непоясненої дисперсії є незміщеною оцінкою істинного значення випадкової величини.

7. Для тестування автокореляції застосовують:

а) графічний метод;

б) порівняльний аналіз;

в) критерій Дарбіна-Уотсона;

г) метод найменших квадратів.

8. При застосуванні критерію Дарбіна-Уотсона додатна автокореляція виникає, якщо:

а) $0 \leq DW < d_l$;

б) $d_u \leq DW < 4 - d_u$;

в) $d_l \leq DW < d_u$;

г) $4 - d_u \leq DW < 4 - d_l$.

9. При застосуванні критерію Дарбіна-Уотсона автокореляція відсутня, якщо:

а) $0 \leq DW < d_l$;

б) $d_u \leq DW < 4 - d_u$;

в) $d_l \leq DW < d_u$;

г) $4 - d_u \leq DW < 4 - d_l$.

10. Обов’язковою умовою побудови якісної кореляційно-регресійної моделі є:

а) однакова дисперсія $\sigma_i^2 = \sigma_j^2$ випадкових величин, тобто дисперсія кожної ε_i залишається сталою незалежно від малих чи великих значень фактора;

б) незалежність факторної ознаки від результативної;

в) залежність будь-якого ε_i значення випадкової величини від значення x_i;

г) жодної правильної відповіді.

11. Гетероскедастичність виявляється в тому, що:

а) наявним є щільний взаємозв’язок між різними факторними ознаками;

б) умовна дисперсія випадкових величин не є сталою, тобто значення σ_{ε}^{2} залежать від x;

в) явище кореляції значень результуючої змінної, що виникає унаслідок залежності значень випадкової величини ε_i;

г) жодної правильної відповіді.

12. Причини виникнення гетероскедастичності:

а) ефект масштабу у варіаційних рядах;

б) застосування динамічних рядів;

в) відсутність достатньої кількості статистичних даних;

г) жодної правильної відповіді.

13. До наслідків гетероскедастичності належать:

а) зміщеність оцінок коефіцієнтів моделі;

б) зменшення дисперсії оцінок параметрів;

в) зміщеність дисперсії оцінок параметрів регресії;

г) висновки, отримані на підставі відповідних t- і F-статистик, а також інтервальні оцінки будуть ненадійними.

14. До методів тестування гетероскедастичності належать:

а) графічний аналіз випадкових відхилень;

б) метод найменших квадратів;

в) критерій Фішера;

г) жодної правильної відповіді.

15. Серед аналітичних методів тестування гетероскедастичності слід виділити такі:

а) графічний;

б) методи, що здійснюють тестування гетероскедастичності у довільних неперервних, монотонних функціях;

в) методи, що здійснюють тестування гетероскедастичності у певних функціях;

г) жодної правильної відповіді.

16. За критерієм Бартлетта гетероскедастичність наявна, якщо:

а) $b^c \geq \chi_{кр}^2$;

б) $b^c < \chi_{кр}^2$;

в) $t < F_{кр}$;

г) жодної правильної відповіді.

17. Під час використання тесту рангової кореляції Спірмена припускають:

а) дисперсія випадкових відхилень ε монотонно збільшується або зменшується зі зростанням значень незалежної змінної x;

б) дисперсія випадкових відхилень ε монотонно зменшується зі зростанням значень незалежної змінної x;

в) дисперсія результативної ознаки суттєво впливає на дисперсію значень незалежної змінної x;

г) жодної правильної відповіді.

18. За критерієм Спірмена гетероскедастичність наявна, якщо:

а) $t \geq t_{\alpha}^{кр}$;

б) $t < t_{\alpha}^{кр}$;

в) $t < F_{кр}$;

г) жодної правильної відповіді.

19. Під час використання тесту Гольдфельда-Квандта розглядається випадок, коли:

а) дисперсія залишків зростає пропорційно квадрату незалежної змінної;

б) дисперсія залишків зростає обернено пропорційно незалежній змінній;

в) дисперсія залишків зростає пропорційно кубу незалежної змінної;

г) жодної правильної відповіді.

20. За критерієм Гольдфельда-Квандта гетероскедастичність наявна, якщо:

а) $R^* \leq F_{табл., v_1, v_2}$;

б) $R^* > F_{табл., v_1, v_2}$;

в) $t < t_{\alpha}^{кр}$;

в) $t < F_{кр}$.

21. За критерієм Парка гетероскедастичність наявна, якщо:

а) $t < t_{\alpha}^{кр}$;

б) $t > t_{\alpha}^{кр}$;

в) $t < F_{кр}$;

г) $R^* > F_{табл., v_1, v_2}$.

22. Для усунення гетероскедастичності у моделі $E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = k^2 \cdot x^2$ слід:

а) поділити модель на x;

б) помножити модель на $\sqrt{x}$;

в) поділити модель на y;

г) жодної правильної відповіді.

23. Для усунення гетероскедастичності у моделі $E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = k^2 \cdot x_i$ слід:

а) поділити модель на $\sqrt{x}$;

б) помножити модель на $\sqrt{x}$;

в) поділити модель на y;

г) жодної правильної відповіді.

24. Для усунення гетероскедастичності у моделі $E(\varepsilon_i)^2 = \sigma_{\varepsilon_i}^2 = = k^2(a_0 + a_1 x_i)^2$ слід:

а) поділити модель на $\sqrt{(a_0 + a_1 x_i)^2} = a_0 + a_1 x_i$;

б) помножити модель на $\sqrt{x}$;

в) помножити модель на y;

г) жодної правильної відповіді.

25. Мультиколінеарність – це явище, що виникає:

а) у разі наявності лінійного взаємозв'язку між двома (чи більше) факторними ознаками;

б) за умови наявності щільного взаємозв’язку між різними факторними ознаками;

б) за умови дисперсії випадкових величин, що не є сталою;

г) жодної правильної відповіді.

26. Наслідками мультиколінеарності є:

а) зміщення оцінок коефіцієнтів регресії;

б) відсутність можливості визначення чіткого впливу різних факторних ознак на результативну;

в) незміщеність оцінок коефіцієнтів регресії;

г) жодної правильної відповіді.

27. Серед основних причин виникнення мультиколінеарності слід зазначити такі:

а) одночасне змінення в одному напрямку різних показників;

б) використання лагових значень однієї і тієї самої ознаки під час побудови економетричних моделей;

в) значна залежність між результативною та факторною ознаками;

г) жодної правильної відповіді.

28. Серед теоретичних наслідків мультиколінеарності слід зазначити такі:

а) оцінки параметрів множинної лінійної регресійної моделі за МНК залишаються незмінними;

б) дисперсія оцінки може бути великою порівняно із значенням параметра;

в) коефіцієнти регресії, що оцінені за МНК, не є чутливими до змін вибіркових даних;

г) стандартні помилки стають нечутливими до змін вибіркових даних.

29. Серед практичних наслідків мультиколінеарності слід зазначити такі:

а) одночасне зростання коефіцієнта кореляції дисперсії параметрів, обчислених на основі МНК, та модуля коефіцієнта коваріації цих параметрів;

б) збільшення коефіцієнта кореляції між факторними ознаками призводить до збільшення значень середньоквадратичних відхилень параметрів регресії, що призводить до збільшення інтервалу довіри для них;

в) параметри моделі є значимими;

г) жодної правильної відповіді.

30. Оберіть серед наведених нижче методи тестування мультиколінеарності:

а) визначення коефіцієнта множинної детермінації та аналіз значень парних коефіцієнтів кореляції для факторних ознак;

б) алгоритм Феррара – Глобера;

в) метод найменших квадратів;

г) критерій Гурвіца.

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Економетрія. Частина 1 : [навчальний посібник] / А. О. Азарова, Н. В. Сачанюк-Кавецька, О. М. Роїк, Ю. В. Міронова. – Вінниця : ВНТУ, 2011. – 108 с.
2. Економетрія. Частина 2 : [навчальний посібник] / А. О. Азарова, Н. В. Сачанюк-Кавецька, О. М. Роїк, Ю. В. Міронова. – Вінниця : ВНТУ, 2011. – 118 с.
3. Айвазян С. А. Прикладная статистика. Основы эконометрики : в 2-х т. : учебник для вузов / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ-ДАНА, 2001. – Т. 1 : Теория вероятностей и прикладная статистика. – 2001. – 656 с.
4. Айвазян С. А. Прикладная статистика. Основы эконометрики : в 2-х т. : учебник для вузов / Айвазян С. А. – М. : ЮНИТИ-ДАНА, 2006. – Т. 2 : Основы эконометрики. – 2006. – 432 с.
5. Бабешко Л. О. Основы эконометрического моделирования : учеб. пособие / Бабешко Л. О. – [2-е изд., исправленное]. − М. : КомКнига, 2006. – 432 с.
6. Берндт Э. Практика эконометрики : классика и современность / Берндт Э. – М. : Юнити-Дана, 2005. – 848 с.
7. Валландер С. С. Заметки по эконометрике. Часть 1 : учебное пособие / Валландер С. С. − СПб. : Изд. Европ. ун-та в С.-Петербурге, 2002. − 46 с.
8. Гладилин А. В. Эконометрика : учебное пособие / А. В. Гладилин, А. Н. Герасимов, Е. И. Громов. – М. : КНОРУС, 2008. – 232 с.
9. Грубер И. Эконометрия : в 2-х т. : учебное пособие для студентов економических специальностей / Грубер И. − К., 1996. − Т. 1 : Введение в эконометрию. 1996. − 397 с.
10. Данко П. Е. Высшая математика в упражнениях и задачах / П. Е. Данко, А. Г. Попов − М. : Высшая школа, 1974. – 280 с.
11. Дорохина Е. Ю. Сборник задач по эконометрике : учебное пособие для студентов экономических вузов / Дорохина Е. Ю., Преснякова Л. Ф., Тихомиров Н. П. − М. : Издательство “Экзамен”, 2003. − 224 с.
12. Доугерти К. Введение в эконометрику / Доугерти К. : пер. с англ. – М. : ИНФРА-М, 2009. – 402 с.
13. Елисеева И. И. Практикум по эконометрике : учеб. пособие / Елисеева И. И. – М. : Финансы и статистика, 2009. – 344 с.
14. Елисеева И. И. Эконометрика : учебник / Елисеева И. И., Курышева С. В., Костеева Т. В. и др. ; под ред. И. И. Елисеевой. – [2-е изд.]. – М. : Финансы и статистика, 2006. − 576 с.

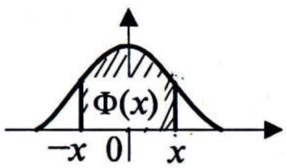
15. Елисеева И. И. Эконометрика : учебник / Елисеева И. И., Курышева С. В., Костеева Т. В. и др. ; под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2009. – 576 с.
16. Замков О. О. Математические методы в экономике / Замков О. О., Толстопятенко А. В., Черемных Ю. Н. − М. : ДИС, 1998. – 385 с.
17. Здрок В. В. Економетрія : підручник / В. В. Здрок, Т. Я. Лагоцький. – К. : Знання, 2010 – 541 с.
18. Катышев П. К. Сборник задач к начальному курсу эконометрики / Катышев П. К., Магнус Я. Р., Пересецкий А. А. – М. : Дело, 2002. – 208 с.
19. Кремер Н. Ш. Эконометрика / Н. Ш. Кремер, Б. А. Путко. – М. : Юнити, 2004. – 311 с.
20. Кремер Н. Ш. Эконометрика / Н. Ш. Кремер, Б. А. Путко. – М. : Юнити-Дана, 2003-2004. – 311 с.
21. Кремер Н. Ш. Эконометрика : учебник для вузов / Н. Ш. Кремер, Б. А. Путко; под ред. проф. Н. Ш. Кремера. – М. : ЮНИТИ-ДАНА, 2008. – 311 с.
22. Кулинич Е. И. Эконометрия / Кулинич Е. И. – М. : Финансы и статистика, 2001. – 304 с.
23. Леонтьев В. В. Экономические эссе. Теория, исследования, факты и политика / Леонтьев В. В. ; пер. с англ. – М. : Политиздат, 2008. – 324 с.
24. Лотов А. В. Введение в экономико-математическое моделирование / Лотов А. В. − М. : Наука, 1984. − 392 с.
25. Луговская Л. В. Эконометрика в вопросах и ответах : учебное пособие / Луговская Л. В. – М. : ТК Велби, Изд-во Проспект, 2006. – 208 с.
26. Лук’яненко І. Економетрика : підручник / І. Лук’яненко, Л. Краснікова. – К. : Товариство “Знання”, КОО, 1998. – 494 с.
27. Магнус Я. Р. Эконометрика. Начальный курс / Магнус Я. Р., Катышев П. К., Пересецкий А. А. – М. : Дело, 2007. – 504 с.
28. Магнус Я. Р. Эконометрика. Начальный курс : учебник / Магнус Я. Р., Катышев П. К., Пересецкий А. А. – М. : Дело, 2007. – 400 с.
29. Магнус Я. Р. Эконометрика. Начальный курс / Магнус Я. Р., Катышев П. К., Пересецкий А. А. – [5-е изд., исправленное]. – М. : Дело, 2006. – 400 с.
30. Мардас А. Н. Эконометрика / Мардас А. Н. – СПб. : Питер, 2005. – 144 с.
31. Наконечний С. І. Економетрія / Наконечний С. І., Терещенко Т. О., Романюк Т. П. – К. : КНЕУ, 2006. – 528 с.
32. Орлов А. И. Эконометрика : учеб. пособие для вузов / Орлов А. И. – М. : Издательство “Экзамен”, 2002. – 576 с.

33. Приходько А. И. Практикум по эконометрике : регрессионный анализ средствами Excel / Приходько А. И. – Ростов н/Д. : Феникс, 2007. – 256 с.
34. Суслов В. И. Эконометрия / [Суслов В. И., Ибрагимов Н. М., Талышева Л. П., Цыплаков А. А.]. – Новосибирск : СО РАН, 2005. – 744 с.
35. Сытник В. Ф. Математические модели в планировании и управлении предприятиями / В. Ф. Сытник, Е. А. Карагодова. − Киев : Выща школа, 1985. − 214 с.
36. Терехов Л. Л. Экономико-математические методы / Терехов Л. Л. − М. : Статистика, 1972. – 250 с.
37. Тихомиров Н. П. Эконометрика : учебник / Тихомиров Н. П., Дорохина Е. Ю. – М. : Издательство “Экзамен”, 2005. – 512 с.
38. Толбатов Ю. А. Економетрика : підручник / Толбатов Ю. А. – Л. : Четверта хвиля, 1997. − 362 с.
39. Тутубалин В. Н. Границы применимости (вероятностно-статистические методы и их возможности) / Тутубалин В. Н.– М. : Знание, 2002. – 64 с.
40. Шалабанов А. К. Практикум по эконометрике с применение MS Excel / А. К. Шалабанов, Д. А. Роганов. – Казань : Издательский центр Академии управления “ТИСБИ”, 2008 – 53 с.
41. Шалабанов А. К. Эконометрика : учебно-методическое пособие / А. К. Шалабанов, Д. А. Роганов. – Казань : Издательский центр Академии управления “ТИСБИ”, 2008. – 198 с.
42. Яновский Л. П. Введение в эконометрику : учебное пособие / Л. П. Яновский, А. Г. Буховец. – М. : КНОРУС, 2009. – 256 с.

Д О Д А Т К И

ДОДАТОК А

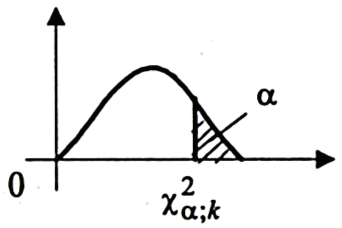
Значення функції Лапласа $\Phi(x) = \dfrac{2}{\sqrt{2\pi}} \int\limits_0^x e^{-t^2/2}\, dt$



Цілі та десяткові частки x	Соті частки x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1663	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,5	0,3829	0,3899	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6679	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7984	0,8029
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9392	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9533
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9331	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9951	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,2	0,9986	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,3	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,4	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,5	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997
3,6	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998
3,7	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

ДОДАТОК Б

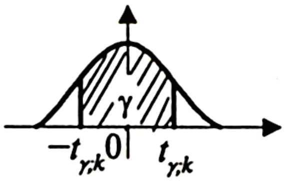
Значення χ^2 критерію Пірсона



Число ступенів вільності, k	Ймовірність, α												
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,00	0,00	0,00	0,02	0,06	0,15	0,45	1,07	1,64	2,71	3,84	5,41	6,64
2	0,02	0,04	0,10	0,21	0,45	0,71	1,39	2,41	3,22	4,60	5,99	7,82	9,21
3	0,11	0,18	0,35	0,58	1,00	1,42	2,37	3,66	4,64	6,25	7,82	9,84	11,3
4	0,30	0,43	0,71	1,06	1,65	2,20	3,36	4,88	5,99	7,78	9,49	11,7	13,3
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,06	7,29	9,24	11,1	13,4	15,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,23	8,56	10,6	12,6	15,0	16,8
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,0	14,1	16,6	18,5
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,0	13,4	15,5	18,2	20,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2
11	3,05	3,61	4,58	5,58	6,99	8,15	10,3	12,9	14,6	17,3	19,7	22,6	24,7
12	3,57	4,18	5,23	6,30	7,81	9,03	11,3	14,0	15,8	18,5	21,0	24,1	26,2
13	4,11	4,76	5,89	7,04	8,63	9,93	12,3	15,1	17,0	19,8	22,4	25,5	27,7
14	4,66	5,37	6,57	7,79	9,47	10,8	13,3	16,2	18,1	21,1	23,7	26,9	29,1
15	5,23	5,98	7,26	8,55	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6
16	5,81	6,61	7,96	9,31	11,1	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0
17	6,41	7,26	8,67	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4
18	7,02	7,91	9,39	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8
19	7,63	8,57	10,1	11,6	13,7	15,3	18,3	21,7	23,9	27,2	30,1	33,7	36,2
20	8,26	9,24	10,8	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6
21	8,90	9,92	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9
22	9,54	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,7	44,3
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6
27	12,9	14,1	16,1	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6
30	14,9	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9

ДОДАТОК В

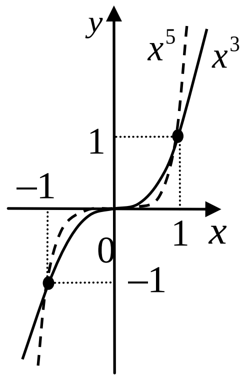
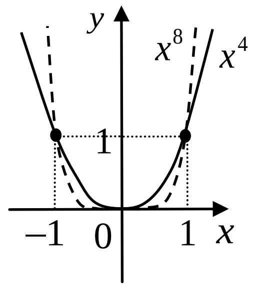
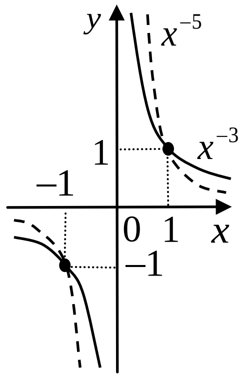
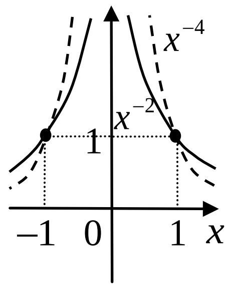
Значення t-критерію Стьюдента

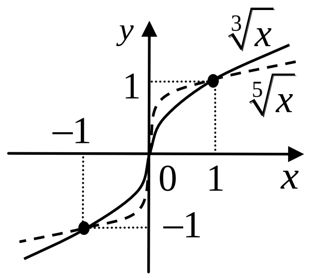
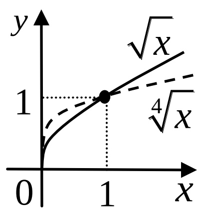
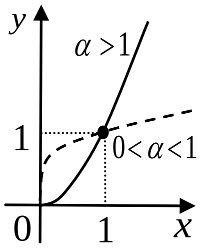
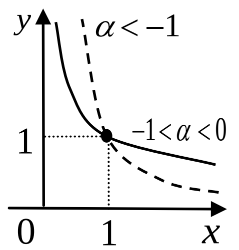


Число ступенів вільності, k	Ймовірність, γ											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
1	0,16	0,32	0,51	0,73	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66
2	14	29	44	62	0,82	06	34	1,89	2,92	4,30	6,96	9,92
3	14	28	42	58	76	0,98	25	64	35	3,18	4,54	5,84
4	13	27	41	57	74	94	19	53	13	2,78	3,75	4,60
5	13	27	41	56	73	92	16	48	01	57	36	03
6	0,13	0,26	0,40	0,55	1,72	1,91	1,13	1,44	1,94	2,45	3,14	3,71
7	13	26	40	55	71	90	12	41	89	36	00	50
8	13	26	40	55	70	89	11	40	86	31	2,90	35
9	13	26	40	54	70	88	10	38	83	26	82	25
10	13	26	40	54	70	88	09	37	81	23	76	17
11	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11
12	13	26	39	54	69	87	08	36	78	18	68	05
13	13	26	39	54	69	87	08	35	77	16	65	01
14	13	26	39	54	69	87	08	34	76	14	62	2,98
15	13	26	39	54	69	87	07	34	75	13	60	95
16	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,34	1,75	2,12	2,58	2,92
17	13	26	39	53	69	86	07	33	74	11	57	90
18	13	26	39	53	69	86	07	33	73	10	55	88
19	13	26	39	53	69	86	07	33	73	09	54	86
20	13	26	39	53	69	86	06	32	72	09	53	84
21	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,08	2,52	2,83
22	13	26	39	53	69	86	06	32	72	07	51	82
23	13	26	39	53	68	86	06	32	71	07	50	81
24	13	26	39	53	68	86	06	32	71	06	49	80
25	13	26	39	53	68	86	06	32	71	06	48	79
26	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,71	2,06	2,48	2,78
27	13	26	39	53	68	85	06	31	70	05	47	77
28	13	26	39	53	68	85	06	31	70	05	47	76
29	13	26	39	53	68	85	05	31	70	04	46	76
30	13	26	39	53	68	85	05	31	70	04	46	75
40	0,13	0,25	0,39	0,53	0,68	0,85	1,05	1,30	1,68	2,02	2,42	2,70
60	13	25	39	53	68	85	05	30	67	00	39	66
120	0,13	0,25	0,39	0,53	0,68	0,84	1,04	1,29	1,66	1,98	2,36	2,62
∞	13	25	38	52	67	84	04	28	64	96	33	58

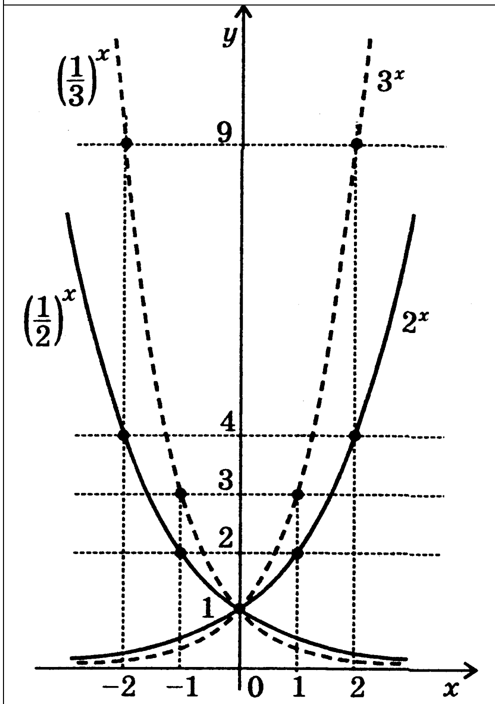
ДОДАТОК Г
ЕЛЕМЕНТАРНІ БАЗИСНІ ФУНКЦІЇ

1. Степенева функція

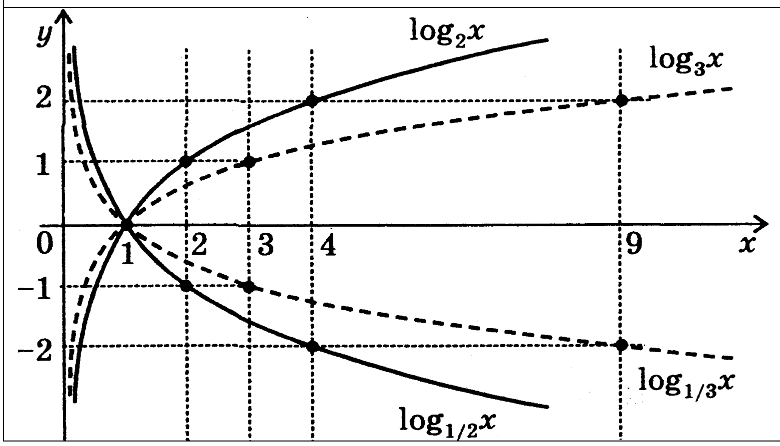
$y = x^n, \quad n \in N$		$y = x^{-n}, \quad n \in N$	
n непарне	n парне	n непарне	n парне
			
$x \in R,\ y \in R.$	$x \in R,\ y \ge 0.$	$x \ne 0,\ y \ne 0.$	$x \ne 0,\ y > 0.$

$y = \sqrt[n]{x}, \quad n \in N$		$y = x^{\alpha}, \quad \alpha \in R$	
n непарне	n парне	$\alpha > 0$	$\alpha < 0$
			
$x \in R,\ y \in R.$	$x \in R,\ y \ge 0.$	$x \ge 0,\ y \ge 0.$	$x > 0,\ y > 0.$

2. Показникова функція

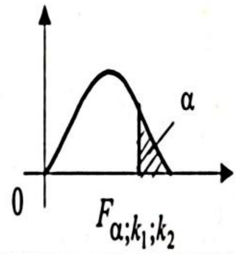
<table>
<tr><td colspan="2">$y = a^x, \quad a > 0$
(при $a = 1$ вироджується в $y = 1$).</td></tr>
<tr><td>
</td><td>1. Область визначення $x \in R$;
2. Область значень $y > 0$;
3. Зростає при $a > 1$;
4. Спадає при $0 < a < 1$.</td></tr>
</table>

3. Логарифмічна функція

<table>
<tr><td colspan="2">$y = \log_a x, \quad a > 0, \quad a \neq 1$</td></tr>
<tr><td>
</td><td>1. Обл. визначення $x > 0$;
2. Обл. значень $y \in R$;
3. Зростає при $a > 1$;
4. Спадає при $0 < a < 1$.</td></tr>
</table>

ДОДАТОК Д

Значення $F_{\alpha,\, k_1,\, k_2}$ - критерію Фішера – Снедекора



α=0,05																			
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	240	242	244	246	248	249	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6.59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84

ДОДАТОК Е
ЕЛЕМЕНТИ ТЕОРІЇ ВИЗНАЧНИКІВ

Визначником матриці ***n-го порядку (n ≥ 2) (детермінантом)*** називається алгебраїчна сума $n!$ членів, кожен з яких є добутком n елементів даної матриці, взятих по одному з кожного рядка і стовпця зі знаком $(-1)^t$, де t − число інверсій в перестановці других індексів, якщо перші записані в порядку зростання. Позначається визначник матриці A: $|A|$ або $\det A$.

Зауваження. Ситуація в перестановці чисел, коли більше число стоїть перед меншим, називається *інверсією.*

Як окремий випадок розглянемо обчислення визначників другого та третього порядку. Визначник другого порядку обчислюється як різниця добутків елементів, що стоять на головній та сторонній діагоналях:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}.$$

Визначник матриці 3-го порядку шукають за методом трикутника:

$$\begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{matrix} \qquad\qquad \begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{matrix}$$

$$\boldsymbol{a}^{+} \qquad\qquad\qquad \boldsymbol{a}^{-}$$

$$|A|_{3\times 3} = a^{+} - a^{-} = \left(a_{11}\cdot a_{22}\cdot a_{33} + a_{21}\cdot a_{13}\cdot a_{32} + a_{12}\cdot a_{23}\cdot a_{31}\right) -$$
$$= \left(a_{13}\cdot a_{22}\cdot a_{31} + a_{21}\cdot a_{12}\cdot a_{33} + a_{11}\cdot a_{23}\cdot a_{32}\right).$$

Визначники третього та вищих порядків можна обчислювати шляхом розвинення за елементами ***будь-якого*** рядка або стовпця:

$|\mathbf{A}| = \sum_{k=1}^{n} a_{kj} A_{kj}$ (формула розвинення визначника за елементами *j*-го стовпця);

$|\mathbf{A}| = \sum_{k=1}^{n} a_{ik} A_{ik}$ (формула розвинення визначника за елементами *i*-го рядка),

де A_{ij} – *алгебраїчне доповнення* елемента a_{ij}, $i = \overline{1, n}$, $j = \overline{1, n}$:

$$A_{ij} = (-1)^{i+j} \cdot M_{ij},$$

M_{ij} – мінор елемента a_{ij}, $i = \overline{1, n}$, $j = \overline{1, n}$.

Мінором M_{ij} елемента a_{ij} матриці *A n*-го порядку є визначник матриці *(n-1)*-го порядку, яка отримується шляхом вилучення з $|A|$ *i*-го рядка та *j*-го стовпця.

Властивості визначника n-го порядку (n ≥ 2)

1. Якщо змінити місцями два рядки (два стовпці) визначника, то визначник змінить свій знак на протилежний. Наприклад,
$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = -\begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix}$$
2. Якщо елементи деякого рядка (стовпця) визначника дорівнюють 0, то визначник дорівнює **0**.
3. Якщо елементи деякого рядка (стовпця) визначника дорівнюють відповідно елементам іншого рядка (стовпця) визначника, то такий визначник дорівнює **0**.
4. Загальний множник елементів деякого рядка (стовпця) можна винести за знак визначника:
$$\begin{vmatrix} ka_{11} & ka_{12} \\ a_{21} & a_{22} \end{vmatrix} = k\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$
5. Якщо елементи деякого рядка (стовпця) визначника пропорційні елементам іншого рядка (стовпця), то визначник дорівнює **0**:
$$\begin{vmatrix} a_{11} & ka_{11} \\ a_{21} & ka_{21} \end{vmatrix} \xrightarrow{(4)} k\begin{vmatrix} a_{11} & a_{11} \\ a_{21} & a_{21} \end{vmatrix} \xrightarrow{(3)} 0.$$

6. Визначник не зміниться, якщо до елементів деякого рядка (стовпця) додати елементи іншого рядка (стовпця), помножені на одне й те ж число k.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12}+ka_{11} \\ a_{21} & a_{22}+ka_{21} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & ka_{11} \\ a_{21} & ka_{21} \end{vmatrix} \xrightarrow{(5)} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

Приклад. Обчислити визначник $\begin{vmatrix} a & b & c & d \\ 1 & 2 & 3 & 1 \\ 2 & -1 & 4 & 0 \\ 0 & 1 & 3 & 1 \end{vmatrix}$.

Розв'язання

Обчислимо даний визначник шляхом розвинення його за елементами першого рядка:

$$\begin{vmatrix} a & b & c & d \\ 1 & 2 & 3 & 1 \\ 2 & -1 & 4 & 0 \\ 0 & 1 & 3 & 1 \end{vmatrix} = a\cdot\begin{vmatrix} 2 & 3 & 1 \\ -1 & 4 & 0 \\ 1 & 3 & 1 \end{vmatrix} - b\cdot\begin{vmatrix} 1 & 3 & 1 \\ 2 & 4 & 0 \\ 0 & 3 & 1 \end{vmatrix} + c\cdot\begin{vmatrix} 1 & 2 & 1 \\ 2 & -1 & 0 \\ 0 & 1 & 1 \end{vmatrix} - d\cdot\begin{vmatrix} 1 & 2 & 3 \\ 2 & -1 & 4 \\ 0 & 1 & 3 \end{vmatrix} =$$

$=a(2\cdot4\cdot1+3\cdot0\cdot1+(-1)\cdot3\cdot1-1\cdot4\cdot1-3\cdot0\cdot2-(-1)\cdot3\cdot1)-b(1\cdot4\cdot1+3\cdot0\cdot0+2\cdot3\cdot1-0\cdot4\cdot1-3\cdot0\cdot1-2\cdot3\cdot1)+c(1\cdot(-1)\cdot1+2\cdot0\cdot0+2\cdot1\cdot1-0\cdot(-1)\cdot1-1\cdot0\cdot1-2\cdot2\cdot1)-$
$-d(1\cdot(-1)\cdot3+2\cdot4\cdot0+2\cdot1\cdot3-0\cdot(-1)\cdot3-1\cdot4\cdot1-2\cdot2\cdot3)=4a-4b-3c+13d$.

Зауваження: детермінанти третього порядку обчислені за методом трикутника.

Слідом квадратної матриці A n-го порядку (позначення $tr(A)$ від англійського слова "trace") називають суму її діагональних елементів:

$$tr(A)=a_{11}+a_{22}+...+a_{nn}=\sum_{i=1}^{n}a_{ii}\,.$$

Властивості сліду матриць:

1. $tr(E_n)=n$.
2. $tr(\lambda A)=\lambda tr(A)$.
3. $tr(A^T)=tr(A)$.
4. $tr(A+B)=tr(A)+tr(B)$.

5. $tr(AB) = tr(BA)$.

Зокрема, якщо A – $(n \times 1)$ вектор-стовпець, $B = A^T$, то

$$tr\left(AA^T\right) = tr\left(A^T A\right),$$

де AA^T, $A^T A$ – квадратні матриці n-го та першого порядків.

Обернена матриця

Матриця A^{-1} називається ***оберненою*** до матриці ***A***, якщо

$$AA^{-1} = A^{-1}A = E.$$

N.B. *Обернену матрицю шукають лише для квадратної матриці!*

Теорема про існування та обчислення оберненої матриці. Якщо визначник квадратної матриці A не дорівнює 0, тобто $|\mathbf{A}| \neq 0$, то існує обернена до неї матриця A^{-1}, яка обчислюється за формулою:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ & & \cdots & \\ A_{31} & A_{32} & \cdots & A_{nn} \end{pmatrix}^{\mathrm{T}},$$

де A_{ij} – алгебраїчне доповнення елемента a_{ij}, $i = \overline{1, n}$, $j = \overline{1, n}$.

Приклад. Знайти матрицю, обернену до матриці $A = \begin{pmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 4 & 3 & 1 \end{pmatrix}$.

Розв’язання

$$|A| = \begin{vmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 4 & 3 & 1 \end{vmatrix} = 2 \cdot \begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} - 3 \cdot \begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix} + \begin{vmatrix} 3 & 2 \\ 4 & 3 \end{vmatrix} = 2(2-6) - 3(3-8) + 9 - 8 = 8,$$

$$A_{11} = \begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} = -4, \qquad A_{12} = -\begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix} = 5, \qquad A_{13} = \begin{vmatrix} 3 & 2 \\ 4 & 3 \end{vmatrix} = 1,$$

$$A_{21}=-\begin{vmatrix}3 & 3\\ 1 & 1\end{vmatrix}=0,\qquad A_{22}=\begin{vmatrix}2 & 1\\ 4 & 1\end{vmatrix}=-2,$$

$$A_{23}=-\begin{vmatrix}2 & 3\\ 4 & 3\end{vmatrix}=6,$$

$$A_{31}=\begin{vmatrix}3 & 1\\ 2 & 2\end{vmatrix}=4,\qquad A_{32}=-\begin{vmatrix}2 & 1\\ 3 & 2\end{vmatrix}=-1,$$

$$A_{33}=\begin{vmatrix}2 & 3\\ 3 & 2\end{vmatrix}=-5.$$

$$A^{-1}=\frac{1}{8}\begin{pmatrix}-4 & 0 & 4\\ 5 & -2 & -1\\ 1 & 6 & -5\end{pmatrix}.$$

Перевірка:

$$A^{-1}A=\frac{1}{8}\begin{pmatrix}-4 & 0 & 4\\ 5 & -2 & -1\\ 1 & 6 & -5\end{pmatrix}\cdot\begin{pmatrix}2 & 3 & 1\\ 3 & 2 & 2\\ 4 & 3 & 1\end{pmatrix}=\frac{1}{8}\begin{pmatrix}-8+16 & -12+12 & -4+4\\ 10-6-4 & 15-4-3 & 5-4-1\\ 2+18-20 & 3+12-15 & 1+12-5\end{pmatrix}=E.$$

ДОДАТОК Ж

СИСТЕМИ ЛІНІЙНИХ РІВНЯНЬ ТА МЕТОДИ ЇХ РОЗВ'ЯЗУВАННЯ

Розглянемо систему n рівнянь з n невідомими виду:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1; \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2; \\ \vdots \qquad \vdots \qquad \cdots \qquad \vdots \qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n. \end{cases} \tag{1}$$

де a_{ij} – це коефіцієнти при невідомих $x_1,\ldots,x_n$, $i=\overline{1,n}$, $j=\overline{1,n}$;
$b_1,\ldots,b_n$ – вільні члени.

Виникає декілька запитань:

- чи існують розв'язки системи (1);
- як знайти розв'язки системи (1);
- чи є розв'язок даної системи єдиним?

Зважаючи на пераховані запитання відмітимо, що система лінійних рівнянь (1) називається *сумісною*, якщо вона має принаймні один розв'язок, і *несумісною* у іншому випадку.

Складемо матрицю ***A*** з коефіцієнтів при невідомих, матрицю-стовпець ***B*** (стовпець вільних членів) і матрицю-стовпець ***X*** невідомих таким чином:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}; \qquad \mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}; \qquad \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Якщо система має визначник, який не є **0**, тобто $|\mathbf{A}| \neq 0$, то вона є *невиродженою* і завжди має *єдиний розв'язок*.

Розглянемо найбільш уживані методи розв'язування систем лінійних рівнянь: метод Крамера, метод Гаусса та метод звичайних жорданівських виключень.

МЕТОД КРАМЕРА

Теорема Крамера. Нехай Δ – визначник матриці ***A*** ($\Delta \neq 0$), а Δ_j – визначник матриці, яка отримується з матриці ***A*** шляхом заміни j-го стовпця стовпцем вільних членів $\boldsymbol{B}$. Тоді система (1) має єдиний розв'язок

$$X = \begin{pmatrix} \frac{\Delta_1}{\Delta} \\ \frac{\Delta_2}{\Delta} \\ \vdots \\ \frac{\Delta_n}{\Delta} \end{pmatrix}.$$

Приклад. Розв'язати методом Крамера систему лінійних рівнянь:

$$\begin{cases} 2x_1 + 3x_2 + x_3 = 2 \\ 3x_1 + 2x_2 + 2x_3 = 2 \\ 4x_1 + 3x_2 + x_3 = -2 \end{cases}.$$

Розв'язання

$$\Delta = \begin{vmatrix} 2 & 3 & 1 \\ 3 & 2 & 2 \\ 4 & 3 & 1 \end{vmatrix} = 2 \cdot \begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} - 3 \cdot \begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix} + \begin{vmatrix} 3 & 2 \\ 4 & 3 \end{vmatrix} = 2(2-6) - 3(3-8) + 9 - 8 = 8.$$

$$\Delta_1 = \begin{vmatrix} 2 & 3 & 1 \\ 2 & 2 & 2 \\ -2 & 3 & 1 \end{vmatrix} = 2 \cdot \begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} - 3 \cdot \begin{vmatrix} 2 & 2 \\ -2 & 1 \end{vmatrix} + \begin{vmatrix} 2 & 2 \\ -2 & 3 \end{vmatrix} = 2(2-6) - 3(2+4) + 6 + 4 = -16.$$

$$\Delta_2 = \begin{vmatrix} 2 & 2 & 1 \\ 3 & 2 & 2 \\ 4 & -2 & 1 \end{vmatrix} = 2 \cdot \begin{vmatrix} 2 & 2 \\ -2 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix} + \begin{vmatrix} 3 & 2 \\ 4 & -2 \end{vmatrix} = 2(2+4) - 2(3-8) - 6 - 8 = 8.$$

$$\Delta_3 = \begin{vmatrix} 2 & 3 & 2 \\ 3 & 2 & 2 \\ 4 & 3 & -2 \end{vmatrix} = 2 \cdot \begin{vmatrix} 2 & 2 \\ 3 & -2 \end{vmatrix} - 3 \cdot \begin{vmatrix} 3 & 2 \\ 4 & -2 \end{vmatrix} + 2 \cdot \begin{vmatrix} 3 & 2 \\ 4 & 3 \end{vmatrix} = 2(-4-6) - 3(-6-8) + 2(9-8) = 24.$$

$$x_1 = \frac{\Delta_1}{\Delta} = \frac{-16}{8} = -2;$$

$$x_2 = \frac{\Delta_2}{\Delta} = \frac{8}{8} = 1;$$

$$x_3 = \frac{\Delta_3}{\Delta} = \frac{24}{8} = 3.$$

МЕТОД ГАУССА ТА ЖОРДАНА-ГАУССА

Метод Гаусса базується на використанні елементарних перетворень системи (1), до яких належать такі перетворення:

- будь-які з рівнянь можна міняти місцями;
- множення довільного рівняння на будь-яке число відмінне від нуля;
- додавання до будь-якого з рівнянь іншого, помноженого на деяке число (результат додавання записується на місці початкового рівняння).

Виконуючи такі перетворення, отримаємо нову систему, що еквівалентна початковій системі. Замість того, щоб здійснювати елементарні перетворення над системою (1), їх здійснюють над розширеною матрицею цієї системи:

$$\mathbf{(A,B)} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & \mathrm{b}_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & \mathrm{b}_2 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & \mathrm{b}_n \end{pmatrix}. \qquad (2)$$

Елементарні перетворення над системою (1) будуть, відповідно, перетвореннями над рядками (стовпцями) розширеної матриці цієї системи. Нехай $a_{11} \neq 0$ (цього завжди можна досягти, переставивши місцями довільні стовпці), тоді перший рядок розширеної матриці ділимо на a_{11}. На наступному етапі вилучаємо невідоме x_1 з інших $(n-1)$ рядків:

$$\mathbf{(A,B)^{(1)}} = \begin{pmatrix} 1 & a^{(1)}{}_{12} & a^{(1)}{}_{13} & \cdots & \mathrm{a}^{(1)}{}_{1n} & \mathrm{b}^{(1)}{}_1 \\ 0 & a^{(1)}{}_{22} & a^{(1)}{}_{23} & \cdots & \mathrm{a}^{(1)}{}_{2n} & \mathrm{b}^{(1)}{}_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & a^{(1)}{}_{n2} & a^{(1)}{}_{n3} & \vdots & a^{(1)}{}_{nn} & \mathrm{b}^{(1)}{}_n \end{pmatrix}.$$

Оскільки при елементарних перетвореннях системи визначник розширеної матриці може змінити знак (при заміні місцями рядків чи стовпців матриці) або збільшитися в одне й те ж саме число разів, то визначник новоутвореної матриці не є нулем.

Якщо в результаті перетворень деякі рядки матриці будуть однаковими, то система є *виродженою*. Якщо однаковими є лише коефіцієнти при невідомих, а вільні члени – різними, то система не має розв’язків взагалі, тобто є *несумісною*.

У результаті n кроків необхідно отримати трикутну матрицю вигляду:

$$
\mathbf{(A,B)}^{(n)} = \begin{pmatrix} 1 & a^{(n)}_{12} & a^{(n)}_{13} & \cdots & a^{(n)}_{1n} & b^{(n)}_{1} \\ 0 & 1 & a^{(n)}_{23} & \cdots & a^{(n)}_{2n} & b^{(n)}_{2} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & b^{(n)}_{n} \end{pmatrix}.
$$

Далі знову повертаємось до системи лінійних рівнянь. Знайдемо спочатку x_n із співвідношення $x_n = b^{(n)}{}_n$. Підставляючи x_n у попереднє рівняння знайдемо x_{n-1}. Отримані значення знайдених невідомих підставляємо у попереднє рівняння і т. д. В результаті виконання послідовності вказаних операцій ми отримаємо матрицю $\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, яка і є розв’язком системи (1).

Метод Жордана-Гаусса (метод звичайних жорданівських виключень) базується на методі Гаусса, але в результаті m кроків необхідно отримати діагональну матрицю:

$$
\boldsymbol{(A, B)^{(n)}} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & b^{(m)}{}_1 \\ 0 & 1 & 0 & \dots & 0 & b^{(m)}{}_2 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & b^{(m)}{}_n \end{pmatrix}.
$$

Звідси, $x_n=b^{(m)}{}_n, \dots, x_1=b^{(m)}{}_1$.

Зауваження. У випадку, коли $a_{11} = 0$, існує два варіанти розв’язування системи (1):

- якщо довільний елемент $a_{i1} \neq 0$, то в зведеній матриці (2) можна поміняти місцями перший та i-ий рядки і далі виконувати всі необхідні перетворення;

- якщо довільний елемент $a_{1j} \neq 0$, то в зведеній матриці (2) можна поміняти місцями перший та j-ий стовпці і далі знову виконувати всі необхідні перетворення. При цьому слід пам’ятати, що, *переставивши стовпці, ми змінили місце відповідних невідомих в системі лінійних рівнянь*.

Приклад. Розв'язати систему лінійних рівнянь методом Гаусса та Жордана-Гаусса

$$\begin{cases} 2x_1 + 3x_2 + x_3 = 2 \\ 3x_1 + 2x_2 + 2x_3 = 2 \\ 4x_1 + 3x_2 + x_3 = -2 \end{cases}.$$

Розв'язання

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 2 \\ 3 & 2 & 2 & 2 \\ 4 & 3 & 1 & -1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 3 & 2 & 2 \\ 2 & 2 & 3 & 2 \\ 1 & 3 & 4 & -2 \end{array}\right)^{\bullet(-2)} \sim \left(\begin{array}{ccc|c} 1 & 3 & 2 & 2 \\ 0 & -4 & -1 & -2 \\ 0 & 0 & -2 & 4 \end{array}\right).$$

Зауваження:

1. Поміняємо місцями перший та третій стовпці зведеної матриці (при цьому змінюється порядок змінних);
2. Перший рядок переписуємо без змін;
3. Перший рядок домножимо на (-2) і додамо до другого рядка (результат запишемо у другий рядок нової зведеної матриці);
4. Від першого рядка віднімемо третій і результат запишемо у третій рядок нової зведеної матриці.

Тоді маємо:

$$\Rightarrow \quad \begin{array}{l} x_3 + 3x_2 + 2x_1 = 2 \Rightarrow x_3 = 3 \\ 4x_2 + x_1 = 2 \Rightarrow x_2 = 1 \\ -2x_1 = 4 \Rightarrow x_1 = -2. \end{array}$$

Приклад розв'язування системи методом Жордана-Гаусса.

Скористаємося кінцевим результатом методу Гаусса:

$$\left(\begin{array}{ccc|c} 1 & 3 & 2 & 2 \\ 0 & 4 & 1 & 2 \\ 0 & 0 & -2 & 4 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 3 & 0 & 6 \\ 0 & 4 & 0 & 4 \\ 0 & 0 & -2 & 4 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 3 & 0 & 6 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 2 \end{array}\right) \cdot 3 \sim$$

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 2 \end{array}\right).$$

Зауваження:

1. Третій рядок переписуємо без змін;

2. Додамо до другого рядка третій, поділений на 2 (результат запишемо у другий рядок новоутвореної матриці);

3. До першого рядка додамо третій рядок (результат запишемо у перший рядок новоутвореної матриці);

4. Поділимо третій рядок на 2, а другий на 4;

5. Від першого рядка віднімемо другий, помножений на 3.

Звідси, $x_1 = -2;\ x_2 = 1;\ x_3 = 3$.

ДОДАТОК И
ЕЛЕМЕНТИ ТЕОРІЇ МАТРИЦЬ

Матриця – це упорядкований масив елементів, який позначається:

$$\boldsymbol{A=(a_{ij})}, \text{ де } i=\overline{1,m},\ j=\overline{1,n},\ A=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Перший індекс елемента (i) вказує номер рядка, в якому він розташований, а другий індекс (j) – номер стовпця. Розмірністю або порядком матриці є кількість рядків та стовпців, що її складають. У нашому випадку матриця A є матрицею розмірності ($m\times n$), тобто A містить m рядків та n стовпців.

Особливі види матриць

1. Прямокутна матриця A – матриця розмірності ($m\times n$).
2. Квадратна матриця A – матриця розмірності ($n\times n$).
3. Діагональна матриця – квадратна матриця, у якої всі її ненульові елементи містяться на головній діагоналі, що проходить з лівого верхнього кута у правий нижній.

$$A=\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

4. Одинична матриця – діагональна матриця, у якої всі елементи дорівнюють одиниці. Одинична матриця позначається $\boldsymbol{E}$.
5. Нульова матриця – матриця, всі елементи якої дорівнюють нулю. Нульова матриця позначається $\boldsymbol{O}$.
6. Трикутна матриця – така квадратна матриця, у якої елементи, розташовані вище або нижче головної діагоналі дорівнюють нулю.

Алгебраїчні дії над матрицями

1. Рівність матриць

Матриці A та B називають рівними, якщо вони мають однакову розмірність і кожний елемент матриці A дорівнює відповідному елементу матриці B.

2. Транспонування матриць

Транспонованою матрицею відносно матриці A називається матриця A^T, яку можна отримати, замінивши в матриці A рядки стовпцями і навпаки. Наприклад,

$$A=\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix},\ A^T=\begin{pmatrix}1 & 4\\ 2 & 5\\ 3 & 6\end{pmatrix}.$$

3. *Множення матриці на число*

Добутком матриці A на число $\lambda \neq 0$ називається матриця B, утворена з матриці A шляхом множення всіх її елементів на λ, тобто $B=(\lambda a_{ij})$.

4. *Додавання матриць*

Сумою двох матриць однакової розмірності називається матриця C такої ж розмірності, кожен елемент якої дорівнює сумі відповідних елементів матриць A та B, тобто $C=A+B=(a_{ij}+b_{ij})$.

5. *Віднімання матриць*

Зауваження. Операцію віднімання матриць можна ввести на основі операцій 3 та 4, тобто $A+(-1)B=(a_{ij}+(-1)b_{ij}) =(a_{ij}-b_{ij})$.

Основні властивості операції додавання матриць

$A+B=B+A$ (комутативність)

$(A+B)+C=A+(B+C)$ (асоціативність)

$(A+B)^T=A^T+B^T$ (транспонування)

$A+(-A)=0$ (додавання протилежної матриці)

$A+0=A$ (додавання нульової матриці)

Приклад. Обчислити $2\begin{pmatrix}1 & 2 & 3\\ 0 & -1 & 2\\ 1 & 2 & 5\end{pmatrix}-3\begin{pmatrix}6 & 7 & 11\\ 3 & 8 & 12\\ 1 & 1 & 1\end{pmatrix}+\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$.

Розв'язання

$$2\begin{pmatrix}1 & 2 & 3\\ 0 & -1 & 2\\ 1 & 2 & 5\end{pmatrix}-3\begin{pmatrix}6 & 7 & 11\\ 3 & 8 & 12\\ 1 & 1 & 1\end{pmatrix}+\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}=$$

$$=\begin{pmatrix}2 & 4 & 6\\ 0 & -2 & 4\\ 2 & 4 & 10\end{pmatrix}+\begin{pmatrix}-18 & -21 & -33\\ -9 & -24 & -36\\ -3 & -3 & -3\end{pmatrix}+\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}=$$

$$=\begin{pmatrix} 2-18+1 & 4-21+0 & 6-33+0 \\ 0-9+0 & -2-24+1 & 4-36+0 \\ 2-3+0 & 4-3+0 & 10-3+1 \end{pmatrix}=\begin{pmatrix} -15 & -17 & -27 \\ -9 & -25 & -32 \\ -1 & 1 & 8 \end{pmatrix}=$$

$$-\begin{pmatrix} 15 & 17 & 27 \\ 9 & 25 & 32 \\ 1 & -1 & -8 \end{pmatrix}.$$

7. Множення матриць

Множення матриць проводиться за правилом “рядок на стовпець”. Для того, щоб отримати елемент добутку матриць A та B (матриця C), розташований в i-му рядку та j-му стовпці, необхідно додати добутки відповідних елементів i-го рядка матриці A та відповідних елементів j-го стовпця матриці B:

$$(a_{i1}\ \ a_{i2}\ \dots\ \ a_{in})\cdot\begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix}$$

Якщо добуток матриць позначити $C=AB$, то $c_{ij}=a_{i1}\cdot b_{1j}+a_{i2}\cdot b_{2j}+\ldots+a_{in}\cdot b_{nj}$.

N.B. Добуток двох матриць існує *тільки для випадку*, коли *кількість стовпців першої матриці дорівнює кількості рядків другої матриці*. Це означає, що множення двох матриць, узагалі, не є комутативною операцією $AB\neq BA$.

Основні властивості операції множення матриць

$(AB)C=A(BC)$ (асоціативність)

$A(B+C)=AB+AC$ (дистрибутивність множення відносно додавання)

$(AB)^T=B^TA^T$ (транспонування добутку)

$AE=A$ (множення матриці на одиничну матрицю)

$AO=O$ (множення матриці на нульову матрицю)

Приклад. Дано матриці $A=\begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$ та $B=\begin{pmatrix} 1 & -1 & 3 \\ 4 & 3 & 2 \\ 1 & -2 & 5 \end{pmatrix}$.

Обчислити добутки AB та BA.

Розв’язання

$$AB=\begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}\cdot\begin{pmatrix} 1 & -1 & 3 \\ 4 & 3 & 2 \\ 1 & -2 & 5 \end{pmatrix}=$$

$$=\begin{pmatrix} 1\cdot 1+1\cdot 4+(-1)\cdot 1 & 1\cdot(-1)+1\cdot 3+(-1)\cdot(-2) & 1\cdot 3+1\cdot 2+(-1)\cdot 5 \\ 2\cdot 1+1\cdot 4+0\cdot 1 & 2\cdot(-1)+1\cdot 3+0\cdot(-2) & 2\cdot 3+1\cdot 2+0\cdot 5 \\ 1\cdot 1+(-1)\cdot 4+1\cdot 1 & 1\cdot(-1)+(-1)\cdot 3+1\cdot(-2) & 1\cdot 3+(-1)\cdot 2+1\cdot 5 \end{pmatrix}=$$

$$=\begin{pmatrix} 4 & 4 & 0 \\ 6 & 1 & 8 \\ -2 & -6 & 6 \end{pmatrix}.$$

$$BA=\begin{pmatrix} 1 & -1 & 3 \\ 4 & 3 & 2 \\ 1 & -2 & 5 \end{pmatrix}\cdot\begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}=$$

$$=\begin{pmatrix} 1\cdot 1+(-1)\cdot 2+3\cdot 1 & 1\cdot 1+(-1)\cdot 1+3\cdot(-1) & 1\cdot(-1)+(-1)\cdot 0+3\cdot 1 \\ 4\cdot 1+3\cdot 2+2\cdot 1 & 4\cdot 1+3\cdot 1+2\cdot(-1) & 4\cdot(-1)+3\cdot 0+2\cdot 1 \\ 1\cdot 1+(-2)\cdot 2+5\cdot 1 & 1\cdot 1+(-2)\cdot 1+5\cdot(-1) & 1\cdot(-1)+(-2)\cdot 0+5\cdot 1 \end{pmatrix}=$$

$$=\begin{pmatrix} 2 & -3 & 2 \\ 12 & 5 & -2 \\ 2 & -6 & 4 \end{pmatrix}.$$

Навчальне видання

Азарова Анжеліка Олексіївна
Сачанюк-Кавецька Наталія Василівна
Роїк Олександр Митрофанович
Міронова Юлія Володимирівна

ЕКОНОМЕТРІЯ

Навчальний посібник

Редактор Т. Старічек
Оригінал-макет підготовлено А. Азаровою

Підписано до друку р.
Формат 29,7 × 42 1/4. Папір офсетний.
Гарнітура Times New Roman.
Друк різографічний. Ум. друк. арк.
Наклад 300 (1-й запуск – 100) прим. Зам. №

Вінницький національний технічний університет,
навчально-методичний відділ ВНТУ.
21021, м. Вінниця, Хмельницьке шосе, 95,
ВНТУ, ГНК, к. 114.
Тел. (0432) 59-85-32.
Свідоцтво суб’єкта видавничої справи
серія ДК №3516 від 01.07.2009 р.

Віддруковано у Вінницькому національному технічному університеті
в комп’ютерному інформаційно-видавничому центрі
21021, м. Вінниця, Хмельницьке шосе, 95.
ВНТУ, ГНК, к. 114.
Тел. (0432) 59-87-38.
Свідоцтво суб’єкта видавничої справи
Серія ДК №3516 від 01.07.2009 р.