**ISBN 978-617-7195-53-4 (PDF)**

Міністерство освіти і науки України

Одеська державна академія будівництва і архітектури

# НАВЧАЛЬНИЙ ПОСІБНИК

з курсу

«**ЕКОНОМЕТРІЯ**»

для студенів спеціальності 076

«Підприємництво, торгівля та біржова діяльність»

денної форми навчання

ОДЕСА 2017

**УДК 330.43(075.8)**
**О 49**

Затверджено Вченою радою факультету економіки і управління будівництвом ОДАБА (протокол № 4 від 1.12. 2016 р.)

Навчальний посібник розглянутий на засіданні науково-методичної комісії ОДАБА (протокол № 3 від 29.11. 2016 р.).

Навчальний посібник з економетрії для студенів спеціальності 076 «Підприємництво, торгівля та біржова діяльність» денної форми навчання розглянутий та рекомендований до друку на засіданні кафедри прикладної та обчислювальної математики і САПР (протокол № 4 від 10.11.2016 р.).

**О 49** **Окара Д. В.**
**Економетрія** : навч. посіб. / Окара Д. В. та ін. — Одеса : ОДАБА, 2018. — 144 с. : іл. **ISBN 978-617-7195-53-4 (PDF)**

Укладачі:

Денисенко Вікторія Юріївна, к. техн. н., доцент ОДАБА,

Ковальова Ірина Лаврентіївна, к. техн. н., доцент ОДАБА,

Молчанюк Ірина Володимирівна, к. ф.-м. н., доцент ОДАБА,

Окара Діана Василівна, к. ф.-м. н., доцент ОДАБА,

Хрістіченко Олег Павлович, ст. викладач ОДАБА,

Чернишев Валентин Геральдович, к. ф.-м. н., доцент ОНЕУ.

Рецензенти:

Кічмаренко Ольга Дмитрівна, канд. ф.-м. н., завідувач кафедри оптимального керування та економічної кібернетики Одеського національного університету ім. І.І.Мечникова;

Мацкул Валерій Миколайович, канд. ф.-м. н., завідувач кафедри математичних методів аналізу економіки Одеського національного економічного університету, доцент.

У навчальному посібнику розглядаються лінійні та нелінійні регресійні моделі, а також особливі випадки у багатофакторному кореляційно-регресійному аналізі.

Відповідальна за випуск: Молчанюк І.В., канд.ф.-м.н, завідувач кафедри прикладної та обчислювальної математики і САПР, доцент.

**УДК 330.43(075.8)**
**О 49**

**ISBN 978-617-7195-53-4 (PDF)**

## ЗМІСТ

# ВСТУП

## *Предмет і задачі економетрії*

Одним із основних завдань економічного аналізу є моделювання розвитку економічних явищ і процесів при створенні тих чи інших умов. Математичне моделювання економічних процесів пов'язано з кардинальними змінами світової економічної системи, з розвитком «нової економіки», з руйнованими економічними та фінансовими кризами та деструкціями, що мали місце в останні десятиліття. Нестабільність світової та національної економіки настільки велика, що мова йде про кризу сучасної економічної системи у цілому. Сучасна ринкова економіка націлена на покращення використання статистичної та економічної інформації, яка характеризує результати економічної діяльності. Створення надійної інформаційної бази для менеджменту в усіх галузях економіки неможливе без врахування дії різних чинників, які позитивно або негативно впливають на результати господарювання. Доцільно виділити окремий вплив тих чинників, що безпосередньо залежать від прийняття управлінських рішень даним об'єктом, і вплив чинників, що від менеджменту на даному господарському об'єкті не залежать. При відсутності необхідних законів для перехідного періоду до ринкової економіки, або при зміні цін, тарифів, економічних нормативів, податків, інфляції (що безперечно не залежить від конкретного об'єкта господарювання) відповідно погіршуються економічні результати роботи цього об'єкта. Усунення впливу таких чинників у економічних розрахунках і характеристика їх впливу шляхом відповідних обчислень дозволяє більш правильно прогнозувати результати господарської діяльності в майбутньому періоді.

Галузь економічної науки, яка вивчає методи кількісного вимірювання взаємозв'язків між соціально-економічними процесами та явищами, називають *економетрією* [1].

Визначення предмета економетрії в різних виданнях подається по-різному. Можна виділити такі підходи до визначення економетрії, що характерні для зарубіжної економетричної літератури [23]:

економетрія – це наука, що вивчає вимірювання зв'язків у відповідному економічному аналізі (Л. Клейн);

економетрію Г. Тінтнер ототожнює з економічною статистикою;

економетрія – це застосування математичних і статистичних методів у економіці (Г. Хансен);

економетрія – це синтез економічної теорії і математики;

економетрія –синтез економічної теорії, математики і статистики.

Кожне з цих означень предмета економетрії має своє раціональне зерно. Надалі будемо використовувати таке *класичне визначення*: *економетрія – це наука, що вивчає кількісні закономірності та взаємозв'язки економічних об'єктів і процесів із викори-станням математико-статистичних методів і моделей.*

Під *об'єктом економетрії* розуміють економічні системи й об'єкти різного рівня складності: від окремого підприємства чи фірми до економіки галузей, регіонів, держави й світу в цілому.

*Предмет економетрії* - це методи побудови й дослідження математико-статистичних моделей економіки, проведення кількісних досліджень економічних явищ, пояснення й прогнозування розвитку економічних процесів.

*Основна задача економетрії* – наповнення емпіричним змістом апріорних (від лат. слова a priori – спочатку, до досліду) економічних міркувань (Л. Клейн). Вона поєднує сукупність методів і моделей, що дозволяють на базі економічної теорії, економічної статистики й математико-статистичного інструментарію додавати кількісні вираження якісним залежностям.

Економетрія є одночасно нашими телескопом і мікроскопом для вивчення навколишнього економічного світу (Ц. Грихелес).

*Метою економетрії* є емпіричне виведення економічних законів.

*Місце курсу «Економетрія» серед дисциплін фундаментальної підготовки бакалаврів з економічних спеціальностей*

Економетрія як сучасний науковий напрям швидко розвивається. Досягнення сучасної економічної науки пред'являють нові вимоги до вищої освіти майбутніх економістів. Без знання економетричних методів неможливо ні дослідження і теоретичне узагальнення емпіричних залежностей економічних змінних, ані побудова надійного прогнозу. Мета вивчення курсу «Економетрія» - навчитися аналізувати інформаційні потоки в соціально-економічних системах, прогнозувати їх поводження, оцінювати і будувати економетричні моделі різного рівня. Економетрія як навчальна дисципліна належить до провідних у фундаментальній підготовці бакалаврів з економіки. Вона будується на основі математичних та економічних знань. Економетричні моделі можуть використовуватися, наприклад, для прогнозування або оцінювання впливу прийнятих рішень чи урядових постанов на подальші зміни цін, податків чи стан справ будь-якого підприємства. Економетричні розрахунки допомагають краще зрозуміти економічні явища й процеси, що, у свою чергу, допомагає більш достовірно формулювати поради і давати прогнози. Це вимагає умілої й реалістичної економічної політики, і, навпаки, ефективна політика вимагає кращого розуміння різних взаємозв'язків між чинниками й результатами господарської діяльності.

Економетрія як сучасний науковий напрям застосовує для кількісного економічного аналізу науково-обгрунтовані засоби, що сформувалися у фундаментальних й прикладних напрямках наукових досліджень, зокрема, у математичній статистиці, математичній економіці і дослідженні операцій. Важливим є знання економічних категорій і понять. Під час вивчення економетрії важливо вміти використовувати комп'ютерну техніку (на базі курсу «Інформатика та комп'ютерна техніка»), що допомагає суттєво спростити розв'язування задач обчислювального характеру, дає можливість глибше зануритись у гнучкій процес економетричного моделювання і допомагає візуально ефектно відображати результати досліджень.

*Зв'язок економетрії із суміжними дисциплінами*

Економетрія є синтезною дисципліною, що поєднує в собі економічну теорію, математичну економіку, економічну і математичну статистику.

Основні результати економічної теорії носять якісний характер, а економетрія вносить у них *емпіричний зміст*. *Математична економіка* виражає економічні закони у вигляді математичних співвідношень, а економетрія *здійснює досвідчену перевірку цих законів*. Математична економіка стає економетрією, коли символічно подані в рівняннях коефіцієнти замінюють конкретними числовими оцінками, отриманими на базі відповідних статистичних даних методами математичної статистики. Економічна й математична статистики дають інформаційне забезпечення досліджуваного процесу у вигляді вихідних даних і економічних показників, а економетрія *проводить аналіз кількісних взаємозв'язків між цими показниками*. Багато базових понять економетрії мають два визначення – «економічне» і «математичне». Дослідник у значній мірі залежить від інформації, яка не може бути прямо проконтрольована. Така інформація часто має помилки вимірювання, тому треба розробляти спеціальні методи для аналізу подібних помилок.

Економетричні методи застосовуються, зокрема, для побудови великих економетричних систем моделей, що описують економіку тієї чи іншої країни і складених елементів, що включають виробничу функцію, інвестиційну функцію, а також рівняння, що характеризують рух зайнятості доходів, цін і процентних ставок та інші блоки. Серед найбільш відомих економетричних систем подібного роду, за якими ведуться розрахунки на ЕОМ, - Брукингська модель (Brookings model, США; призначена для аналізу структури ділових циклів і тенденцій економічного зростання, зміни цін і таке інше), Голландська модель (призначена для прогнозування й розробки економічної політики), Уортонська модель (Wharton model, США; призначена для щоквартального прогнозування економічної активності й рівня безробіття).

*Структура дисципліни*

Економетрію умовно можна розділити на дві частини:

1) економетричні методи;
2) економетричні моделі економічних процесів і явищ.

Економетричні методи, у свою чергу, розбивають на такі групи:

1) методи оцінки параметрів класичної економетричної моделі, зокрема, метод найменших квадратів (МНК);
2) методи оцінки параметрів узагальненої моделі, коли порушуються деякі припущення застосування МНК;
3) методи оцінювання параметрів динамічних моделей;
4) статистичний аналіз часових рядів;
5) методи оцінювання параметрів економетричних моделей, що будуються на підставі системи одночасних рівнянь.

*Історія і розвиток економетрій*

Уперше назву «економетрія» у 1910 році запровадив львівський учений П. Чомпа, опублікувавши книгу «Нариси економетрії і природної теорії бухгалтерії, яка ґрунтується на політичній економії». Однак це поняття в той час не було досить поширеним, оскільки не було фундаментальних праць у цій галузі науки.

У 1926 році термін «економетрія» [econometrics] «знову» використовується норвезьким економістом Рагнаром Фрішем (1895 – 1973) (у буквальному перекладі «econometrics» означає «вимірювання економіки»). Замість терміна «економетрія» часто використовують термін «економетрика»; відомий журнал цього напрямку, заснований у 1933 році Р.Фрішем, теж називається «Econometrica». Термін не відображає дійсного положення речей: поняття економетрії набагато ширше, хоча вимір економіки залишається однією з її частин. Як окрема галузь науки економетрія відома під такою назвою з 1930 року, коли було засновано «Міжнародне товариство для розвитку економічної теорії і її зв'язку зі статистикою і математикою».

Джерела економетрії беруть початок у «політичній арифметиці» англійського економіста В. Петті (1623 – 1687), у наукових працях французького математика, економіста А. О. Курно (1801 – 1877), німецького статистика Е. Енгеля (1821 – 1896). Наприкінці XIX століття були опубліковані роботи італійського економіста В.Парето (1848 – 1923) (рівняння гіперболи для опису розподілу приросту населення), англійського статистика Р. Гукера і російського статистика О.О. Чупрова (1874 – 1926) (кореляційний аналіз економічних процесів). Наприкінці XIX – початку XX століття були розроблені і почалося використання таких статистичних методів, як множинна регресія, статистична перевірка гіпотез, теорія помилок, вибіркові методи, що були розроблені англійським статистиком-біологом К. Пірсоном (1857 – 1936) та основоположником математичної генетики англійцем Р. Е. Фішером (1890 – 1962) та іншими.

На початку XX століття у деяких країнах здійснювалися спроби створення так названого «барометра розвитку». Найбільш відомий з них – «гарвардський барометр», за допомогою якого в двадцяті роки намагалися передбачати поводження товарного і грошового ринку. Гарвардська школа вважалася в той час центром економічних досліджень. Тут уперше почали системно вивчати низку економічних показників (з урахуванням взаємозв'язку між ними) і на їхній підставі досліджувати тенденції і цикли економічних процесів.

Економічна криза 1929 – 1933 років змусила вчених критично переглянути методи аналізу економіки, що використовувалися в той час. Облік в економічних дослідженнях випадкових аспектів економічних явищ дозволив сформуватися економетрії як галузі економічної науки.

Як самостійна дисципліна економетрія сформувалась у 20 – 30-х роках XX століття завдяки працям американських економістів Г. Мура (1869 – 1958; показав, що складні математичні побудови, які наповнені фактичними даними, можуть скласти основу для розробки соціальної стратегії) й Т.В. Шульца (1902 – 1998; проаналізував роль сільського господарства в економіці, розробив теорію людського капіталу). У перших економетричних роботах досліджувалися аналітико-статистичні моделі (рівняння лінійної регресії з параметрами, що оцінюються за допомогою МНК).

У роботах англійських математиків-економістів Р. Аллена (1906 – 1983) і А. Маршалла (1842 – 1924) використовується моделювання структур попиту і споживчих витрат і їх емпірична оцінка; у роботах Е. Уоркинга формулюється задача ідентифікації; економістами Н. Д. Кондратьєвым (1892–1938), Є.Є. Слуцьким (1880–1948), Р. Фрішем розробляється статистичне моделювання ділового циклу. У 1928 році американськими економістами і статистиками Ч. Коббом (1875 – 1949) і П. Дугласом (1892 – 1976) була побудована одна з перших виробничих функцій, що згодом узагальнена американським економістом Р. Солоу (1924 – 1987). Згодом регресійні моделі виявили свою обмеженість при описі узагальнених модельних економічних комплексів і нормативних моделей (моделей, що призначені для знаходження бажаного стану об'єкту, наприклад, оптимального).

Засновниками економетрії вважають Р. Фріша, американського економіста Й. Шумпетера (1883–1950), нідерландського економіста Я.Тінбергена (1903 – 1994) – послідовників неокласичної економічної школи і кейсіанства. Вони одними з перших цілеспрямовано намагалися об'єднати економічну теорію з математичними і статистичними методами.

Спочатку вчені обмежилися вивченням деяких моделей попиту і пропозиції. Починаючи з 30-х років, Я. Тінберген, американський економіст Л. Клейн (1920 - 2013), англійський економіст Р.Стоун (1913–1991) та інші розробили моделі економіки, що описують статистичні зв'язки виробництва, кінцевого індивідуального і державного попиту, цін, податків, зовнішньої торгівлі. Моделі складалися з декількох рівнянь, у зв'язку з чим значно ускладнилися проблеми оцінювання невідомих параметрів. Виникла необхідність використання нового математичного апарата, що розширив можливості практичного застосування економетрії.

Після Другої світової війни важливим центром розвитку економетрії стала Комісія Коулса (діяла в 50–60-ті роки) – об'єднання вчених - економістів США, що поставило метою розвиток якісних методів вивчення економіки. Його девізом був: «Наука є вимірювання»; пізніше він перемінився і став менш категоричним: «Теорія і вимірювання». Комісія названа ім'ям А. Коулса (1891 – 1984), американського бізнесмена, що фінансував її роботу (за це в наукових колах він одержав прізвисько Лоренцо Медічі для економетрії). До діяльності Комісії були притягнуті представники не тільки США, але і ряду інших країн. Завдяки роботі Комісії почалося вивчення комплексних економетричних моделей на макрорівні. Економетрія одержала новий інструментарій у результаті розробки моделей одночасних рівнянь Т. Хаавельмо (1911 – 1999; Норвегія), Т. Купмансом (1910 – 1975; США; народився в Нідерландах), Г. Тейлом (1924 – 2000; Нідерланди). Значний внесок у розвиток економетричних методів, особливо на початковому етапі, внесли також радянські економісти Л. В. Канторович (1912 – 1986), В. В. Новожилов (1892 – 1970), А. Г. Гранберг (1936 – 2010), незважаючи на тривале офіційне третирування економетрії як «буржуазної», «антимарксистської» і «шкідливої лженауки». Методи економетрії здатні були виявити ті чи інші небажані для влади тенденції економічного розвитку, вивчення цих напрямків не заохочувалося. Велика роль у реабілітації економетрії належала В. С. Немчинову (1894 – 1964) – написана їм у 1965 році стаття «Економетрія» з'явилася свого роду відкриттям для широкої громадськості.

Свідченням всесвітнього визнання економетрії є присудження за видатні розробки в цій області премій «Пам'яті Нобеля» у галузі економіки, зокрема, таким вченим:

у 1969 році – Р. Фрішу і Я. Тінбергу за розробку економетричних моделей прийняття рішень;

у 1975 році – Т. Купмансу разом з Л. В. Канторовичем за внесок у розвиток теорії оптимального розподілу ресурсів;

у 1979 році – Т. Шульцу за праці в області аналізу ефективности людського капіталу;

у 1980 році – Л. Клейну за розробку складних економетричних моделей і їх використання для аналізу кон'юнктурних коливань і економічної політики;

у 1983 році – американському економісту і математику французького походження Жерару Дебре (1921 - 2004), Президенту Економетричного суспільства 1971 року, за розробку математичної моделі аналізу умов, що впливають на загальну ринкову рівновагу;

у 1989 році – Т. Хаавельмо за розробку і використання теоретико-ймовірносних методів для аналізу взаємозалежних економетричних структур;

у 2000 році – Дж. Хекману і Д. Макфаддену (США) за розробку мікроеконометрії і методів статистичного аналізу;

у 2003 році – Р. Інглу (США) і К. Грейнджеру (Великобританія) за розробку нових статистичних методів для двох ключових параметрів часового ряду в економіці;

у 2010 році – П.Даймонду, Д.Мортенсену (США) і К.Писсарідесу (Великобританія) за дослідження ринків з моделями пошуку;

у 2013 році – Ю. Фама, Л. П. Хансену і Р. Шиллеру (Великобританія) за емпіричний аналіз зміни ціни активів.

# ЧАСТИНА 1

## РОЗДІЛ 1. ПАРНА ЛІНІЙНА РЕГРЕСІЯ

### 1.1. Типи залежностей між змінними

Одна з головних задач наукового дослідження полягає у вивченні зв'язків між різними явищами. Така залежніть виявляється за допомогою строгих логічних доведень і не має потреби в дослідній перевірці. Велика частина традиційних економічних теорій, у яких зв'язки між економічними категоріями відображаються за допомогою алгебраїчних формул, має справу з точними функціональними співвідношеннями. У математичному аналізі поняття функціональної залежності між змінними $X$ і $Y$ являє собою математичну абстракцію реальних зв'язків між величинами. За визначенням *функціональної залежності*, кожному значенню одної змінної відповідає визначене значення іншої. Геометрична интерпретація функціональної залежності між величинами $X$ і $Y$ зображена на рис.1.1: усі побудовані точки за даними значенням величин $X$ і $Y$ у декартовій системі координат розташовуються на деякій кривій, що є графіком функції $y = f(x)$.

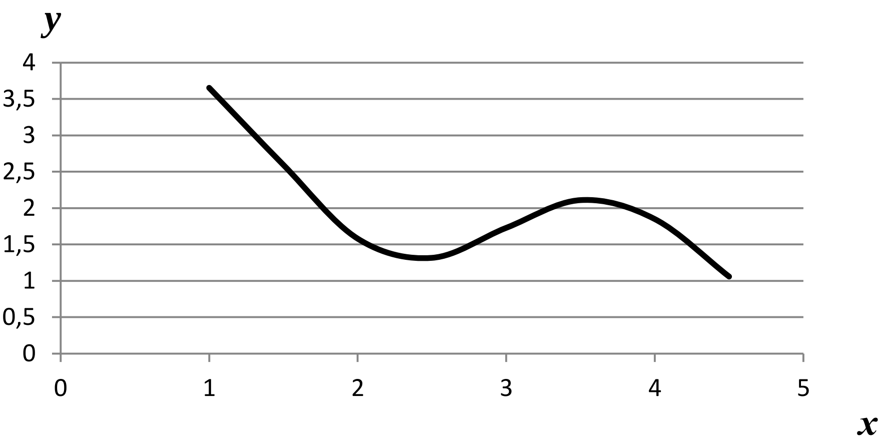



Рис. 1.1

Функціональні залежності вивчаються в математиці, фізиці та інших науках. Наприклад, $S = \pi R^2$ – формула площі кругу; $S = v_0\, t + a\, t^2/ 2$ – формула шляху для рівноприскореного руху.

Якщо $X$ – детермінована величина, яка набуває цілком визначені значення, то і функціонально залежна від неї величина $Y$ теж є детермінованою; якщо ж $X$ – випадкова величина, то і $Y$ – також випадкова величина.

У соціально-економічних явищах, що відрізняються складністю і різноманіттям існуючих взаємозв'язків, функціональні залежності між величинами зустрічаються вкрай рідко. Навіть елементарне знайомство з економічними даними показує, що їхні окремі значения не можуть укладатися на деяку гладку лінію. У цьому разі між величинами $X$ і $Y$ існує зв'язок особливого роду, при якому зі зміною однієї величини змінюється *розподіл* іншої. Такий зв'язок називаеться *стохастичним* або *ймовірносним* і припускає, що для заданого значения незалежної змінної $X$ можна вказати ряд значень залежної змінної $Y$, випадково розсіяних у деякому інтервалі. Стохастична залежність між $Y$ і $X$ пояснюється тим, що залежна змінна, крім виділеної змінної $X$, підпадає під вплив також ряду неврахованих факторів.

Оскільки значения залежної змінної характеризуються випадковим розподілом, то вони не можуть бути передбачені з достатньою точністю, а тільки з деякою ймовірністю. Отримані значення залежної змінної $Y$ є реалізаціями випадкової величини. Так, рівень витрат обсягу в магазинах залежить не тільки від розміру товарообігу магазина, хоча цей фактор має основне значения. На нього впливають і такі фактори, як режим роботи магазина, організація його постачання, умови для зберігання товару, особисті якості персоналу й багато інших. Прикладами кореляційних залежностей можуть бути також: залежність між витратами фірми на рекламу та її прибутками, залежність між собівартістю та обсягами випуску продукції.

У зв'язку з цим в економіці говорять не про функціональні, а про статистичні залежності.

Залежність між двома випадковими величинами $X$ і $Y$ називається *статистичною*, якщо кожному значенню однієї з них відповідає визначений розподіл іншої.

Геометричним зображенням статистичної залежності є сукупність точок, що розташовані в околі деякої лінії (рис.1.2).

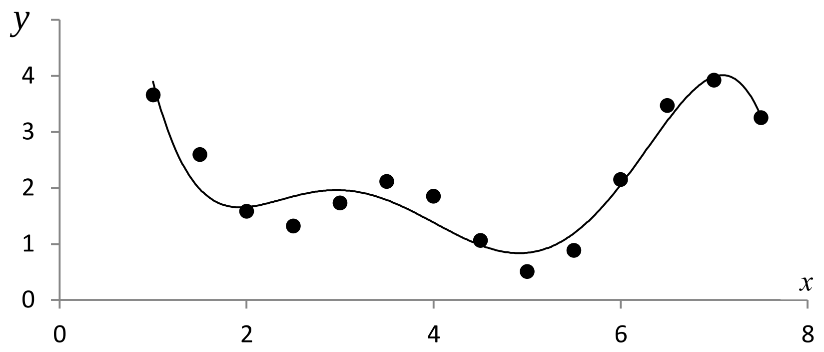



Рис. 1.2

Завдяки неоднозначності статистичної залежності між $X$ та $Y$, становить інтерес *залежність між змінними, при якій кожному значенню одної змінної відповідає визначене середнє значення* (*умовне математичне сподівання*) *іншої*. Така статистична залежність є кореляційною.

*Кореляційною залежністю* між двома змінними називається функціональна залежність між значеннями однієї з них і середнім значенням (умовним математичним сподіванням) іншої.

Кореляційні залежності $Y$ на $X$ і $X$ на $Y$ можуть бути записані відповідно у вигляді:

$$\overline{y}_x = \varphi(x), \quad \text{або} \quad M_x(Y) = \varphi(x), \qquad (1.1)$$

$$\overline{x}_y = \psi(y), \quad \text{або} \quad M_y(X) = \psi(y),$$

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

Кореляційний зв'язок між ознаками може виникати різними шляхами. Найважливіший з них - причинна залежність результативної ознаки (її варіації) від варіації факторної ознаки. Дуже важливо зрозуміти суть досліджуваного зв’язку, оскільки кореляційний зв'язок може виникати між двома *наслідками загальної причини*. Класичним є приклад, наведений відомим статистиком XX століття О.О.Чупровим (1874 – 1926). Якщо як ознаку $X$ взяти число пожежних команд у місті, а за ознаку $Y$ - суму збитків у місті від пожеж, то між ознаками $X$ та $Y$ в містах виявиться значна пряма кореляція. У середньому, чим більше пожежників у місті, тим більше збитків від пожеж. В чому ж справа? Дану кореляцію неможливо інтерпретувати як зв'язок причини та наслідку, тому що обидві ознаки - *наслідки загальної причини* - розміру міста. У великих містах більше пожежних частин, але більше і пожеж, і збитків від них за рік, ніж у малих. Такого типу кореляцію називають *хибною кореляцією*.

Хибна кореляція виникає і у випадку, коли кожна з ознак є і *причиною, і наслідком*. Наприклад, при відрядній оплаті праці існує кореляція між продуктивністю праці і заробітною платою. З одного боку, чим вища продуктивність праці, тим вища заробітна плата. З іншого - висока заробітна плата теж є стимулюючим фактором, який примушує робітника працювати більш інтенсивно. Аналогічно, надходження нових книг у наукову бібліотеку не означає, що підвищаться екзаменаційні оцінки студентів.

На початковій стадії аналізу статистичних даних не завжди потрібні кількісні оцінки, досить тільки визначити напрямок і характер зв'язку, з'ясувати форму впливу одних факторів на інші. З цією метою застосовують методи зіставлення паралельних даних, аналітичних групувань та графічний.

Єдиний спосіб, який абсолютно точно показує причину та наслідок, - це *формальний експеримент*, у якому можна змінювати одну або декілька вихідних змінних і оцінювати значення результативної змінної.

Крім того, необхідно виконати порівняння, яке не залежить від зміни вихідних даних. Це може кількісно оцінити та обмежити вплив інших змінних, які не змінюються, але створюють «шум».

У регресійному аналізі розглядається *однобічна залежність* випадкової змінної $Y$ від однієї (чи декількох) невипадкової незалежної змінної $X$. Така залежність може виникнути, наприклад, у випадку, коли при кожному фіксованому значенні $X$ відповідні значення $Y$ піддані випадковому розкиду за рахунок дії ряду неконтрольованих факторів. Залежність $Y$ від $X$ може бути також зображена у вигляді моделі (рівняння) регресії $Y$ на $X$ (1.1). Статистично залежний показник $Y$ (залежну випадкову величину) називають *залежною змінною* (ендогенною змінною, регресандом, результативною ознакою і т.ін.). Статистично незалежний показник $X$ (незалежну випадкову величину), відповідно, називають *незалежною змінною* (екзогенною змінною, регресором, факторною ознакою).

***Задачею регресійного аналізу*** *є вивчення форми зв'язків* між показником і факторами на підставі статистичних даних.

Рівняння (1.1), що являє собою формулу статистичного зв'язку між змінними, називається *рівнянням регресії*, функція $\varphi(x)$ - *функцією регресії*, а її графік - *теоретичною лінією регресії*.

Термін «регресія» впровадив англійський антрополог Ф.Гальтон (1822 - 1911), виражаючи сутність передачі дітям спадкоємних ознак їх батьків. У статистиці цей термін не зовсім удалий. Доцільніше замість «рівняння регресії», «лінія регресії» тощо говорити «кореляційне рівняння», «кореляційна лінія» і т. ін.

У математичній статистиці розроблені методи оцінки коефіцієнтів, що характеризують кореляцію між випадковими величинами або ознаками, і методи перевірки гіпотез про їх значення, що використовують їх вибіркові аналоги. Сукупність таких методів називається *кореляційним аналізом.*

***Задачею кореляційного аналізу*** *є виявлення зв'язку між випадковими величинами та оцінка його тісноти*.

***Задачею дисперсійного аналізу*** *є дослідження і встановлення степені впливу якісних регресорів на регресанд*.

Існує принципова різниця між функціональною залежністю і регресією. При функціональній залежності аргумент $X$ цілком визначає значення функції $Y$. Крім того, при функціональній залежності функція може мати *обернену*. Функція регресії цієї властивості не має. З огляду на це, розрізняють регресію $Y$ на $X$ (якщо досліджують залежність $Y$ від $X$) і регресію $X$ на $Y$ (якщо досліджують залежність $X$ від $Y$). Так, при вивченні механізму зв'язку між ціною товару ($X$) і попитом ($Y$) практичний інтерес представляють обидві постановки задачі: залежність ціни товару від попиту, а також обернена залежність – попиту від ціни товару, так як зміна цін на товари відбивається на попиті населення. У цьому випадку окремо будуються регресії $Y$ на $X$ і $X$ на $Y$. Обидві змінні тут є випадковими. Кожному значенню $X$ відповідає безліч значень $Y$ і навпаки, кожному значенню $Y$ відповідає безліч значень $X$. Якщо ж логічне тлумачення залежності між двома змінними можливо тільки в одному напрямку, як, наприклад, при дослідженні впливу кількості внесених добрив або опадів ($X$) на врожайність сільськогосподарських культур ($Y$), то знаходиться тільки одна функція регресії $Y$ на $X$. Стохастична залежність $X$ від $Y$ у даному випадку не має змісту.

Регресія за формою залежності може бути (рис. 1.3):

- *парною* (одновимірною, простою або однофакторною), якщо $n = 1$ (тут є лише один фактор $X$, що впливає на результат). Загальна форма рівняння парної регресії:

$$Y = f(X),$$

- *множинною* ($n$-вимірною, багатофакторною), якщо $n > 1$. Загальна форма рівняння множинної регресії:

$$Y = f(X_1, X_2, \ldots, X_n).$$

Залежно від типу залежності рівняння парної регресії може бути:

- *лінійним* (пропорційним, де існує лише єдина форма),
- *нелінійним* (де є різні форми залежно від вигляду функції).

За напрямом зв'язок між корелюючи ми величинами може бути:

- *прямим* (зміна факторної ознаки зумовлює *зміну у тому ж напрямі* результативної ознаки). Наприклад, зв'язок між рівнем механізації виробничих процесів та продуктивністю праці;
- *оберненим* (із *збільшенням* факторної ознаки результативна ознака *зменшується* чи, навпаки, із *зменшенням* факторної ознаки результативна *зростає*). Наприклад, зв'язок між продуктивністю праці та собівартістю продукції, собівартістю продукції та рентабельністю виробництва.

Для знаходження рівняння регресії необхідно знати *закон розподілу* залежної змінної $Y$ за умови, що змінна $X$ набуватиме значення $x$, тобто $X = x$. У статистичній практиці таку інформацію одержати, як правило, не вдається, тому що звичайно дослідник має лише вибіркову пару значень $(x_i, y_i)$ обмеженого обсягу $n$. У цьому випадку мова може йти про *оцінку* (наближений вираз, апроксимацію) по вибірці функції регресії. Такою оцінкою *є вибіркова лінія* (*крива*) *регресії*:

$$\overline{y}_x = \hat{\varphi}(x, b_0, b_1, \ldots, b_p), \qquad (1.2)$$

де $\overline{y}_x$ - *групова середня* змінної $Y$ при фіксованому значенні змінної $X = x$; $b_0$, $b_1$, …, $b_p$ – параметри кривої. Відзначимо, що найкращою оцінкою є *лінія середньої квадратичної регресії*.

Рівняння (1.2) називається *вибірковим рівнянням регресії.*

Правильно визначена апроксимуюча функція $\hat{\varphi}(x, b_0, b_1, ..., b_p)$ при *збільшенні обсягу вибірки* ($n \to \infty$) *збігається за імовірністю до функції регресії* $\varphi(x)$.

Визначення функції регресії відбувається за емпіричними даними, які містять *випадковість*. Практично в кожному окремому випадку фактичне значення результативної змінної $y_i$ має вигляд:

$$y_i = \hat{y}_i + \varepsilon_i,$$

де $\hat{y}_i$ – теоретичне значення результативної змінної, знайдене за рівнянням регресії; $\varepsilon_i$ – випадкова величина, що характеризує відхилення реального значення змінної від теоретичного.

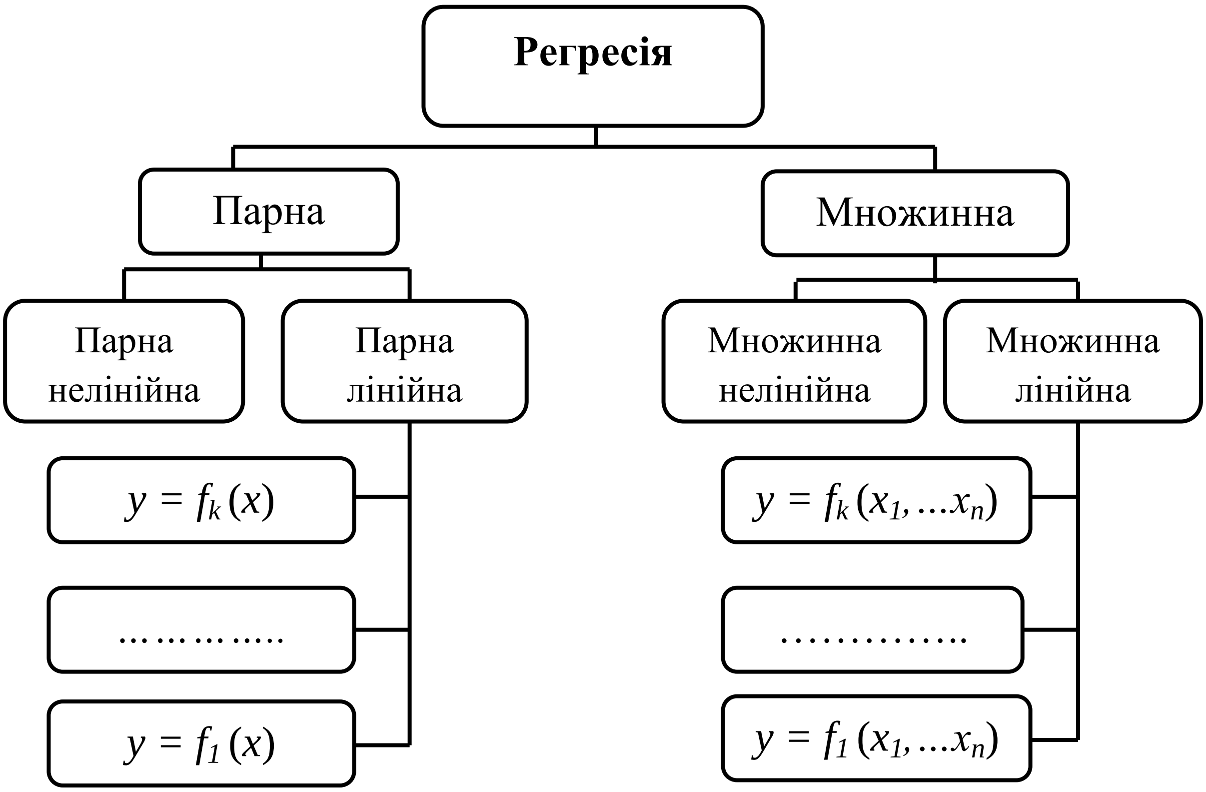



Рис. 1.3.

Випадкова величина

$$\varepsilon = y - \hat{y},$$

називається *збуренням* (*збуреною величиною*) або *залишком*.

Збурення $\varepsilon$ включає вплив неврахованих факторів, випадкових перешкод і помилок спостережень, її значення змінюється для кожного спостереження. Наявність збуреної змінної обумовлена такими причинами:

1. *Уведення в модель не всіх пояснюючих змінних.* Будь-яка регресійна (зокрема, економетрична) модель - це спрощення реальної ситуації. Остання завжди є складною композицією різних факторів, багато з яких у моделі не враховуються, що призводить до відхилення реальних значень залежної змінної від її модельних значень. Наприклад, попит на товар визначається його ціною, цінами

на товари-замінники, на товари, що його доповнюють, прибутком споживачів, їхніми смаками, уподобаннями тощо. Безумовно, перелічити всі пояснюючі змінні практично неможливо. Зокрема, неможливо врахувати такі фактори, як традиції, національні чи релігійні особливості, географічне положення району, погоду та багато інших, вплив яких призводить до деяких відхилень реальних спостережень від модельних. Ці відхилення можуть бути описані як випадкова складова моделі.

У деяких випадках заздалегідь невідомо, які фактори за умов, що склалися, насправді є визначальними, а якими можна знехтувати. Крім того, інколи безпосередньо врахувати якийсь фактор неможливо через відсутність статистичних даних. Наприклад, обсяг заощаджень домогосподарств може визначатися не лише прибутками їх членів, а й станом здоров'я останніх, інформація про яке в розвинутих країнах становить лікарську таємницю. У деяких ситуаціях ряд факторів має принципово випадковий характер, що додає неоднозначності певним моделям, наприклад, погода в моделях, що прогнозують обсяг врожаю.

2. *Неправильний вибір функціональної форми моделі.* Через слабку вивченість досліджуваного процесу або через його мінливість може бути неправильно дібрано функцію, що його моделює. Це спричинить відхилення моделі від реальності, що позначиться на величині випадкової складової. Наприклад, виробнича функція ($Y$) одного фактора ($X$) може моделюватися функцією $Y = b_0+b_1X,$ хоча мала б використовуватися інша модель: $Y = b_0 X^{b1}$ $(0 < b_1 < 1)$, що враховує закон спадної ефективності. Неправильним може бути також добір пояснюючих змінних.

3. *Агрегування змінних.* У багатьох моделях розглядаються залежності між факторами, що самі є складною комбінацією інших, простіших змінних. Так, при вивченні сукупного попиту аналізується залежність, у якій змінна (сукупний попит) є складною композицією

індивідуальних попитів, що може виявитися причиною відхилення реальних значень від модельних.

4. *Помилки вимірювань.* Якою б якісною не була модель, помилки вимірювання змінних впливатимуть на розбіжності між модельними та емпіричними даними, що також позначиться на величині випадкового члена.

*5. Обмеженість статистичних даних.* Найчастіше будуються моделі, що описуються неперервними функціями. Для оцінювання параметрів моделі використовується набір даних, що має дискретну структуру. Це знаходить відображення у випадковому відхиленні.

*6. Непередбачуваність людського фактора.* Ця причина може "зіпсувати" найякіснішу модель. Дійсно, при правильному виборі форми моделі, скрупульозному доборі пояснюючих змінних неможливо спрогнозувати поведінку кожного індивідуума.

Хоча дослідник має у своєму розпорядженні значення пояснюваної та пояснюючих змінних у результаті одночасних спостережень над ними, однак значення збурення $\varepsilon$ безпосередньо одержати не можна, оскільки воно являє собою конгломерат багатьох випадкових впливів. Лише після кількісної оцінки функції регресії можна знайти значення збуреної змінної $\varepsilon$.

## 1.2. Аналітичне групування

Одним із найпростішим способів дослідження взаємозв'язку між результуючою змінною та факторними ознаками є *аналітичне групування.* Групування за однією факторною змінною називають *простим*, а за кількома - *складним*. Розглянемо просте аналітичне групування, яке надалі називатимемо аналітичним групуванням.

*Аналітичне групування* - статистична таблиця, у якій записані інтервали значень факторної ознаки, згідно з якими згруповано одиниці сукупності, а також наведено групові середні значення результуючої змінної.

Кількість груп (інтервалів) аналітичного групування можна встановити емпірично або за формулою американського статистика Г.Стерджеса (1882-1958) залежно від особливостей об'єкта, який досліджують. Формула Стерджеса має вигляд:

$$k = 1 + 3{,}322\ lg\ n,$$

де $k$ - кількість груп аналітичного групування; $n$ - кількість одиниць сукупності. З наведеної формули випливає, що вибір кількості груп залежить від обсягу сукупності. Якщо груп виявляється багато та вони містять мале число одиниць, то групові показники можуть стати ненадійними. У цьому випадку використовується багатомірне групування, яке здійснюється за комплексом ознак одночасно.

Величиною інтервала є різниця між максимальним $\mathrm{x}_{\max}$ та мінімальним $\mathrm{x}_{\min}$ значеннями факторної ознаки, тобто $\mathrm{x}_{\max} - \mathrm{x}_{\min}$. При побудові аналітичного групування треба чітко визначати закінчення проміжків (закритий інтервал, напівзакритий інтервал).

Групування з рівними інтервалами будують, коли досліджуються кількісні відмінності у величині ознаки всередині руп однакової якості, а також якщо розподіл має більш-менш рівномірний характер. Якщо можна заздалегідь визначити певну кількість груп, то величина рівного інтервалу розраховується за формулою

$$\boldsymbol{i = (x_{max} - x_{min})/k}.$$

Якщо не потрібно попереднього встановлення числа груп, то використовується формула

$$i = (x_{max} - x_{min})/(1 + 3{,}322\ lg\ N),$$

де $N$ – кількість спостережень.

Зазначимо, що при розрахунку величини рівного інтервалу за даною формулою, треба знаменник дробу попередньо округлити до цілого числа (як правило, *завжди більшого*, тобто $\lceil k \rceil$), оскільки кількість груп не може бути дробовим числом.

Кількість груп, яка відповідає обсягу сукупності $n$ і розрахована за формулою Стерджеса, наведено у таблиці 1.1:

Таблиця 1.1.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Обсяг сукупності** | **інтервал** | 10 - 24 | 25 - 44 | 45 - 89 | 90 - 179 | 180 - 359 | 360 - 719 |
| | $n$ | 15 | 20 | 45 | 90 | 180 | 360 |
| **Число груп** | $k$ | 4,91 | 5,32 | 6,13 | 7,49 | 8,49 | 9,49 |
| | $\lceil k \rceil$ | 5 | 6 | 7 | 8 | 9 | 10 |

Результати аналітичного групування доцільно подавати графічно у вигляді емпіричної лінії регресії .

Аналітичне групування дає можливість оцінити ступінь узгодженості змін факторної ознаки та групового середнього значення результуючої змінної. Якщо така узгодженість має місце, то вважають, що між змінними є кореляційний зв’язок. Аналітичне групування дозволяє зробити висновок про те, *як змінюється середнє значення результуючої змінної при збільшенні значень факторної ознаки: зростає, спадає, коливається тощо*.

Позитивними якостями аналітичного групування є його простота та наочність, а основним недоліком є низька точність. У цьому легко переконатися, якщо змінити кількість груп.

### 1.3. Основні завдання кореляційно - регресійного аналізу

До основних завдань кореляційно-регресійного аналізу належать:

1. *Специфікація моделі* (*встановлення причинно-наслідкового зв'язку між досліджуваними економічними змінними*). Потрібно визначити, які із змінних є причинами (незалежними змінними, факторами), а які - наслідками (залежними, результуючими змінними). Зазвичай це здійснюють за допомогою положень економічної теорії, а в складніших випадках можна використати методи причинно-наслідкового аналізу. Від того, наскільки вдало розв’язана проблема специфікації моделі, в значній мірі залежить успіх всього моделювання.

2. *Параметризація моделі* (*визначення типу і форми кореляційно - регресійної моделі*).Тип моделі насамперед визначають за кількістю змінних, між якими досліджують зв'язок. Якщо моделюють кореляційну залежність між двома змінними - факторною та результуючою, то модель називають *парною,* або *однофакторною* (використовують ще термін "проста модель"). Якщо маємо кореляційну залежність однієї результуючої змінної від кількох факторних, то кореляційно - регресійну модель називають *множинною,* або *багатофакторною.* Парні та множинні кореляційно - регресійні моделі належать до класу функціональних моделей, симультативні моделі є структурними.

Форму кореляційно - регресійної моделі визначають загальним видом функції регресії і обґрунтовують на основі положень економічної теорії, а в складніших випадках - за допомогою експертних та математичних методів. При моделюванні економічних систем використовують *лінійні* та *нелінійні* функції регресії.

3. *Вибір методу оцінювання невідомих параметрів моделі та побудова моделі.* При побудові моделі висувають певні припущення щодо змінних, які входять в модель, та випадкових факторів, що впливають на ці змінні в реальному житті. Якщо ці припущення виконуються, то кореляційно-регресійні моделі називають *класичними*. Для побудови класичні лінійні моделі, а також класичні нелінійні моделі, які заміною змінних зводять до лінійних (*лінеаризовані, квазілінійні*), найчастіше використовують МНК. Параметри класичних нелінійних регресійних моделей оцінюють за допомогою ітеративних методів, методу Тейла, методу трьох точок. Моделі, під час побудови яких порушують припущення класичного регресійного аналізу, розробляють за допомогою різноманітних модифікацій МНК (методу зважених найменших квадратів, узагальненого МНК, методу непрямих найменших квадратів, двокрокового МНК, трикрокового МНК), а також інших методів.

4. *Оцінювання сили кореляційного зв'язку між змінними*. Тісноту кореляційної залежності результуючої змінної $y$ від факторних ознак оцінюють за допомогою величини розсіювання значень $y$ навколо його умовного середнього значення. Для оцінювання сили кореляційного зв’язку між змінними розраховують параметри зв’язку: коефіцієнти регресії, детермінації, кореляції.

5. *Ідентифікація моделі* (*перевірка моделі на точність*). Перевірка надійності результатів моделювання (діагностику моделі) здійснюють шляхом обчислення статистичних оцінок моделі та параметрів зв’язку: стандартної похибки моделі, вибіркових похибок параметрів моделі, вибіркової похибки прогнозного значення результуючої змінної, похибки кореляційного відношення, довірчих інтервалів для істинних значень параметрів моделі та зв’язку, а також шляхом перевірки нульових гіпотез.

6. *Верифікація моделі* (*перевірка адекватності моделі*). Під адекватністю моделі розуміють відповідність моделі об'єкта чи процесу, що модулюється. Адекватність якоюсь мірою є умовним поняттям, тому що повної відповідності моделі реальному об'єкту бути не може: інакше це була б не модель, а сам об'єкт. При моделюванні мається на увазі адекватність не взагалі, а за тими властивостями моделі, що для дослідження вважаються істотними. Проблема адекватності економічних моделей ускладнюється труднощами виміру економічних величин. З'ясовується, наскілько вдало розв’язані проблеми специфікації, ідентифікації моделі, точність розрахунків за даною моделлю. Адекватність багатьох моделей можна перевірити за допомогою коефіцієнту детермінації. Однак якщо коефіцієнт детермінації має нечітко виражене граничне значення, то використовують інші критерії перевірки моделі на адекватність, наприклад, критерій Фішера, критерій Стьюдента тощо.

7. *Вибір “найкращої” моделі.* Різноманітність і складність економічних процесів зумовлює різноманіття моделей, які

використовують для економетричного аналізу. Усе це істотно ускладнює процес знаходження максимально адекватної (“якісної”, “найкращої”) моделі, а, насамперед, форми економетричної моделі.

Для парної регресії підбір форми моделі звичайно здійснюють на основі аналізу розташування точок спостережень на кореляційному полі. Проте інколи виникають ситуації, коли розташування точок спостережень відповідає кільком функціям і потрібно вибрати найкращу з них.

Наприклад, нелінійну залежність у випадку прискореного зростання можна апроксимувати степеневою, показниковою або поліноміальною функціями.

Під час першої побудови кореляційно - регресійної моделі зазвичай невідомо, яка з них найліпше змальовує реальний процес, і тому часто підбирають модель, що найточніше відповідає реальним даним. При цьому необхідно враховувати, що ідеальної моделі немає.

Завдання вибору “найкращої” моделі є багатокритерійним. Обґрунтований висновок щодо якості моделі здебільшого отримують після детального порівняльного аналізу кількох побудованих та досліджених моделей. Враховують такі критерії “якості” моделі:

- *узгодженість з теорією*. Модель повинна враховувати теоретичні засади функціонування економічних об’єктів;

- *простота.* Модель повинна бути максимально простою, тобто містити якомога меншу кількість змінних, мати простішу форму функціональної залежності та меншу кількість невідомих параметрів. Цей критерій вирізняється тим, що модель не відображає дійсність ідеально,а є її спрощенням. Простота моделі забезпечує використання порівняно нескладних процедур оцінювання параметрів моделі;

- *однозначність.* У будь-якому наборі статистичних даних невідомі параметри моделі потрібно оцінювати однозначно;

- *максимальна адекватність.* Чим більшу частину варіації результуючої змінної зможе пояснити модель, тим кращою вона буде.

Саме тому прагнуть побудувати модель із максимально можливим коефіцієнтом детермінації;

- *прогнозні якості.* Модель може бути якісною, якщо отримані на її основі прогнози підтверджує реальність. Теоретичним критерієм оцінювання прогнозних якостей моделі може бути *коефіцієнт варіації*.

Якість моделі не потрібно абсолютизувати хоча б тому, що навіть «найкраща» модель є результатом «підтасування» специфікації моделі під вибірку. Тому може виникнути ситуація, коли дослідники, які використовують різні групи даних, будують різні кореляційно-регресійні моделі для опису одного і того самого об'єкта дослідження. Іншою проблемою є використання моделі для прогнозування. Іноді моделі, добротні з погляду діагностичних тестів, мають досить низькі прогнозні якості. Враховуючи динаміку зміни економічних процесів та явищ, їхню мінливу поведінку, можна зробити висновок, що практично немає моделей, які є постійно високоякісними. Нові умови функціонування економічних систем вимагають перегляду навіть досить «якісних» моделей.

8. *Аналіз результатів моделювання, їх економічна інтерпретація та практичне використання.* Параметри моделей зазвичай мають чітку та змістовну економічну інтерпретацію. Результати кореляційно-регресійного аналізу використовують здебільшого для прогнозування.

## 1.4. Лінійна кореляційна залежність.

Можна довести, що кореляційна залежність між величинами, розподіленими за нормальним законом, є лінійна. Саме тому лінійна кореляційна залежність $Y$ від $X$ на практиці зустрічається дуже часто.

Нехай випадкові величини $X$ та $Y$ зв'язані лінійною кореляційною залежністю. Рівняння теоретичної лінії регресії $Y$ на $X$ будемо шукати в лінійній формі, тобто у вигляді

$$\overline{y}_x = b_0 + b_1 x, \text{ або } y^{\text{т}} = b_0 + b_1 x, \ \hat{y} = b_0 + b_1 x. \quad (1.3)$$

Параметр $b_0$ - вільний член рівняння регресії, початкове значення результуючої змінної $\bar{y}_x(0)$. Значення параметра $b_0$ має сенс *середнього рівня результуючої змінної при нульовому значенні факторної ознаки* (хоча у багатьох випадках економетричного аналізу соціально-економічних об'єктів $\bar{y}_x(0)$ оцінити важко або взагалі неможливо).

Параметр $b_1$ – *коефіцієнт регресії* (тангенс кута нахилу прямої). Він виявляє, на *скільки одиниць у середньому змінюється величина Y при збільшенні величини X на одну одиницю*. Він має розмірність розглянутої величини. Якщо $b_1 > 0$, то кореляційна залежність є прямою; якщо $b_1 < 0$ - оберненою; якщо $b_1 = 0$, то кореляційна залежність відсутня. *Абсолютне значення параметра $b_1$* показує *величину зміни результуючої змінної у на одну одиницю приросту факторної ознаки х*: $\hat{y}(x + 1) - \hat{y}(x) = b_0 + b_1(x + 1) - (b_0+b_1x) = b_1$.

Парна лінійна модель (1.3) відображає зв'язок між змінними $\hat{y}$ та $x$ в проміжку $[x_{min}, x_{max}]$, визначеному емпіричними значеннями факторної ознаки. Цей проміжок називають *областю існування моделі*. Для ви користування моделі за межами області існування потрібно обґрунтувати той факт, що форма кореляційної залежності результуючої змінної $y$ від факторної ознаки $x$ не зміниться.

Мірою якості знайдених оцінок можуть бути визначені композиції відхилень $\varepsilon_i$, $i = \overline{1,n}$. Так, коефіцієнти $b_0$ і $b_1$ рівняння регресії можуть бути оцінені за умови мінімізації однієї з таких сум:

$$\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} \left(y_i - \hat{b}_0 - \hat{b}_1 x\right);$$

$$\sum_{i=1}^{n} |\varepsilon_i| = \sum_{i=1}^{n} |y_i - \hat{y}_i| = \sum_{i=1}^{n} \left|y_i - \hat{b}_0 - \hat{b}_1 x\right|;$$

$$\sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{b}_0 - \hat{b}_1 x\right)^2.$$

Однак перша сума не може бути мірою якості знайдених оцінок через те, що існує безліч прямих (зокрема, $Y = \bar{y}$), для яких $\sum_{i=1}^{n} \varepsilon_i = 0$

Метод визначення оцінок коефіцієнтів за умови мінімізації другої суми називається *методом найменших модулів* (МНМ).

Найпоширенішим і теоретично обґрунтованим є метод визначення коефіцієнтів, при якому мінімізується третя сума. Він дістав назву методу найменших квадратів (МНК). МНК – найпростіший з обчислювальної точки зору. Крім того, оцінки коефіцієнтів регресії, знайдені цим методом при визначених передумовах мають ряд оптимальних властивостей (незміщеність, ефективність, обґрунтованість). Для знаходження невідомих параметрів моделі (1.3) використовуємо МНК**.** Основоположниками МНК є К.Гаусс (1777 – 1851) та П.Лаплас (1749 – 1827).

Розглянемо емпіричні дані, які містять $n$ пар значень $(x_i, y_i)$, кожна з яких спостерігалася один раз (це еквівалентно припущенню, що пари значень $(x_i, y_i)$, що спостерігаються, є рівноймовірними). Ці дані можна записуються у вигляді такої таблиці:

| | | | | | | |
|---|---|---|---|---|---|---|
| $X$ | $x_1$ | $x_2$ | … | $x_i$ | … | $x_n$ |
| $Y$ | $y_1$ | $y_2$ | … | $y_i$ | … | $y_n$ |

За МНК *невідомі параметри* $b_0$ *і* $b_1$ *вибираються таким чином, щоб сума квадратів відхилень емпіричних значень* $y_i$ *від значень* $y_i^{т} = b_0 + b_1 x_i$, *знайдених за рівнянням регресії* (1.3) (рис. 1.4) , *була мінімальною:*

$$S(b_0, b_1) = \sum_{i=1}^{n} (y_i^{т} - y_i)^2 = \sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2 \rightarrow \min.$$

Використовуємо необхідну умову екстремуму функції двох змінних. Знайдемо частинні похідні функції $S(b_o, b_1)$ за $b_0$ і $b_1$ і прирівняємо їх до нуля:

$$\begin{cases} \partial S / \partial b_0 = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i)(-1) = 0 \\ \partial S / \partial b_1 = 2\sum_{i=1}^{n}(b_0 + b_1 x_i - y_i)(-x_i) = 0, \end{cases}$$

Після перетворень одержимо систему *нормальних рівнянь*:

$$\begin{cases} b_1 \sum_{i=1}^{n} x_i^2 + b_0 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \\ b_1 \sum_{i=1}^{n} x_i + b_0 n = \sum_{i=1}^{n} y_i. \end{cases} \tag{1.4}$$

Розділивши обидва рівняння системи (1.4) на $n$, запишемо її у вигляді:

$$\begin{cases} b_1 \overline{x^2} + b_0 \overline{x} = \overline{xy}, \\ b_1 \overline{x} + b_0 = \overline{y}, \end{cases} \tag{1.5}$$

де $\overline{x^2} = \sum_{i=1}^{n} x_i^2 \Big/ n$, $\overline{x} = \sum_{i=1}^{n} x_i \Big/ n$, $\overline{y} = \sum_{i=1}^{n} y_i \Big/ n$, $\overline{xy} = (\sum_{i=1}^{n} x_i y_i) \Big/ n$.

Розв'язуючи систему рівнянь (1.5), одержимо

$$\begin{cases} b_0 = \overline{y} - b_1 \overline{x}. & (1.6) \\ b_1 = (\overline{xy} - \overline{x} \cdot \overline{y}) \Big/ (\overline{x^2} - (\overline{x})^2) = (\overline{xy} - \overline{x} \cdot \overline{y}) \Big/ s_x^2 = cov(x, y) / var(x), & (1.7) \end{cases}$$

де $var(x) = s_x^2 = \overline{x^2} - (\overline{x})^2$ – вибіркова дисперсія величини $X$;

$cov(x, y) = (\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})) \Big/ n = \overline{xy} - \overline{x} \cdot \overline{y}$ – *вибіркова коваріація між величинами x та y* (*вибірковий кореляційний момент*, або *вибіркова кореляція*).

Вибіркова дисперсія в середньому знижує значення теоретичної дисперсії. Оскільки вибіркове середнє знаходиться в центрі вибірки, то відхилення від неї в середньому менше відхилень від теоретичного середнього значення.

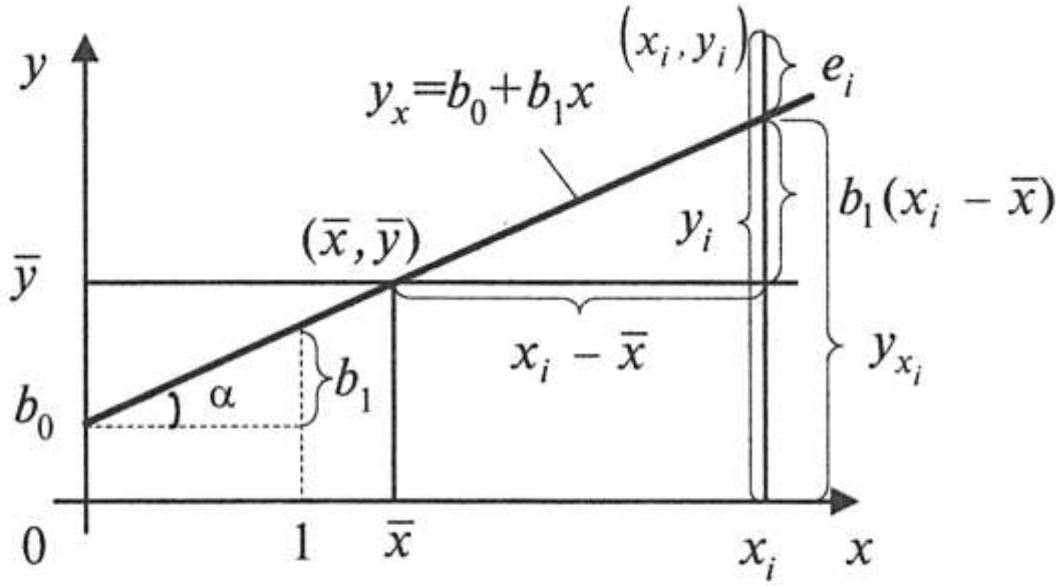



Рис. 1.4.

Вибіркова коваріація – це величина, що показує степень лінійної залежності між двома вибірками. Якщо значення вибіркової коваріації додатнє, то із зростанням значень однієї випадкової величини значення другої мають тенденцію зростання, а якщо значення вибіркової коваріації від'ємне – тенденцію спадання.

Очевидно, що

$$var(x) = cov(x,x),\ cov(x,y) = cov(y,x).$$

Для лінійно незалежних даних коваріація дорівнює нулю.

Вибіркова дисперсія і вибіркова коваріація мають від'ємне зміщення, тобто дають значення оцінок, які менші теоретичних величин. Незміщеними оцінками теоретичної дисперсії і теоретичної коваріації є відповідно оцінки $n \cdot var(x) / (n-1)$ і $n \cdot cov(x) / (n-1)$.

Доказ достатньої умови існування екстремуму (мінімуму) функції $S$ у точці $(b_0, b_1)$ наведено у [6],[24] та ін.

Таким чином, параметри $b_0$ (1.6) і $b_1$ (1.7) визначені. Рівняння теоретичної лінії регресії (1.3) можна записати в іншому вигляді, якщо використовувати формули (1.6) і (1.7):

$$\overline{y}_x = b_1 x + \overline{y} - b_1 \overline{x}; \qquad \overline{y}_x - \overline{y} = b_1 (x - \overline{x});$$

$$\overline{y}_x - \overline{y} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{s_x^2}(x - \overline{x}),$$

або

$$\overline{y}_x - \overline{y} = \frac{cov(x,y)}{var(x)}(x - \overline{x}). \tag{1.8}$$

З рівняння (1.8) як рівняння прямої, що проходить через дану точку в даному напрямку, випливає, що лінія регресії проходить через точку $(\overline{x}, \overline{y})$ і має кутовий коефіцієнт $b_1$, що позначається через $\rho_{y/x}$.

Отже, рівняння регресії $Y$ на $X$ має вигляд:

$$\overline{y}_x - \overline{y} = \rho_{y/x}(x - \overline{x}), \tag{1.9}$$

$$\rho_{y/x} = (\overline{xy} - \overline{x} \cdot \overline{y}) \big/ s_x^2 = cov(x,y) / var(x). \tag{1.10}$$

Величина $\rho_{y/x}$ називається *коефіцієнтом регресії $Y$ на $X$* .

Виведення рівняння теоретичної лінії регресії $Y$ на $X$ проводилося в припущенні однозначної відповідності між значеннями $x_i$ та $y_i$.

Аналогічно можна записати рівняння теоретичної лінії регресії $X$ на $Y$ у вигляді:

$$\bar{x}_y - \bar{x} = \rho_{x/y}(y - \bar{y}), \tag{1.11}$$

$$\rho_{x/y} = (\overline{xy} - \overline{x} \cdot \overline{y}) \big/ s_y^2 = cov(x,y) / var(y). \tag{1.12}$$

Величина $\rho_{x/y}$ називається *коефіцієнтом регресії $X$ по $Y$*.

Значення $\overline{x}_y = \overline{x}$ *и* $\overline{y}_x = \overline{y}$ задовольняють обом рівнянням регресії (1.9) і (1.11). Отже, лінії регресії перетинаються у точці $(\overline{x}, \overline{y})$.

## 1.5. Основні твердження регресійного аналізу. Оцінка параметрів парної регресійної моделі.

Як відзначалося вище, розглянута в регресійному аналізі залежність $Y$ на $X$ може бути записана у вигляді рівняння регресії (1.1). Завдяки впливу неврахованих випадкових факторів і причин, окремі спостереження змінної $Y$ будуть у більшій чи меншій мірі відхилятися від функції регресії $\varphi(X)$. У цьому випадку рівняння взаємозв'язку двох змінних (*парна регресійна модель*) може бути записане у вигляді:

$$Y = \varphi(X) + \varepsilon, \qquad (1.13)$$

де $\varepsilon$ – випадкова змінна, що називається *залишковою помилкою*, або *збурюванням*. Вона характеризує відхилення від функції регресії.

Таким чином, у регресійній моделі залежна змінна $Y$ є деякою функцією $\varphi(X)$ з точністю до випадкового збурювання $\varepsilon$.

Розглянемо лінійний регресійний аналіз, для якого функція $\varphi(X)$ є лінійною щодо оцінюваних параметрів:

$$M_x(Y) = \beta_0 + \beta_1 x\ . \qquad (1.14)$$

Припустимо, що для оцінки параметрів лінійної функції регресії (1.14) узята вибірка, що містить $n$ пар значень змінних $(x_i, y_i)$. Лінійна парна регресійна модель має вигляд:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ . \qquad (1.15)$$

*Основні припущення регресійного аналізу.*

1. У моделі (1.15) збурювання $\varepsilon_i$ (чи залежна змінна $y_i$) є величиною випадковою, а пояснююча змінна $x_i$ – величиною невипадковою.

2. Математичне сподівання збурювання $\varepsilon_i$ дорівнює нулю:

$$M(\varepsilon_i) = 0, \qquad (1.16)$$

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

3. Дисперсія збурювання $\varepsilon_i$ (або залежної змінний $y_i$) є сталою:

$$D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2 \qquad (1.17)$$

(або $D(y_i) = \sigma^2$) – умова *гомоскедастичності* збурювання.

4. Збурювання $\varepsilon_i$ і $\varepsilon_j$ (змінні $y_i$ і $y_j$) не корелюються між собою, тобто

$$M(\varepsilon_i, \varepsilon_j) = 0 \ \ (i \neq j). \tag{1.18}$$

При виконанні припущення 4 вимога некорельованості рівносильна незалежності змінних $\varepsilon_i$ і $\varepsilon_j$ ($y_i$ і $y_j$).

5. Збурювання $\varepsilon_i$ (або залежна змінна $y_i$) являє собою *нормально розподілену випадкову величину*.

В усіх передумовах $i = \overline{1,n}$. У цьому випадку модель (1.15) називається *класичною нормальною лінійною регресійною моделлю.*

Для одержання рівняння регресії досить виконання припущень 1 - 4.

Оцінкою моделі (1.15) за вибіркою є рівняння регресії (1.3). Параметри цього рівняння $b_0$ і $b_1$ визначаються на основі МНК.

Вплив неврахованих випадкових факторів і помилок спостережень у моделі (1.15) визначається за допомогою дисперсії або збурювань *залишкової дисперсії* $\sigma^2_{залиш}$. Незміщеною оцінкою цієї дисперсії є *вибіркова залишкова дисперсія* (або дисперсія випадкової складової):

$$\sigma^2 = \left(\sum_{i=1}^{n} (y_i^{\text{т}} - y_i)^2\right) \Big/ (n-2) = \left(\sum_{i=1}^{n} e_i^2\right) \Big/ (n-2), \tag{1.19}$$

де $e_i = y_i^{\text{Т}} - y_i$ - вибіркова оцінка збурювання $\varepsilon_i$, або *залишок регресії*.

Виникає питання, чи є оцінки $b_0$, $b_1$, $\sigma^2$ параметрів $\beta_0$, $\beta_1$ і $\sigma^2$ «найкращими»? Відповідь на це питання дає така теорема.

*Теорема Гаусса-Маркова.* Якщо регресійна модель (1.15) задовольняє припущенням 1 - 4, то оцінки $b_0$ (1.6) і $b_1$ (1.7) мають найменшу дисперсію в класі всіх лінійних незміщених оцінок.

Таким чином, найкращі оцінки $b_0$ і $b_1$, отримані за допомогою МНК, мають такі властивості:

1) *незміщеність*: $M(b_0) = \beta_0$, $M(b_1) = \beta_1$.

(означає відсутність систематичної помилки в положенні лінії регресії);

2) *ефективність*: мають найменшу дисперсію в класі всіх лінійних незміщених оцінок, яка дорівнює

$$D(b_0) = \overline{x^2}\sigma^2 \big/ (n \cdot var(x)), \quad D(b_1) = \sigma^2 \big/ (n \cdot var(x));$$

3) *обґрунтованість*: $\lim\limits_{n\to\infty} D(b_0) = 0,\ \lim\limits_{n\to\infty} D(b_1) = 0$.

(при досить великому $n$ оцінки $b_0$ і $b_1$ близькі відповідно $\beta_0$ і $\beta_1$).

Розглянемо докладніше припущення 1 - 4, що лежать в основі регресійного аналізу.

*Перше припущення* є особливо важливим. Якщо воно не виконується, то оцінки коефіцієнтів регресії виявляються *зміщеними* і *необґрунтованими*. Порушення цього припущення може бути зв'язане з помилками виміру змінних або з використанням лаговых змінних.

У регресійному аналізі замість першого припущення часто використовується більш слабка *умова про незалежність розподілів* (*некорельованість*) *пояснюючої змінної та випадкового збурювання*. Одержувані при цьому оцінки коефіцієнтів регресії мають такі ж самі основні властивості.

*Друге припущення* означає, що збурювання не повинне мати систематичного зміщення. Якщо постійний член включений у рівняння регресії, то ця умова виконується автоматично.

*Третє припущення* означає, що дисперсія відхилення в кожному спостереженні має тільки одне значення.

Під дисперсією $\sigma^2$ мається на увазі можливе поводження збурювання до того, як зроблена вибірка. Величина $\sigma^2$ невідома, і одна із задач регресійного аналізу складається в тому, щоб її оцінити.

Умова *незалежності* дисперсії випадкового відхилення від номера спостереження називається *гомоскедастичністю* (що означає однаковий розкид): $D(\varepsilon_i) = \sigma^2$. *Залежність* дисперсії випадкового відхилення від номера спостереження, у свою чергу, називається *гетероскедастичністю*: $D(\varepsilon_i) = \sigma_i^2$. Характерні діаграми розсіювання для випадків гомоскедастичності та гетероскедастичності показані на рис. 1.5 а) і б) відповідно.

Якщо умова гомоскедастичності не виконується, то оцінки коефіцієнтів кореляції будуть *неефективними*, хоча і *незміщеними*.

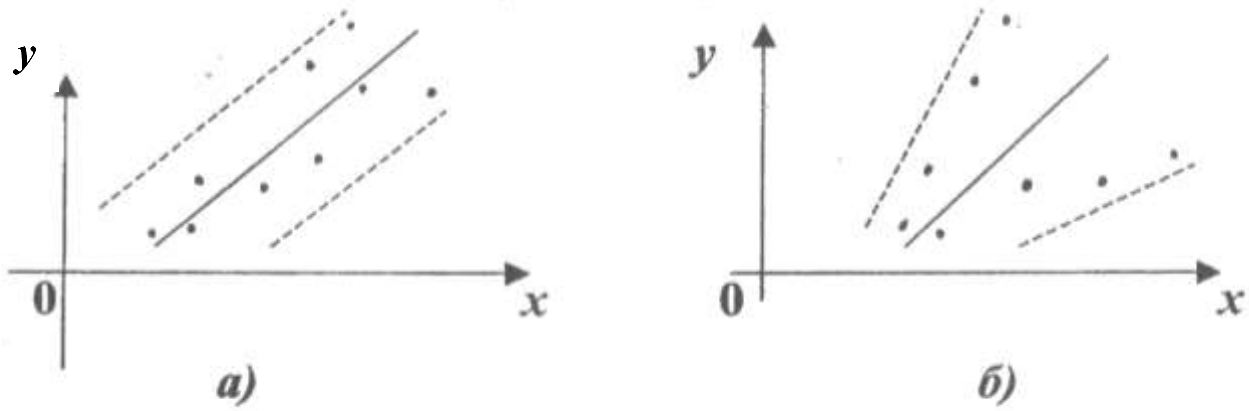



Рис. 1.5

Існують спеціальні методи діагностування й усунення гетероскедастичності.

*Четверте припущення* вказує на некорельованість випадкових відхилень для різних спостережень. Ця умова часто порушується, якщо дані є часовими рядами. У випадку його невиконання має місце *автокореляція залишків*.

Автокореляція – це кореляція між значеннями результуючої змінної, яка виникає унаслідок залежності значень випадкової величини в різних постереженнях. Головними причинами автокореляції можуть бути помилка специфікації, отримання числових даних при дослідженні економічного явища з великими похибками, інерційність в зміні економічних показників, ефект павутини (у багатьох сферах економічні показникі реагують на зміну економічних умов із запізненням – числовим лагом), згладжування даних, перетворення початкової специфікації моделі до лінійної форми. Типовий вигляд даних при наявності автокореляції показаний на рис. 1.6. Якщо умова незалежності збурювань не виконується, то оцінки коефіцієнтів регресії, отримані за МНК, виявляються *неефективними*, хоча і *незміщеними*. Існують методи діагностування й усунення автокореляції.

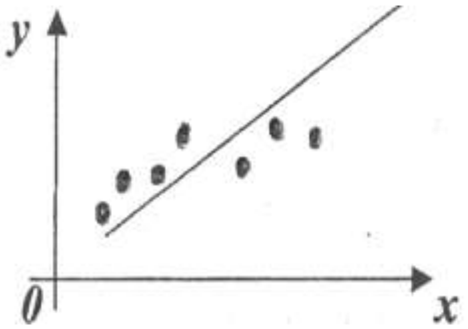



Рис.1.6.

*П'яте припущення* необхідне для інтервальної оцінки функції регресії і її параметрів.

Дотепер були використані оцінки параметрів, отримані за допомогою МНК. Розглянемо ще один метод одержання оцінок, який використовується у економетриці, - *метод максимальної правдоподібності* (ММП).

Для застосування ММП треба мати *вигляд закону розподілу імовірностей* наявних вибіркових даних.

Припустимо, що виконано *п'яте припущення* регресійного аналізу. Будемо розглядати значення $y_i$ як незалежні нормально розподілені випадкові величини з математичним сподіванням $M(y_i) = \beta_0+\beta_1 x_i$ , що є функцією від $x_i$ , і постійною дисперсією $\sigma^2$. Щільність нормально розподіленої випадкової величини $y_i$ має вигляд:

$$\varphi_N(y_i) = e^{-(y_i-\beta_0-\beta_1 x_i)^2 / (2\sigma^2)} \Big/ (\sigma\sqrt{2\pi})$$

Функція правдоподібності (від англ. likelihood – правдоподібність), що виражає щільність імовірності спільної появи результатів вибірки, має вигляд:

$$L(y_1, x_1; ...; y_n, x_n; \beta_0, \beta_1; \sigma^2) = \prod_{i=1}^{n} e^{-\sum\limits_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2 / 2\sigma^2)} \Big/ (\sigma\sqrt{2\pi})^n$$

Згідно з ММП, оцінками параметрів $\beta_0$ , $\beta_1$ і $\sigma^2$ вважаються значення $\hat{\beta}_0, \hat{\beta}_1$ і $\hat{\sigma}^2$, що максимізують функцію правдоподібності $L$.

При заданих значеннях $x_1, x_2, \ldots, x_n$ пояснюючої змінної $X$ и постійної дисперсії $\sigma^2$ функція правдоподібності $L$ досягає максимуму, коли показник степені при експоненті буде мінімальним за абсолютною величиною, тобто за умови мінімуму функції $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$, що збігається з умовою

$$S(b_0, b_1) = \sum_{i=1}^{n}(y_i^{\text{т}} - y_i)^2 = \sum_{i=1}^{n}(b_0 + b_1 x_i - y_i)^2 \to \min$$

знаходження оцінок $b_0$ і $b_1$ за допомогою МНК.

Отже, оцінки $b_0$ (1.6) і $b_1$ (1.7) параметрів $\beta_0$ і $\beta_1$ збігаються з оцінками ММП.

Знайдемо частинну похідну функції правдоподібності за $\sigma^2$ і прирівняємо до нуля:

$$\frac{\partial L}{\partial \sigma^2} = \frac{1}{2\sigma^4}\prod_{i=1}^{n}\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 - n\sigma^2}{(\sigma\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \Big/ (2\sigma^2)} = 0,$$

звідки

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2,$$

де параметри $\beta_0$ і $\beta_1$ замінені їх оцінками $b_0$ і $b_1$ . Порівнюючи $\hat{\sigma}^2$ з отриманою раніше незміщеною оцінкою $s^2$ (1.19), бачимо, що оцінка $\hat{\sigma}^2$ є *зміщеною*. Відповідно до властивостей оцінок ММП оцінки ($b_0$, $b_1$) і $\hat{\sigma}^2$ (а також і $\sigma^2$) є обґрунтованими оцінками. При виконанні п'ятого припущення регресійного аналізу про нормальний закон розподілу збурювання $\varepsilon_i$ $(i = \overline{1, n})$ ці оцінки є *незалежними*.

## 1.6. Оцінка адекватності лінійної регресії

### 1. 6. 1. Коефіцієнт кореляції

Головним недоліком коваріації як показника зв’язку є те, що його значення залежить від одиниць виміру вихідних даних та не має критичних значень, що ускладнює порівняння сили зв’язку різних сукупностей і робить неможливим встановлення критичних значень коваріації. Цей недолік усуває такий показник сили зв’язку, як коефіцієнт кореляції. Він є відносним показником сили зв’язку та також характеризує силу і напрямок лінійного зв’язку двох ознак.

Із рівностей (1.10) і (1.11) можна бачити, що коефіцієнти регресії $\rho_{y/x}$ і $\rho_{x/y}$ мають однакові знаки, тому що чисельники дробів рівні, а знаменники – додатні величини. Отже, знаки $\rho_{y/x}$ і $\rho_{x/y}$ збігаються зі знаком величини коваріації $cov(x,y)$. Для уточнення коефіцієнтів регресії як показників тісноти зв'язку потрібна така стандартна система одиниць вимірювання, в якій дані, що мають різні характеристики, виявилися б порівняними між собою.

Оцінка тісноти зв'язку залежить від величини кута $\varphi$, що утворюється прямими регресії. Чим менше цей кут $\varphi$, тим тісніше кореляційний зв'язок між $X$ і $Y$. Для зручності рис.1.8 зображує випадок прямої кореляційної залежності, тобто, коли $\rho_{y/x} > 0$ і $\rho_{x/y} > 0$.

Для більшої наочності на рис. 1.7 показане положення прямих регресії щодо нової системи координат з початком у точці $O_1(\bar{x}, \bar{y})$– перетину цих прямих. Тут: $k_1 = \rho_{y/x} = tg\ \alpha$ де $\alpha$ - кут, що утворений прямою $l_1$ з віссю $Ox$; $k_2 = 1 / \rho_{x/y} = tg\ \beta$, де $\beta$ - кут, що утворений прямою $l_2$ з віссю $Ox$.

Кут між двома прямими обчислюється за формулою:

$$tg\ \varphi = (k_2 - k_1) / (1 + k_1 \cdot k_2) = (1 - \rho_{y/x}\, \rho_{x/y})(\rho_{y/x} + \rho_{x/y}) \qquad (1.20)$$

З формули (1.20) випливає, що при $1 - \rho_{y/x}\, \rho_{x/y} \to 0$ величина кута $\varphi \to 0$ і навпаки.

Показником тісноти лінійної кореляційної залежності є *лінійний коефіцієнт кореляції*

$$r = \pm\sqrt{\rho_{y/x} \cdot \rho_{x/y}} \qquad (1.21)$$

Маючи на увазі (1.10) і (1.12), одержуємо такий вираз лінійного коефіцієнта кореляції:

$$r = \frac{cov(x, y)}{\sqrt{var(x) \cdot var(y)}} \,. \qquad (1.22)$$

Перед дробом залишений тільки один знак «+», тому що збіг знаків *r* і коефіцієнтів регресії вже забезпечено.

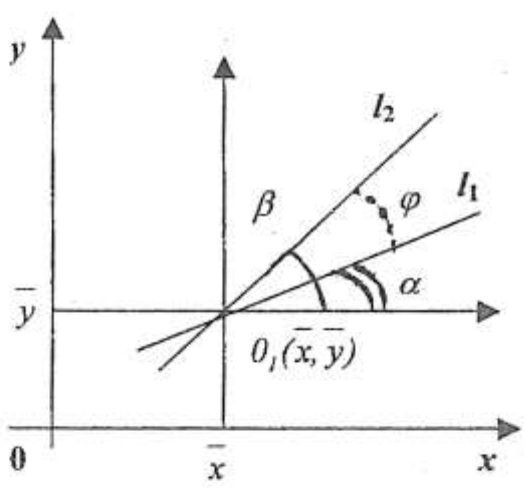



Рис. 1.7

Формула (1.22) дозволяє виразити кожний коефіцієнт регресії через лінійний коефіцієнт кореляції і навпаки. Дійсно,

$$\rho_{y/x} = \frac{cov(x, y)}{var(x)} = \sqrt{var(y)/var(x)}\,\frac{cov(x, y)}{\sqrt{var(x) \cdot var(y)}} = r\sqrt{var(y)/var(x)},$$

звідки

$$r = \rho_{y/x}\sqrt{var(x)/var(y)}. \qquad (1.23)$$

Аналогічно

$$\rho_{x/y} = \frac{cov(x, y)}{var(y)} = \sqrt{var(x)/var(y)}\,\frac{cov(x, y)}{\sqrt{var(x) \cdot var(y)}} = r\sqrt{var(x)/var(y)},$$

звідки

$$r = \rho_{x/y}\sqrt{var(y)/var(x)}. \qquad (1.24)$$

Формули (1.23) і (1.24) мають самостійне значення. З їх допомогою рівняння лінійної регресії $Y$ на $X$ і $X$ на $Y$ можна записати у вигляді:

$$\overline{y}_x - \overline{y} = r\sqrt{var(y)/var(x)}\,(x - \overline{x}), \tag{1.25}$$

$$\overline{x}_y - \overline{x} = r\sqrt{var(x)/var(y)}\,(y - \overline{y}). \tag{1.26}$$

Величину

$$r = \rho_{y/x}\sqrt{var(x)/var(y)} = \rho_{x/y}\sqrt{var(y)/var(x)}$$

іноді називають *лінійним коефіцієнтом кореляції*.

*Властивості лінійного коефіцієнта кореляції*

1) Абсолютна величина лінійного коефіцієнта кореляції не перевищує одиниці: $|r| \le 1$, чи $-1 \le r \le 1$.

Для якісної оцінки тісноти зв'язку можна скористатися такою таблицею:

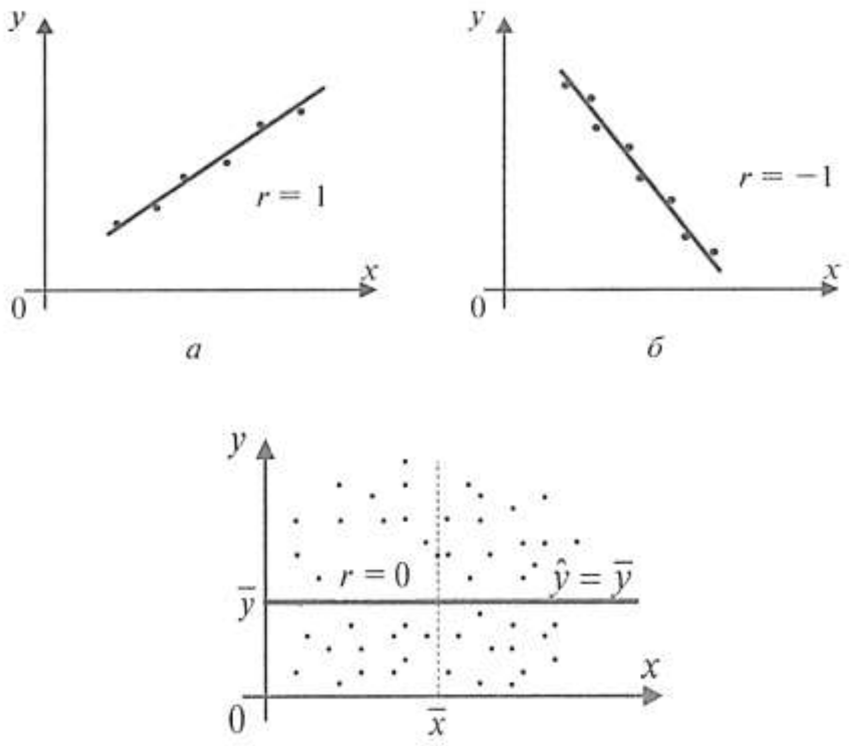
| Величина $\|r\|$ | Характеристика тісноти зв'язку |
|---|---|
| до 0,5 | дуже слабкий зв'язок |
| 0,5 - 0,6 | слабкий зв'язок |
| 0,6 - 0,7 | помірний зв'язок |
| 0,7 - 0,8 | середній зв'язок |
| 0,8 - 0,9 | тісний зв'язок |
| більш 0,9 | дуже тісний зв'язок |

2) Якщо всі значення змінних збільшити (зменшити) на одне й те саме число або в одне й те саме число раз, то величина $r$ не зміниться.

3) При $r = -1$ або $r = 1$ кореляційний зв'язок являє собою *лінійну функціональну залежність*, тобто варіація $Y$ обумовлена тільки впливом $X$. При цьому теоретичні лінії регресії $Y$ на $X$ і $X$ на $Y$ збігаються і всі емпіричні значення величин, що спостерігаються, розташовуються на спільній прямій (рис. 1.8 а), б)). Рівність лінійного коефіцієнта кореляції 1 чи –1 є необхідною і достатньою ознакою лінійної функціональної залежності між $X$ та $Y$.

4) При $r = 0$ лінійна кореляційна залежність відсутня, тому що в цьому випадку, принаймні, один з кутів $\alpha$ чи $\beta$ дорівнює нулю. При цьому групові середні змінних збігаються із їх загальними середніми, а теоретичні лінії регресії $Y$ на $X$ і $X$ на $Y$ паралельні осям координат і кут між ними $\varphi = 90^0$.

Рівність $r = 0$ говорить лише про *відсутність лінійної кореляційної залежності, але не взагалі про відсутність кореляційної, а тим більше статистичної залежності.*




*в*)

Рис. 1.8 [5]

Так, наприклад, для залежностей, що зображені на рис. 1.9 а) і б) *r=0* і лінії регресії $Y$ на $X$ паралельні осі абсцис. Однак за розташуванням точок кореляційного поля чітко проглядається взаємозв'язок між змінними, який є відмінним від лінійної кореляційної. Так, у випадку а) – це нелінійна кореляційна (майже функціональна) залежність; у випадку б) – статистична залежність, що виявляється в даному випадку в тому, що зі зміною $x$ групові середні $\bar{y}_x$ не змінюються, а змінюється лише розсіювання точок поля щодо лінії регресії.

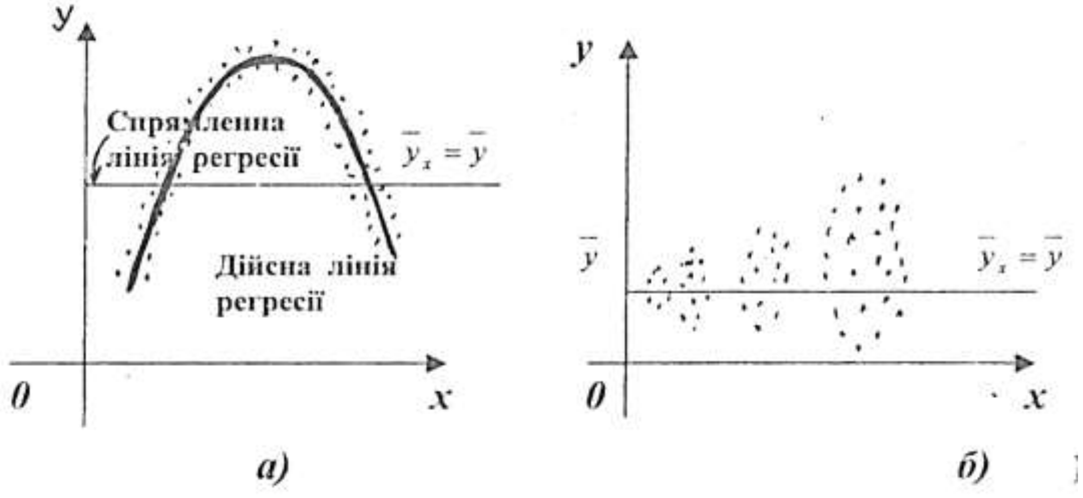



Рис. 1.9 [5]

Вибірковий коефіцієнт кореляції $r$ був уведений виходячи з оцінки близькості точок кореляційного поля до прямої регресії $Y$ на $X$. Однак $r$ є безпосередньо *оцінкою генерального коефіцієнта кореляції* між $X$ і $Y$ лише у випадку двовимірного нормального закону розподілу випадкових величин $X$ і $Y$. У випадках, коли розподіли $X$ та $Y$ відхиляються від нормального, або коли одна з досліджуваних величин не є випадковою, вибірковий коефіцієнт кореляції не треба розглядати як строгу міру взаємозв'язку змінних.

### 1.6.2. Індекс кореляції. Коефіцієнт детермінації

Найзагальнішим поняттям і показником тісноти кореляційного зв'язку служить *індекс кореляції* (*кореляційне відношення*):

$$i_r = \sqrt{\sigma^2_{регр} / \sigma^2_{заг}} = \sqrt{(\sigma^2_{заг} - \sigma^2_{залиш}) / \sigma^2_{заг}} = \sqrt{1 - \sigma^2_{ост} / \sigma^2_{заг}}. \quad (1.27)$$

$$\sigma^2_{заг} = \sigma^2_{регр.} + \sigma^2_{залиш.}\ , \quad (1.28)$$

$\sigma^2_{заг} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \Big/ n$ - загальна дисперсія результуючої змінної, що характеризує *міру відхилень всіх емпіричних* (*табличних*) *значень* $y_i$ *від середнього значення* $\bar{y}$;

$\sigma^2_{регр} = \sum_{i=1}^{n}(y_i^{т} - \overline{y})^2 \Big/ n$ - дисперсія теоретичних значень результуючої змінної, яка пояснює регресію, що характеризує *відхилення розрахункових значень* $y_i^{т} = \overline{y}_i(x_i) = b_0 + b_1 x_i$ *від середнього значення* $\overline{y}$ і показує *вплив пояснюючої змінної X на залежну змінну*;

$\sigma^2_{залиш} = (\sum_{i=1}^{n}(y_i^{т} - y_i)^2) \Big/ n$ - дисперсія випадкових відхилень (помилок), або *залишкова* (непояснена) *дисперсія*, що характеризує *відхилення розрахункових* (*теоретичних*) *значень* $y_i^{т}$ *від емпіричних значень* $y_i$ і показує *вплив на залежну змінну всіх тих факторів, що не увійшли в рівняння регресії.*

Рівність (1.28) називають формулою *декомпозиції загальної дисперсії*: загальна дисперсія результуючої змінної подається як сума поясненої та непоясненої дисперсій.

Порівнюючи у формулі декомпозиції загальної дисперсії пояснену та непояснену складові, можна зробити висновок: чим більша пояснена дисперсія і, відповідно, менша непояснена, тим точніше кореляційно-регресійна модель пояснює зв'язок між змінними. І навпаки: чим менша пояснена дисперсія і, відповідно, більша непояснена, тоді більша у варіації результуючої змінної складова випадкових відхилень, які неможливо пояснити кореляційно-регресійною моделлю. Отже, кожна зі складових загальної дисперсії результуючої змінної є характеристикою точності кореляційно-регресійної моделі.

*Стандартною похибкою* кореляційно-регресійної моделі (стандартною похибкою оцінки за рівнянням регресії) називають величину $\sigma_{залиш}$.

Стандартна позибка моделі характеризує розсіювання фактичних значень результуючої змінної навколо теоретичних, знайдених за рівнянням регресії.

Символ $i_r$ підкреслює, що індекс кореляції, узагалі кажучи, використовується для оцінки тісноти зв'язку *парної нелінійної кореляційної залежності*. Формула (1.27) є характеристикою тісноти зв'язку, як для парної, так і для множинної, як для лінійної, так і для нелінійної форм зв'язку між ознаками. В окремому випадку – при лінійному кореляційному зв'язку – він виступає у формі різних виразів лінійного коефіцієнта кореляції.

Помноживши обидві частини формули декомпозиції загальної дисперсії (1.28) на *n*, отримаємо рівність для сум квадратів

$$SST = SSR + SSE.$$

Відповідні суми квадратів позначають:

$$SST = n\sigma^2_{\text{заг}} \quad (\textit{sum square total}),$$

$$SSR = n\sigma^2_{\text{регр}} \quad (\textit{sum square regression}),$$

$$SSE = n\sigma^2_{\text{залиш}} \quad (\textit{sum square error}).$$

Можливі два крайні випадки: $SST = SSE$; $SST = SSR$. У першому випадку ($SSR = 0$) фактор $\boldsymbol{X}$ не впливає на результат, вся дисперсія $\boldsymbol{Y}$ обумовлена впливом інших факторів, лінія регресії паралельна осі О*x* і $\bar{y} = y^{\mathrm{T}}$. У другому випадку ($SSE = 0$) інші фактори, що не включені в модель, не впливають на результат, тобто залежний показник $\boldsymbol{Y}$ пов'язаний із фактором $\boldsymbol{X}$ функціонально.

Однак на практиці у правій частині рівності сум квадратів присутні обидва доданків. Придатність лінії регресії для прогнозу залежать від того, яка частина загальної варіації *Y* припадає на пояснену варіацію.

Якщо $SSR > SSE$, то рівняння регресії статистично значимо і фактор $\boldsymbol{X}$ істотно впливає на результат $\boldsymbol{Y}$. Це рівносильне тому, що коефіцієнт детермінації буде наближатися до одиниці .

Оцінку якості побудованої моделі дає також *коефіцієнт* (*індекс*) *детермінації* $R^2$, що є однією з найефективніших оцінок адекватності регресійної моделі, мірою якості рівняння регресії (або, як кажуть, мірою якості «припасування» регресійної моделі до значень $y_i$, що *спостерігаються*), характеристикою прогностичної сили аналізованої регресійної моделі.

Коефіцієнт детермінації визначається за однією з формул:

$$R^2 = 1 - SSE/SST = 1 - \sigma^2_{залиш} \Big/ \sigma^2_{заг} = 1 - \sum_{i=1}^{n}(y_i^{т} - y_i)^2 \Big/ \sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$R^2 = SSR/SST = \sigma^2_{регр} \Big/ \sigma^2_{заг} = \sum_{i=1}^{n}(y_i^{т} - \bar{y})^2 \Big/ \sum_{i=1}^{n}(y_i - \bar{y})^2.$$

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

*Властивості коефіцієнта детермінації $R^2$.*

1. Оскільки $0 \le \sigma^2_{регр} \le \sigma^2_{заг}$, то $0 \le R^2 \le 1$.

2. Чим ближче $R^2$ до одиниці, тим краще регресія апроксимує емпіричні дані, тим тісніше спостереження примикають до лінії регресії. Якщо $R^2=1$**,** то емпіричні точки ($x_i$, $y_i$) лежать на лінії регресії (рис. 1.9 а) і б)), між змінними $\boldsymbol{Y}$ і $\boldsymbol{X}$ існує лінійна функціональна залежність.

3. Якщо $R^2=0$, то варіація залежної змінної цілком обумовлена впливом неврахованих у моделі змінних, і лінія регресії паралельна осі абсцис (рис. 1.9 в)).

Коефіцієнт $R^2$ має сенс розглядати тільки при *наявності вільного члена в рівнянні регресії.* У загальному випадку коефіцієнт детермінації може бути і відємним. Це свідчить про крайню неадекватність моделі: просте середнє наближує краще. У випадку парної лінійної регресійної моделі коефіцієнт детермінації дорівнює квадрату лінійного коефіцієнта кореляції або індексу кореляції, тобто $R^2 = r^2 = i_r^2$.

## 1.6.3. Поняття *F*- критерію Фішера. Оцінка значущості рівняння регресії

Адекватність простої лінійної регресійної моделі можна перевірити за допомогою коефіцієнта детермінації. Якщо його значення близькі до одиниці, то можна вважати, що модель адекватна. Якщо ж його значення близьке до нуля, то модель неадекватна, тобто не має лінійного зв'язку між залежною і незалежною змінними. Однак, який висновок можна зробити, якщо значення коефіцієнта кореляції має нечітко виражене граничне значення, наприклад, 0,5, 0,45 і т.д. У таких випадках складно зробити однозначний висновок про наявність зв'язку. Виникає необхідність в іншому критерії, який би однозначно відповідав на запитання про адекватність побудованої моделі. Одним із таких розповсюджених критеріїв є критерій Фішера.

*Поняття про ступені вільності*

Розглянемо тотожність, яка зв’язує суми квадратів:

$$SST = SSR + SSE.$$

Кожна сума квадратів пов’язана з числом, яке називають її *ступенем вільності*. Це число показує кількість потрібних для розрахунку даної суми квадратів незалежних елементів інформації при спостереженні елементів $y_1, y_2, \dots, y_n$. Розглянемо, скільки ступенів вільності має кожна сума квадратів.

Число ступенів вільності (*df - degrees of freedom*) – це число незалежно варійованих значень ознаки. Для *SST* потрібно ($n$ – 1) незалежних відхилень ($df_{заг}$ = $n$ – 1), оскільки $\sum_{i=1}^{n}(y_i - y)^2 = 0$. Вільно варіюються ($n$ – 1) значень, а останнє $n$-е відхилення визначається із загальної суми, яка дорівнює нулю.

Суму квадратів, що пояснює регресію (*SSR*), має одну ступінь вільності $df_{регр}$ = 1. Дійсно, використовується тільки одна незалежна одиниця інформації, яка утворюється з $y_1, y_2, \dots, y_n$, а саме $b_1$. Суму *SSR* можна записати у вигляді:

$$SSR = \sum_{i=1}^{n}(y_i^T - \bar{y})^2 = \sum_{i=1}^{n}[(b_0 + b_1 x) + (b_0 + b_1 \bar{x})]^2 = \sum_{i=1}^{n}(b_1 x - b_1 \bar{x})^2$$

$$= b_1^2 \sum_{i=1}^{n}(x - \bar{x})^2.$$

Сума квадратів помилок ($SSE$) має ($n$ – 2) ступенів вільності $df_{залиш} = n - 2$. Ця сума базується на кількості ступенів вільності, яка дорівнює різниці між кількістю спостережень $n$ і кількістю параметрів $m$, що оцінюються. У разі простої лінійної регресії оцінюються два параметри $b_0$ та $b_1$ (тобто $m$ = 2).

Можна записати рівність і між числами ступенів вільності:

$$df_{заг} = df_{регр} + df_{залиш}, \qquad (n - 1) = 1 + (n - 2).$$

*Рівень значущості* α *– імовірність відкинути правильну гіпотезу за умови, що вона правильна.*

Рівняння регресії є значущим на рівні α, якщо фактичне значення статистики, що спостерігається

$$F_{факт} = \frac{\sigma^2_{\text{регр}}\,(n-m)}{\sigma^2_{\text{зал}}\,(m-1)} > F_{\alpha;k_1;k_2}, \tag{1.29}$$

де $F_{\alpha;k_1;k_2}$ – табличне значення *F-критерію* Фішера-Снедекора, що визначене на рівні значущості α при $k_1 = m - 1$ і $k_2 = n - m$ ступінях вільності. Критичні значення *F-критерію* Фішера-Снедекора наведені в таблиці 1 Додатку.

Якщо відомі індекс кореляції $i_r^2$ або коефіцієнт детермінації $R^2$, то критерій значущості (1.29) рівняння регресії або самого коефіцієнта детермінації може бути записаний у вигляді:

$$F_{факт} = \frac{R^2(n-m)}{(1-R^2)(m-1)} = \frac{i_r^2(n-m)}{(1-i_r^2)(m-1)} > F_{\alpha;k_1;k_2}.$$

У випадку парної лінійної регресії ($m$ = 2) рівняння регресії є значущим на рівні α, якщо

$$F_{факт} = \frac{r^2}{1-r^2}(n-2) > F_{\alpha;1;n-2}.$$

$F_{\alpha;1;n-2} = F_{\text{табл.}}$ – це *максимально можливе значення критерію під впливом випадкових факторів при даних ступінях вільності та рівні значущості* α. На практиці α приймається рівним 0,05 або 0,01. Якщо $F_{\text{табл}} < F_{факт}$, то $H_0$ – гіпотеза про випадкову природу оцінюваних характеристик *відхиляється* і визнається їх *статистична значущість і надійність*. Якщо $F_{\text{табл}} > F_{факт}$, то гіпотеза $H_0$ *не відкидається і визнається статистична незначущість, ненадійність рівняння регресії.*

*Перевірити значущість рівняння регресії – установити, чи відповідає математична модель експериментальним даним і чи достатньо включених у рівняння пояснюючих змінних* (*однієї чи декількох*) *для опису залежної змінної.*

*Значення статистики F показує, якою мірою регресія краще оцінює значення залежної змінної в порівнянні з її середньою*.

Значущість рівняння парної лінійної регресії можна також перевірити за допомогою оцінки значущості коефіцієнта регресії $b_1$, що має $t$-*розподіл* Ст'юдента з $k = n - 2$ ступінями вільності. Критичні значення $t$-статистик Ст'юдента наведені в таблиці 2 Додатку.

Для *парної лінійної моделі обидва способи* перевірки значущості з використанням $F$- *і* $t$-критеріїв *рівносильні*, тому що ці критерії зв'язані співвідношенням $F = t^2$.

У багатьох економічних задачах потрібно оцінити *значущість коефіцієнта кореляції* $r$. При цьому виходять з того, що при відсутності кореляційного зв'язку статистика (*великий обсяг вибірки*, $n > 30$)

$$t = r\sqrt{n-2}\Big/\sqrt{1-r^2} \tag{1.30}$$

має $t$-*розподіл* Ст'юдента з $n - 2$ ступінями вільності.

Коефіцієнт кореляції $r$ значущий на рівні α, якщо

$$|t| = |r|\sqrt{n-2}\Big/\sqrt{1-r^2} > t_{1-\alpha;n-2},$$

де $t_{1-\alpha;n-2}$ - табличне значення $t$- критерію Ст'юдента, що визначене на рівні значущості α при числі ступіней вільності $n - 2$.

При *малому обсязі вибірки* ($n > 30$) використовують статистику:

$$t = r\sqrt{n-1}.$$

### 1.6.4. Інтервальна оцінка функції регресії і її параметрів

*Надійний інтервал для параметрів регресійної моделі*

Згідно з методом Гаусса, потрібно зробити мінімальну різницю між емпіричними значеннями $y_i$ й обчисленими з рівняння $y_i^{т} = b_0 + b_1 x_i \, (i = \overline{1,n})$ теоретичними значеннями регресанта. Для цього використовується МНК; тоді найкращий можливий коефіцієнт регресії $b_1$ знаходиться за формулою $b_1 = r\sqrt{var(x)/var(y)}$, а коефіцієнт $b_0$ - за формулою $b_0 = \overline{y} - b_1\overline{x}$.

Константи $b_0$ і $b_1$ – вибіркові оцінки теоретичних параметрів. Унаслідок цього адекватність моделі визначається надійними інтервалами для них.

Спочатку за формулою (1.19) обчислюють дисперсію та середнє квадратичне відхилення різниці між теоретичними значеннями регресанта і експериментальними даними

$$\sigma = \sqrt{\sum_{i=1}^{n}(y_i^{т} - y_i)^2 \Big/ (n-2)}. \tag{1.31}$$

Із ймовірністю γ = 1 - α (або рівнем значущості α) *інтервальна оцінка параметра $b_1$* має вигляд:

$$\beta_1 \in (b_1 - \Delta b_1,\ b_1 + \Delta b_1),$$

де $b_1$ – знайдена точкова оцінка параметра $\beta_1$,

$$\Delta b_1 = t_{1-\alpha,n-2} \cdot \sigma_{b_1} = t_{1-\alpha,n-2} \cdot \sqrt{\sigma^2_{залиш}/((n-2)var(x))}. \tag{1.32}$$

Тут

$t_{1-\alpha;n-2}$ – значення критерію Ст'юдента, що знаходиться в таблиці 2 Додатка;

$$\sigma^2_{залиш} = \sum_{i=1}^{n}(y_i^{т} - y_i)^2 \Big/ n;$$

$n$ – число пар значень вибірки;

$$var(x) = \overline{x^2} - \overline{x}^2.$$

Надійний інтервал для параметра $\beta_1$ має вигляд:

$$b_1 - t_{1-\alpha,n-2}\sqrt{\sigma^2_{залиш}/((n-2)var(x))} \le \beta_1 \le b_1 + t_{1-\alpha,n-2}\sqrt{\sigma^2_{залиш}/((n-2)var(x))} \quad (1.33)$$

Надійний інтервал для параметра $\beta_0$ знаходиться за формулою:

$$\beta_0 \in (\, b_0 - \Delta b_0\, ;\, b_0 + \Delta b_0\,)\,,$$

де $b_0$ – знайдена точкова оцінка параметра $\beta_0$,

$$\Delta b_0 = t_{1-\alpha,n-2}\cdot\sigma_{b_0} = t_{1-\alpha,n-2}\cdot\sqrt{\sigma^2_{залиш}\sum_{i=1}^{n}x_i^2 \Big/ (n\cdot(n-2)var(x))}. \quad (1.34)$$

*Інтервальна оцінка параметра* $b_0$ на рівні значущості α має вигляд:

$$b_0 - \left(t_{1-\alpha,n-2}\,\sigma\sqrt{\sum_{i=1}^{n}x_i^2}\right) \Big/ \sqrt{n\cdot var(x)} \le \beta_0 \le b_0 + \left(t_{1-\alpha,n-2}\,\sigma\sqrt{\sum_{i=1}^{n}x_i^2}\right) \Big/ \sqrt{n\cdot var(x)} \quad (1.35)$$

*Прогнозуванням* називається наукове передбачення імовірнісних шляхів розвитку явищ і процесів для більш-менш віддаленого майбутнього. Прогнозування, що базується на збереженні загальної тенденції розвитку явищ у часі, можна звести до підбору аналітичних виразів типу $y^{т} = \overline{y}_x = f(X)$ за даними за минулий період і екстраполяції отриманих залежностей.

Прогноз показника одержують шляхом підстановки в знайдене рівняння регресії значень фактора. Результатом є *точкова оцінка середнього значення показника* при даних рівнях факторів. Середнє значення прогнозу показника $y_0^*$ при значенні фактора $x_0$ у випадку лінійної регресії визначається за формулою:

$$\overline{y}_0 = b_0 + b_1 x_0\,.$$

Однією з центральних задач економетричного моделювання є прогнозування значень залежної змінної при певних значеннях пояснювальних змінних. Така задача формулюється як:

1) прогнозування значень умовного математичного сподівання $\overline{y}_x = M_x(Y)$ (*середнього значення*) залежної змінної при певних значеннях пояснювальної змінної;

2) прогнозування певного *конкретного значення* залежної змінної.

*Надійний інтервал для функції регресії*

Побудуємо надійний інтервал для функції регресії, тобто для умовного математичного сподівання $\bar{y}_x = M_x(Y)$, що із заданою надійністю (довірчою імовірністю) $\gamma$ покриває невідоме значення $\bar{y}_x$. Довірчий інтервал для умовного математичного сподівання $M_x$ $(Y)$ з рівнем значущості $\alpha$ має вигляд:

$$M_x(Y) = \beta_0+\beta_1 x+\varepsilon \in ( y^{\text{т}} - \Delta\hat{y} \,; y^{\text{т}} + \Delta\hat{y} ), \tag{1.36}$$

або

$$y^{\text{т}} - \Delta\hat{y} \le M_x(Y) \le y^{\text{т}} + \Delta\hat{y} \;, \tag{1.37}$$

де $y^{\text{т}}$ - знайдене за формулою (1.3) теоретичне значення регресії, а $\Delta\hat{y}$ обчислюється за формулою

$$\Delta\hat{y} = t_{1-\alpha,n-2}\sigma_{\tilde{y}} \;, \tag{1.38}$$

в якій стандартна вибіркова похибка

$$\sigma_{\tilde{y}} = \frac{\sigma}{\sqrt{n}} \sqrt{1+(x-\bar{x})^2 / var(x)} \tag{1.39}$$

(середнє квадратичне відхилення $\sigma$ знаходиться за формулою (1.31)).

Із формули (1.38) випливає, що величина (довжина) надійного інтервалу залежить від значення пояснюючої змінної $x$ : при $x = \bar{x}$ вона мінімальна, а при видаленні $x$ від $\bar{x}$ величина інтервалу збільшується (рис. 1.10). Якщо обсяг вибірки $n$ необмежено зростає $(n \to \infty)$, то $\lim_{n\to\infty} \Delta\hat{y} = 0$.

Прогноз значень залежної змінної $Y$ за рівнянням регресії виправданий, якщо значення $x$ пояснюючої змінної $X$ не виходить за діапазон значень, які отримані за вибіркою (причому тим більше точний, чим ближче $x$ до $\bar{x}$ ). При цьому *екстраполяція* (використання кривої поза межами обстеженого діапазону значень пояснюючої змінної) може призвести до значних похибок. З'єднуючи точки $(x_i, y_i^{\text{т}} - \Delta\hat{y}_i)$ $(i = \overline{1,n})$ відрізками прямої, дістанемо одну ламану лінію, а з'єднуючи відрізками прямої точки $(x_i, y_i^{\text{т}} + \Delta\hat{y}_i)$ $(i = \overline{1,n})$, одержимо іншу ламану лінію. Частина площі, що лежить між цими ламаними, називається *зоною надійності регресії.*

*Надійний інтервал для індивідуальних значень залежної змінної*

Вибіркова похибка парної лінійної кореляційно-регресійної моделі пов'язана з деяким значенням факторної ознаки $X$ і оцінює умовне середнє значення результуючої змінної $Y$ для цього значення $X$. Однак при заданому значенні $X$ результуюча змінна $Y$ набуває кількох різних *індивідуальних, конкретних* значень. Зазвичай у практиці використання економетричних моделей виникає потреба в оцінюванні такого конкретного значення результуючої змінної.

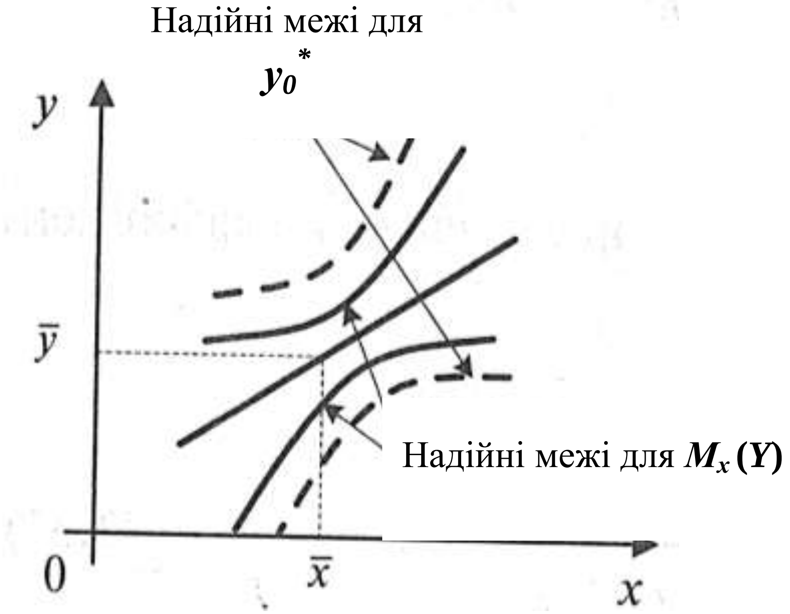



Рис. 1.10

Побудована надійна область для $M_x(Y)$ визначає місце розташування теоретичної лінії регресії (тобто умовного математичного сподівання), але не *окремих можливих значень* залежної змінної, *які відхиляються від середньої*. У зв'язку з цим при визначенні надійного інтервалу для індивідуальних значень $y_0^*$ залежної змінної, треба враховувати ще одне джерело варіації

- *розсіювання навколо лінії регресії*, тобто в оцінку сумарної дисперсії $\sigma^2_{y_т}$ варто включити величину $\sigma^2$, яке обчислюється за формулою (1.31). Оцінка дисперсії індивідуальних значень $y_0$ при $x = x_0$ дорівнює

$$\sigma^2_{\bar{y}_0} = \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) = \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n \cdot var(x)} \right), \qquad (1.40)$$

а відповідний надійний інтервал для прогнозів індивідуальних значень $y_0$* з імовірністю $\gamma = 1 - \alpha$ (або з рівнем значущості $\alpha$) визначається за формулою:

$$\bar{y}_0 - \Delta y_0 \le y_0^* \le \bar{y}_0 + \Delta y_0,$$

де $\Delta y_0 = t_{1-\alpha;n-2} \sigma_{\bar{y}_0}$ ;

стандартна похибка оцінювання індивідуального прогнозу

$$\sigma_{\bar{y}_0} = \sigma \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n \cdot var(x)}} \qquad (1.41)$$

(середнє квадратичне відхилення σ знаходиться за формулою (1.31)).

Із формули (1.41) випливає, що чим далі $x_0$ відстоїть від $\bar{x}$, тим інтервал прогнозу стає *ширшим*, тобто прогноз *погіршується*.

Довірчий інтервал для індивідуального прогнозу геометрично інтерпретується смугою між двома гіперболами.

## 1.7. Коефіцієнт еластичності

Коефіцієнт еластичності часто використовують в задачах для оцінки впливу на показник якого-небудь фактора.

Розглянемо статистичний ряд з базисними даними показника і фактора, що зображені у вигляді:

| $X$ | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| $Y$ | $y_1$ | $y_2$ | … | $y_n$ |

Коефіцієнт еластичності для значення фактора $x_i$ обчислюється за формулою:

$$E_{x_i} = \frac{\Delta y_i}{y_i} : \frac{\Delta x_i}{x_i}.$$

де $\Delta y_i = y_{i+1} - y_i$, $\quad \Delta x_i = x_{i+1} - x_i$.

Виконуючи в рівності операцію граничного переходу при $\Delta x_i \to 0$, одержимо формулу для точкової оцінки коефіцієнта еластичності:

$$E_x = \frac{x}{y} \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = y' \frac{x}{y}.$$

*Еластичністю функції* $y = f(x)$ називається границя відношення відносного приросту функції до відносного приросту аргументу $x$ при $\Delta x \to 0$ та позначається $E_x(y)$:

$$E_x(y) = \lim_{x \to 0} \frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \frac{x}{y} \lim_{x \to 0} \frac{\Delta y}{\Delta x} = y' \frac{x}{y}.$$

*Інтерпретація еластичності*

Еластичність функції показує наближено, на скільки відсотків зміниться значення функції $y = f(x)$ у разі зміни незалежної змінної $x$ на 1%.

$$\frac{\Delta y}{y} \approx E_x(y) \frac{\Delta x}{x}, \quad \Delta x \to 1.$$

Якщо $|E_x(y)| < 1$, то функція називається *нееластичною* (відносний приріст її спадає); якщо $|E_x(y)| > 1$, то функція називається *еластичною* (відносний приріст її зростає); якщо $|E_x(y)| = 1$, то функція називається *нейтральною.*

Для парної лінійної регресії $y^{\text{т}} = b_0 + b_1 x$ похідна

$$(y^{\text{т}}(x))' = (b_0 + b_1 x)' = b_1.$$

Коефіцієнт еластичності обчислюється за формулою:

$$E_x(y) = b_1 x / (b_0 + b_1 x).. \tag{1.42}$$

*Примітка*. Загальне поняття еластичності з'явилося в економіці з природничих наук. Вперше термін «еластичність» був застосований фізиком і хіміком Р. Бойлем (1626 – 1691) при вивченні властивостей газів. Економічне означення еластичності першим застосував у 1885 році А. Маршалл (1842 – 1924), який розвинув роботи відомих класиків А.Сміта (1723 - 1790) та Д.Рікардо (1772-1823). Подальший розвиток теорії еластичності був здійснений відомими економістами Дж. Хіксом (1904 - 1989) і П. Самуельсоном (1915 – 2009).

## 1.8. Середня похибка апроксимації

Фактичні значення результативної ознаки $y$ відрізняються від теоретичних $\hat{y}_x$, що були розраховані за рівнянням регресії. Чим менше ця різниця, тим ближче теоретичні значення до емпіричних даних, краще якість моделі. Величина відхилень фактичних і розрахункових значень результативної ознаки ($y$ - $\hat{y}_x$) за кожним спостереженням і є *похибкою апроксимації* (похибкою наближення). Їх число відповідає обсягу сукупності. В окремих випадках похибка апроксимації може дорівнювати нулю. Відхилення ($y$ - $\hat{y}_x$) не можна порівнювати між собою, виключаючи величину, яка дорівнює нулю. Так, якщо для одного спостереження відхилення $y - \hat{y}_x = 5$, а для другого воно дорівнює $y - \hat{y}_x = 10$, то це не свідчить про те, що у другому випадку модель дає вдвічі гірший результат.

Для порівняння використовуються величини відхилень, які виражені у відсотках до фактичних значень. Оскільки відхилення може бути як додатною величиною, так і від'ємною, то похибки апроксимації для спостережень визначають у відсотках за модулем.

Відхилення $| y - \hat{y}_x |$ є *абсолютною* похибкою апроксимації, а величина $|(y - \hat{y}_x)/y| \cdot 100~\%$ є *відносною* похибкою апроксимації.

Якість моделі із відносних відхилень за кожним спостереженням оцінюють середньою похибкою апроксимації *MAPE* (*mean absolute percentage error*; або середньою абсолютною процентною помилкою):

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

Якщо, наприклад, для першого спостереження $y$ = 20, а для другого $y$ = 50, то похибка апроксимації для першого спостереження дорівнює (5/20)·100% = 25% і (10/50) = 20% для другого спостереження. Порогові значення середньої похибки апроксимації наведені у таблиці:

| Оцінка *MAPE*, % | Характеристика якості регресійної моделі |
|---|---|
| < 10 | висока точність |
| 10 ÷ 20 | добра точність |
| 20 ÷ 50 | задовільна точність |
| > 50 | незадовільна точність |

## 1.9. Приклади парної лінійної економетричної моделі

**Приклад 1.1.** На основі даних про валовий регіональний продукт та величину основних засобів по регіонах України за 2004 рік (таблиця 1.1) необхідно:

1) побудувати аналітичне групування, графічно зобразити його та зробити висновок про залежність між валовим регіональним продуктом та величиною основних засобів;
2) побудувати парну лінійну кореляційно регресійну модель, яка описує залежність величини валового регіонального продукту від величини основних засобів;
3) дати економічну інтерпретацію параметрів побудованої моделі;
4) провести економетричний аналіз випадкових відхилень побудованої моделі;
5) побудувати спряжену парну лінійну кореляційно-регресійну модель;
6) оцінити силу кореляційного зв'язку між величинами основних

засобів по регіонах України;

7) побудувати довірчий інтервал оцінки фактичного значення валового регіонального продукту;
8) знайти коефіцієнт детермінації;
9) побудувати довірчий інтервал для істинних значень параметрів лінійної моделі;
10) обчислити стандартну вибіркову похибку моделі і побудувати довірчий інтервал для фактичного значення валового регіонального продукту, яке відповідає значенню основних засобів, що становлять 20 000 млн. грн.;
11) обчислити стандартну вибіркову похибку та побудувати довірчий інтервал для конкретного значення валового регіонального продукту у регіоні, в якому величина основних засобів дорівнює 20 000 млн. грн.;
12) обчислити коефіцієнт еластичності при $x$ = 20 000 млн.грн.

Таблиця 1.1. Вибіркова сукупність даних про величини валового регіонального продукту та основних засобів по регіонах України [23]

| № п/п | Область | Валовий регіональний продукт у фактичних цінах ($y$), млн. грн. | Основні засоби у фактичних цінах ($x$), млн. грн. |
|---|---|---|---|
| 1 | АР Крим | 9 901 | 43 758 |
| 2 | Вінницька | 8 123 | 25 993 |
| 3 | Волинська | 4 994 | 14 962 |
| 4 | Дніпропетровсь | 30 040 | 128 686 |
| 5 | Донецька | 45 617 | 129 043 |
| 6 | Житомирська | 5 947 | 20 207 |
| 7 | Закарпатська | 5 297 | 15 182 |
| 8 | Запорізька | 15 255 | 59 225 |
| 9 | Івано- | 7 311 | 26 811 |
| 10 | Київська | 11 883 | 42 289 |

| | | | |
|---|---|---|---|
| 11 | Кіровоградська | 5 594 | 19 729 |
| 12 | Луганська | 14 672 | 51 946 |
| 13 | Львівська | 13 992 | 51 890 |
| 14 | Миколаївська | 7 934 | 26 451 |
| 15 | Одеська | 17 029 | 57 173 |
| 16 | Полтавська | 13 983 | 52 493 |
| 17 | Рівненська | 5 599 | 18 964 |
| 18 | Сумська | 6 275 | 26 542 |
| 19 | Тернопільська | 3 948 | 13 913 |
| 20 | Харківська | 20 524 | 75 594 |
| 21 | Херсонська | 5 200 | 18 219 |
| 22 | Хмельницька | 6 344 | 22 399 |
| 23 | Черкаська | 6 623 | 23 238 |
| 24 | Чернівецька | 3 277 | 12 893 |
| 25 | Чернігівська | 6 181 | 27 430 |
| 26 | м. Київ | 61 357 | 129 325 |
| 27 | м. Севастополь | 2 213 | 6 714 |

***Розв'язання***

1) Валовий регіональний продукт є узагальнюючим показником соціального та економічного розвитку регіону і характеризує динаміку і обсяги структурних зрушень в економіці регіонів України. В основу визначення цього показника покладено виробничий метод. Валовий регіональний продукт складається із суми валових доданих вартостей по кожному виду економічної діяльності, скоригованої на величину непрямо вимірюваних послуг фінансованих посередників та податків за винятком субсидій на продукти.

Основні засоби є однією із складових національного багатства країни. Де них належать земельні ділянки, будинки, споруди та передавальні пристрої, машини та обладнання, транспортні засоби, інструменти, багаторічні насадження, робоча і продуктивна худоба тощо. До основних засобів належать основні засоби підприємств, організацій, установ усіх форм власності, а також основні засоби, що перебувають в особистій власності населення. При цьому вважається, що величина основних засобів є причиною (факторною ознакою $x$) отримання деякої величини валового регіонального продукту (наслідку, результуючої змінної $y$).

Згідно із формулою Стерджеса, аналітичне групування повинно містити 6 груп, оскільки вибіркова сукупність даних про валовий регіональний продукт та величину основних засобів по регіонах становить 27 спостережень (24 області, АР Крим та два міста — Київ і Севастополь).

Оскільки $x_{min}$ = 6714 та $x_{max}$ = 129325, то розмір інтервалу

$$h = (129325 – 6714)/2 = 20435{,}1667.$$

Тоді перший інтервал має межі [6714,0000; 6714,0000+20435,1667) або [6714,0000; 27 149,1667),

другий інтервал має межі [27149,1667; 27149,1667+20435,1667), або [27149,1667; 47584,3333) тощо.

При цьому [6714,0000; 27149,1667), [27149,1667; 47584,3333), [47584,3333; 68019,5000), [68019,5000; 88454,6667) – напівзакриті інтервали,

останній інтервал [108889,8333; 129325,0000] - закритий. Аналітичне групування доцільно зобразити у вигляді таблиці 1.2.

Геометрично результати аналітичного групування можна представити як емпіричну лінію регресії (рис.1.11). Щоб її побудувати, треба у декартовій системі координат на площині зобразити ламану лінію, яка сполучає точки $(\bar{x}_i, \bar{y}_i)$, $i = \overline{1,6}$, де $\bar{x}_i = (x_i^l + x_i^p)/2$ ($x_i^l$ та $x_i^p$ - відповідно ліва та права границя $i$-го інтервалу) – середини інтервалів значень основних засобів (факторної ознаки), а $\bar{y}_i$ — середні групові значення валового регіонального продукту (результуючої змінної).

Таблиця 1.2. Аналітичне групування валового регіонального продукту за величиною основних засобів по регіонах України

| Номер з/п | Основні засоби у фактичних цінах, млн. грн | Кількість регіонів | Середнє значення валового регіонального |
|---|---|---|---|
| 1 | 6 714,0000 - 27 149,1667 | 15 | 5 645,2667 |
| 2 | 27 149,1667 - 47 | 3 | 9 321,6667 |
| 3 | 47 584,3333 - 68 | 5 | 14 986,2000 |
| 4 | 68 019,5000 - 88 | 1 | 20 524,0000 |
| 5 | 88 454,6667 - 108 | 0 | — |
| 6 | 108 889,8333 - 129 | 3 | 45 671,3333 |
| **Разом** | | **27** | |
| **У середньому** | | | **12 781,9630** |

Для прикладу

$\bar{x}_1$ = 16 931,5833; $\bar{x}_2$ = 37 366,7500; $\bar{x}_3$=*57* 801,9167;

$\bar{x}_4$ = *78* 237,0833; $\bar{x}_5$ = 98 672,2500; $\bar{x}_6$ = 119 107,4167.

З аналітичного групування можна зробити висновок про те, що між величиною валового регіонального продукту та величиною основних засобів по регіонах України наявна кореляційна залежність, оскільки під час збільшення величини основних засобів зростає середнє значення величини валового регіонального продукту.

*Зауваження.* Середнє значення величини валового регіонального продукту у п’ятій групі аналітичного групування є невизначеним, бо до неї не потрапило жодної одиниці сукупності. У цьому разі точку $(\bar{x}_5, \bar{y}_5)$ в системі координат не зображають, а сполучають точки $(\bar{x}_4, \bar{y}_4)$ та $(\bar{x}_6, \bar{y}_6)$, щоби отримати ламану лінію емпіричної регресії.

2) Використовуючи засоби Excel, зокрема функції ОТРЕЗОК та ЛИНЕЙН (*Мастер функций* $f_x$ → категорія *Статистические*) або *Анализ данняx* → *Регрессия* (рис.1.12) отримаємо такі оцінки $b_0$ та $b_1$ параметрів $\beta_0$ та $\beta_1$:

$$b_0 = -2122{,}6756; \quad b_1=0{,}3527.$$

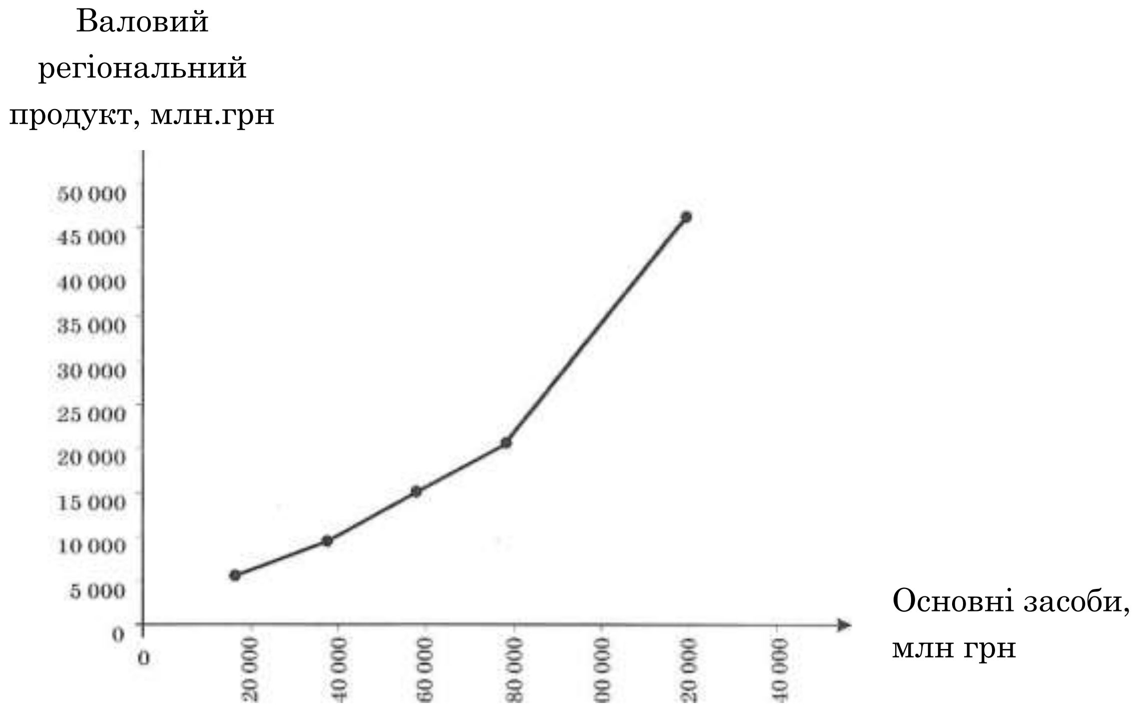



Рис. 1.11. Емпірична лінія регресії залежності валового регіонального продукту від величини основних засобів по регіонах України

Парна лінійна кореляційно-регресійна модель, яка описує залежність величини валового регіонального продукту від величини основних засобів по регіонах України, має вигляд:

$$\hat{y} = -2122{,}6756 + 0{,}3527\,x.$$

Областю існування кореляційно-регресійної моделі є інтервал [6714, 129325]. На рисунку 1.12 зображено регресійне поле та пряма регресії, яка відображає залежність валового регіонального продукту від величини основних засобів.

3) У побудованій парній кореляційно-регресійній моделі коефіцієнт регресії $b_1=0{,}3527$. Оскільки значення коефіцієнта регресії відмінне від нуля, то це вказує на наявність кореляційного зв'язку між величиною валового регіонального продукту та величиною основних засобів. Додатне значення $b_1$ вказує на те, що при збільшенні величини основних засобів у регіонах середнє значення валового регіонального продукту по регіонах України має зростати.

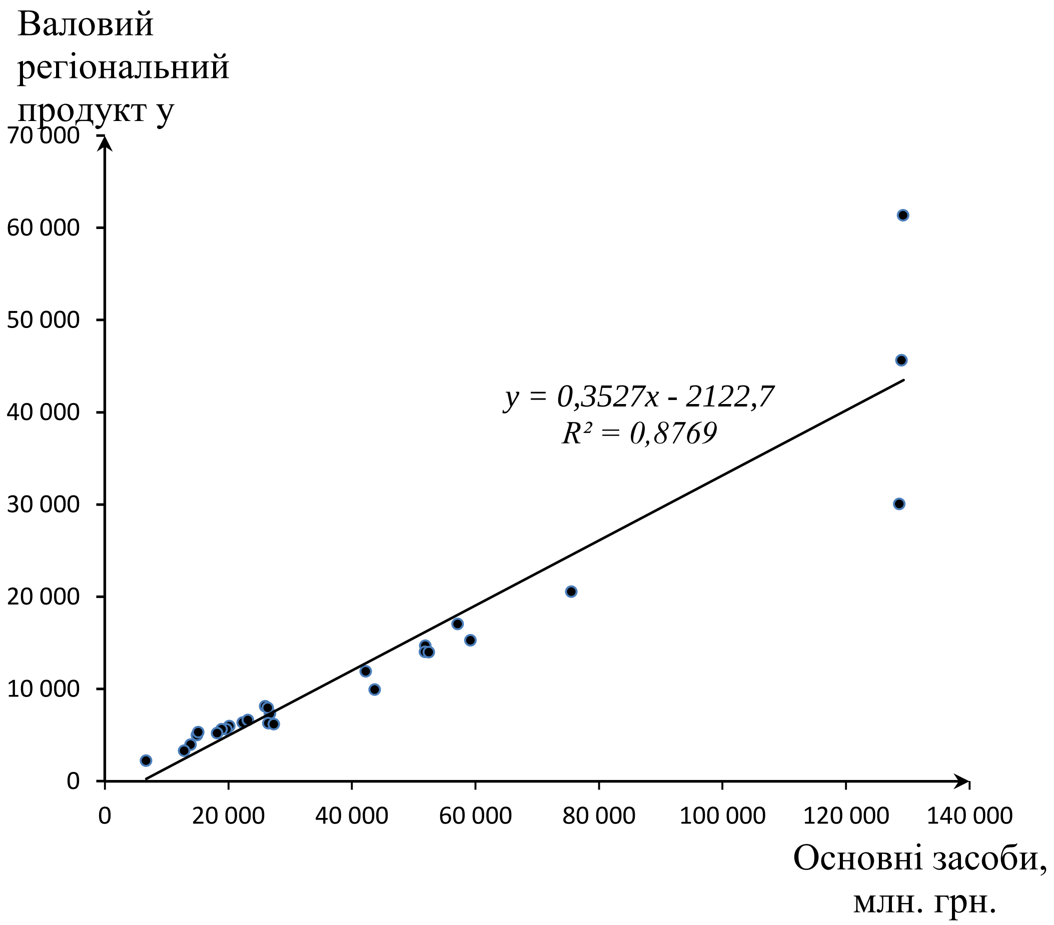



Рис. 1.12

Якщо $b_1$*=0,3527*, то при *збільшенні величини основних засобів на один млн. грн. значення валового регіонального продукту у середньому по регіонах України має зростати на 0,3527 млн. грн*.

Вільний член кореляційно-регресійній моделі $b_0$ = -2122,6756. Це означає, що *при нульовому значенні величини основних засобів валовий регіональний продукт у середньому приймає значення - 2122,6756 млн грн* (цей факт можна економічно пояснити). Однак таке інтерпретування вільного члена є достатньо умовним, оскільки значення факторної ознаки $x$ = 0 не входить в область існування моделі ($x \in$ [6714, 129325]).

4) На підставі побудованої парної лінійної кореляційно-регресійної моделі можна розрахувати значення випадкових відхилень $e_i = y_i - \hat{y}_i$ теоретичних значень результуючої змінної від фактичних. Випадкові відхилення $e_i$ наведемо у таблиці 1.3.

Таблиця 1.3

Випадкові відхилення валового регіонального продукту в 2004 р.

| № з/п | Область | Валовий регіональний продукт (фактичні значення), млн. грн | Валовий регіональний продукт (нормативні значення), млн. грн | Випадкові відхилення, млн. грн |
|---|---|---|---|---|
| 1 | АР Крим | 9 901 | 13 309,6284 | -3 408,6284 |
| 2 | Вінницька | 8 123 | 7 044,3768 | 1 078,6232 |
| 3 | Волинська | 4 994 | 3 154,0311 | 1 839,9689 |
| 4 | Дніпропетровськ | 30 040 | 43 261,5163 | -13 221,5163 |
| 5 | Донецька | 45 617 | 43 387,4209 | 2 229,5791 |
| 6 | Житомирська | 5 947 | 5 003,8057 | 943,1943 |
| 7 | Закарпатська | 5 297 | 3 231,6194 | 2 065,3806 |
| 8 | Запорізька | 15 255 | 18 764,4354 | -3 509,4354 |
| 9 | Івано- | 7 311 | 7 332,8640 | -21,8640 |
| 10 | Київська | 11 883 | 12 791,5505 | -908,5505 |
| 11 | Кіровоградська | 5 594 | 4 835,2276 | 758,7724 |
| 12 | Луганська | 14 672 | 16 197,3222 | -1 525,3222 |
| 13 | Львівська | 13 992 | 16 177,5725 | -2 185,5725 |
| 14 | Миколаївська | 7 934 | 7 205,9014 | 728,0986 |
| 15 | Одеська | 17 029 | 18 040,7486 | -1 011,7486 |
| 16 | Полтавська | 13 983 | 16 390,2348 | -2 407,2348 |
| 17 | Рівненська | 5 599 | 4 565,4320 | 1 033,5680 |
| 18 | Сумська | 6 275 | 7 237,9948 | -962,9948 |
| 19 | Тернопільська | 3 948 | 2 784,0762 | 1 163,9238 |
| 20 | Харківська | 20 524 | 24 537,3543 | -4 013,3543 |
| 21 | Херсонська | 5 200 | 4 302,6900 | 897,3100 |
| 22 | Хмельницька | 6 344 | 5776,8668 | 567,1332 |
| 23 | Черкаська | 6 623 | 6 072,7602 | 550,2398 |
| 24 | Чернівецька | 3 277 | 2 424,3489 | 852,6511 |
| 25 | Чернігівська | 6 181 | 7 551,1692 | -1 370,1692 |
| 26 | м. Київ | 61 357 | 43 486,8749 | 17 870,1251 |

| | | | | |
|---|---|---|---|---|
| 27 | м. Севастополь | 2 213 | 245,1769 | 1 967,8231 |

Якщо випадкове відхилення набуває від'ємне значення, то теоретичне значення результуючої змінної перевищує відповідне фактичне, а отже, фактичне значення валового регіонального продукту у цьому разі є меншим, ніж відповідне нормативне. Отож, щоб збільшити валовий регіональний продукт, у регіоні маємо *резерви*, які дають *можливість ефективніше використовувати основні засоби*.

На основі даних таблиці 1.3 можна стверджувати, що необхідно підвищувати ефективність використання основних засобів

в АР Крим ($e_1$ = -3408,6284 млн грн),

Дніпропетровській ($e_4$ = -13 221,5163),

Запорізькій ($e_8$ = -3509,4354),

Луганській ($e_{12}$ = -1525,3222),

Львівській ($e_{13}$ = -2185,5725),

Одеській ($e_{15}$ = -1011,7486),

Полтавській ($e_{16}$ = -2407,2348),

Харківській ($e_{20}$ = -4013,3543) та інших областях.

Якщо *випадкове відхилення набуває додатне значення, то фактичне значення валового регіонального продукту перевищує відповідне нормативне, а отже, можна стверджувати, що у цьому регіоні основні засоби використовують ефективніше, ніж у середньому по Україні.*

Найефективніше основні засоби використовуються у Києві ($e_{26}$=17870,1251), Донецькій ($e_5$= 2229,5791), Закарпатській областях ($e_7$ = 2065,3806) та у Севастополі ($e_{27}$ = 1967,8231).

5) При побудові спряженої парної лінійної кореляційно-регресійній моделі припускаємо, що факторною ознакою $x$ є величина валового регіонального продукту, а результуючою змінною $y$ – величина основних засобів.

Використовуючи засоби Excel, зокрема функції ОТРЕЗОК та ЛИНЕЙН або *Анализ данных → Регрессия* (рис.1.13) отримаємо такі оцінки $\tilde{b}_0$ та $\tilde{b}_1$ параметрів $\tilde{\beta}_0$ та $\tilde{\beta}_1$ : $\tilde{b}_0$ = *10479,4463;* $\tilde{b}_1$ *=2,4865.*

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ВЫВОД ИТОГОВ | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | *ессионная статистика* | | | | | | | | |
| 4 | Множеств | 0,936442 | | | | | | | |
| 5 | R-квадрат | 0,876924 | | | | | | | |
| 6 | Нормиров | 0,872001 | | | | | | | |
| 7 | Стандартн | 12745,02 | | | | | | | |
| 8 | Наблюдени | 27 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | Дисперсионный анализ | | | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *ачимость F* | | | |
| 12 | Регрессия | 1 | 2,89E+10 | 2,89E+10 | 178,1267 | 7,11E-13 | | | |
| 13 | Остаток | 25 | 4,06E+09 | 1,62E+08 | | | | | |
| 14 | Итого | 26 | 3,3E+10 | | | | | | |
| 15 | | | | | | | | | |
| 16 | | *Коэффициен* | *артная о* | *татисти* | *-Значени* | *ижние 95* | *рхние 95* | *жние 95,0* | *рхние 95,0%* |
| 17 | Y-пересеч | 10479,45 | 3418,613 | 3,065409 | 0,005157 | 3438,681 | 17520,21 | 3438,681 | 17520,21 |
| 18 | Перемен | 2,486501 | 0,186305 | 13,34641 | 7,11E-13 | 2,102799 | 2,870203 | 2,102799 | 2,870203 |

Рис. 1.13

Спряжена парна лінійна кореляційно-регресійна модель (регресія $X$ на $Y$) має вигляд:

$$\tilde{x} = 10479{,}4463 + 2{,}486\, y.$$

Коефіцієнт регресії спряженої моделі $\tilde{b}_1$ *=2,4865* показує, що при збільшенні величини валового регіонального продукту на 1 млн. грн. середнє значення величини основних засобів повинно в середньому зрости на 2,4865 млн. грн.

На рис. 1.14 зображено регресійне поле та пряма регресії, яка відображає залежність основних засобів від валового регіонального продукту.

6) На рис. 1.15 вказано значення коефіцієнта кореляції $r$ (верхня частина таблиці *Регрессионная статистика* рядок

*Множественный*). Величина коефіцієнта кореляції $r$ = 0,9364 свідчить про *дуже тісний* кореляційний зв'язок між валовим регіональним продуктом та величиною основних засобів.

| | A | B |
|---|---|---|
| 1 | Валовий регіональний продукт у фактичних цінах, млн. грн. | Основні засоби у фактичних цінах, млн. грн. |
| 2 | 9 901 | 43 758 |
| 3 | 8 123 | 25 993 |
| 4 | 4 994 | 14 962 |
| 5 | 30 040 | 128 686 |
| 6 | 45 617 | 129 043 |
| 7 | 5 947 | 20 207 |
| 8 | 5 297 | 15 182 |
| 9 | 15 255 | 59 225 |
| 10 | 7 311 | 26 811 |
| 11 | 11 883 | 42 289 |
| 12 | 5 594 | 19 729 |
| 13 | 14 672 | 51 946 |
| 14 | 13 992 | 51 890 |
| 15 | 7 934 | 26 451 |
| 16 | 17 029 | 57 173 |
| 17 | 13 983 | 52 493 |
| 18 | 5 599 | 18 964 |

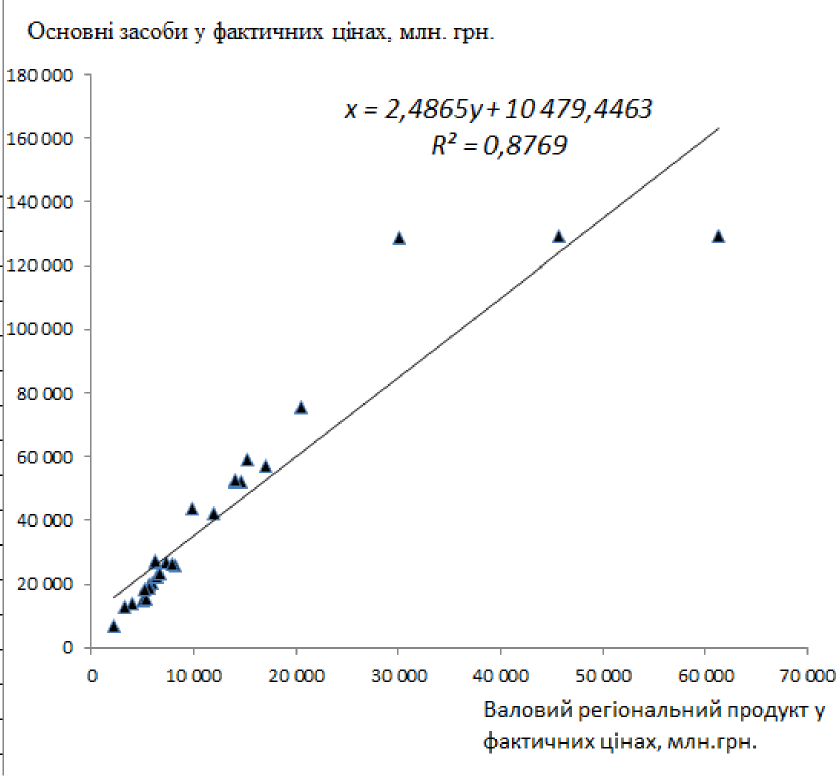



Рис.1.14

ВЫВОД ИТОГОВ

| Регрессионная статистика | |
|---|---|
| Множественный | 0,936442261 |
| R-квадрат | 0,876924107 |
| Нормированный | 0,872001072 |
| Стандартная оши | 4799,905328 |
| Наблюдения | 27 |

Дисперсионный анализ

| | df | SS | MS | F | Значимость F |
|---|---|---|---|---|---|
| Регрессия | 1 | 4103877294 | 4103877294 | 178,1267007 | 7,11342E-13 |
| Остаток | 25 | 575977278,8 | 23039091,15 | | |
| Итого | 26 | 4679854573 | | | |

| | Коэффициенты | Стандартная ошибка | t-статистика | P-Значение | Нижние 95% | Верхние 95% | Нижние 95,0% | Верхние 95,0% |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | -2122,675559 | 1449,288043 | -1,46463332 | 0,155484178 | -5107,540158 | 862,1890403 | -5107,540158 | 862,1890403 |
| Переменная X 1 | 0,352673887 | 0,026424623 | 13,34641153 | 7,11342E-13 | 0,298251359 | 0,407096416 | 0,298251359 | 0,407096416 |

Рис. 1.15

Значення коефіцієнта кореляції $r$ = 0,9364 можна було б отримати за формулою (1.21), тобто

$$r = \sqrt{b_1 \cdot \tilde{b}_1} = \sqrt{\rho_{y/x} \cdot \rho_{x/y}} \approx \sqrt{0{,}3527 \cdot 2{,}4865} \approx 0{,}9364.$$

У середовищі MS Excel для обчислення коефіцієнта кореляції $r$ можна також використати статистичну функцію КОРРЕЛ(X; Y), де X та Y – відповідні діапазони вихідних даних.

7) Обчислимо стандартну похибку парної лінійної кореляційно-регресійної моделі. Оскільки обсяг вибірки малий ($n$ < 30), то під час обчислення стандартної похибки використовуємо формулу (1.19). Отримаємо

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} = \sqrt{\frac{575977278{,}8}{27-2}} \approx 4799{,}9053.$$

Число «575977278,8» взято з рис. 1.15 (розділ «Дисперсионный анализ», перетин рядка «Остаток» і стовпця «SS»). Відповідь «4799,9053» збігається з табличним значенням, розташованим у графі «Регрессионная статистика» рядок «Стандартная ошибка».

Задамо довірчу ймовірність $p$ = 0,95 (рівень значущості α=0,05). Оскільки обсяг вибірки малий, то для того, щоб знайти імовірнісний коефіцієнт $t_p = t_{1-\alpha,n-2}$, використаємо таблиці розподілу Ст'юдента (таблиця 2 Додатку) або функцію Excel СТЬЮДЕНТ.ОБР(0,95;25). За цими таблицями при рівні значущості α = 0,05 та кількості ступенів вільності 25 значення $t_p$ = 1,708. Гранична похибка моделі

$$t_p \sigma = 1{,}708 \cdot 4799{,}9053 = 8198{,}914 \text{ млн. грн.}$$

Довірчий інтервал, наприклад, для $y_1$, має вигляд:

$$13309{,}6284 - 8198{,}914 \le y_1 \le 13309{,}6284 + 8198{,}914$$

$$5110{,}715 \le y_1 \le 21508{,}54.$$

Фактичне значення валового регіонального продукту АР Крим ($y_1$) із ймовірністю $p$ = 0,95 коливається в межах від 5110,715 до 21508,54 (млн. грн.).

8) На рис. 1.15 у розділі «*Регрессионная статистика*» у рядку «*R квадрат*» знаходимо значення коефіцієнту детермінації: $R^2$=0,8769

В середньому по Україні 87,69% зміни величини валового регіонального продукту пояснюються зміною величини основних засобів у цьому регіоні.

Індекс кореляції $i_R$ = 0,9364 > 0,9 свідчить про те, що залежність між величинами валового регіонального продукту та основних засобів є дуже тісною.

9) Дві пари полів (рис.1.15) «Нижние 95%» та «Верхние 95%» - 95%-ві довірчі інтервали для коефіцієнтів $\beta_0$ (-5107,5402; 862,1890) та $\beta_1$(0,2982; 0,4071). У першій парі полів наведені 95%-ні довірчі інтервали, у другій – довірчі інтервали, що вказуються у вікні «Регрессия» (там було встановлено 95%, тому вони й співпали).

Отже, довірчий інтервал для істинного значення параметра $\beta_0$ має вигляд:

$$-5107{,}5402 \le \beta_0 \le 862{,}1890$$

(істинне значення вільного члена $\beta_0$ генеральної сукупності з ймовірністю $p = 0{,}95$ лежить у межах від -5107,5402 до 862,1890 (млн. грн.)).

Довірчий інтервал для істинного значення коефіцієнта регресії $\beta_1$ має вигляд:

$$0{,}2982 \le \beta_1 \le 0{,}4071.$$

(істинне значення коефіцієнта регресії $\beta_1$ генеральної сукупності з ймовірністю $p = 0{,}95$ лежить у межах від 0,2982 до 0,4071(млн. грн.)).

10) Для того, щоб побудувати довірчий інтервал для фактичного значення валового регіонального продукту, яке відповідає значенню основних засобів, що становлять 20 000 млн. грн., знайдемо стандартну вибіркову похибку парної лінійної кореляційно-регресійної моделі за формулою (1.39). Отримаємо:

$$\sigma_{\tilde{y}} = \frac{4799{,}9053}{\sqrt{27}}\sqrt{1 + (20000 - 42261{,}8)^2/1222035980} \approx 1095{,}15.$$

Тут значення $var(x)$=1222035980 знаходиться в Excel за допомогою функції ДИСП.Г(X), де X – діапазон табличних значень вихідних даних.

Теоретичне значення результуючої змінної, яку відповідає значенню факторної ознаки $x$ = 20 000 млн. грн., дорівнює

$$\tilde{y}_{x=20000} = -2122,6756 + 0,3527 \cdot 20\,000 = 4930,8022 \text{ млн. грн.}$$

Гранична вибіркова похибка парної кореляційно-регресійної моделі та довірчий інтервал для фактичного значення результуючої змінної, яке відповідає значенню факторної ознаки $x$ = 20 000 млн. грн., при заданій довірчій імовірності $p$=0,95 має вигляд:

$$\Delta\tilde{y} = t_p \cdot \sigma_{\tilde{y}} = 1,708 \cdot 1095,1482 = 1870,513 \text{ млн. грн.};$$

$$4930,8022 - 1870,513 \le y \le 4930,8022 + 1870,513;$$

$$3060,289 \le y \le 6801,315.$$

Із ймовірністю 0,95 фактичне значення валового регіонального продукту у *всіх* регіонах, де основні засоби оцінюють у 20000 млн. грн., знаходиться в межах від 3060,289 до 6801,315 (млн. грн.).

11) Для того, щоб побудувати довірчий інтервал для конкретного значення валового регіонального продукту у регіоні, в якому величина основних засобів дорівнює 20 000 млн. грн., знайдемо стандартну похибку оцінювання індивідуального прогнозу для $x$ = 20 000 млн. грн. за формулою (1.41)

$$\sigma_{\tilde{y}} = 4799,9053\sqrt{1 + \frac{1}{27} + \frac{1}{27} \cdot \frac{(20000 - 42261,8)^2}{1222035980}} \approx 4923,2551.$$

Теоретичне значення результуючої змінної, яке відповідає значенню факторної ознаки $x$ = 20 000 млн. грн., дорівнює

$$\tilde{y}_{x=20000} = -2122,6756 + 0,3527 \cdot 20\,000 = 4930,8022 \text{ млн. грн.}$$

Граничну вибіркову похибку оцінювання *індивідуального прогнозу* величини валового регіонального продукту у *конкретному регіоні*, в якому основні засоби оцінюють у 20 000 млн. грн., та довірчий інтервал для цього прогнозу при заданій довірчій ймовірності $p = 0{,}95$ мають такий вигляд:

$$\Delta_{\tilde{y}} = t_p \cdot \sigma_{\tilde{y}} = 1{,}708 \cdot 4923{,}2551 = 8408{,}9197 \text{ млн. грн.};$$

$$4930{,}8022 - 8408{,}9197 \leq y \leq 4930{,}8022 + 8408{,}9197;$$
$$-3478{,}12 \leq y \leq 13339{,}72.$$

Із ймовірністю 0,95 фактичне значення валового регіонального продукту в конкретному регіоні з величиною основних засобів, що дорівнює 20 000 млн. грн., знаходиться в межах від -3478,12 до 13339,72 (млн. грн.).

12) Обчислимо коефіцієнт еластичності при $x$ = 20 000 за формулою (1.42):

$$E_{20000} = \frac{0{,}3527 \cdot 20000}{0{,}3527 \cdot 20000 - 2122{,}6756} \approx 1{,}43.$$

*Висновок*. У разі збільшення значення основних засобів на 1% валовий регіональний продукт збільшиться наближено на 1,43%. Оскільки $E_{20000} = 1{,}43 > 1$, то валовий продукт є еластичним (відносний приріст її зростає).

**Приклад 1.2.** Перевірити існування лінійного зв'язку між попитом на товар у магазинах (грош.од.) та його ціною (грош.од.) (вихідні дані наведено у табл. 1.4).

1) Побудувати лінійну економетричну модель.
2) Перевірити статистичну значущість параметрів моделі. Перевірити адекватність моделі за коефіцієнтами кореляції та детермінації і критерієм Фішера. Перевірити статистичну значущість коефіцієнта кореляції за допомогою критерію Стьюдента.

3) Розрахувати прогнозні значення залежної змінної та довірчі інтервали, якщо відомі дані незалежного показника.

4) Обчислити коефіцієнт еластичності при $x = 12$ грош. од.

Таблиця 1.4

| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ціна на товар, $x$ | 5,4 | 7,6 | 2,3 | 5,9 | 11 | 12,6 | 10,4 | 4,9 | 2,4 | 1,6 |
| Попит на товар, $y$ | 13,7 | 18 | 6,2 | 15,5 | 24,1 | 24,8 | 25 | 13 | 8,1 | 6,7 |

***Розв'язання***

1) Припустимо, що між попитом на товар $y$ та його ціною $x$ існує лінійний зв'язок. Лінійна однофакторна економетрична модель має вигляд:

$$y_i = b_0 + b_1 x_i \qquad (\boldsymbol{i = \overline{1,10}}),$$

де $y_i$ - значення залежного фактору для $i$-го спостереження

$b_0$ - вільний параметр; $b_1$ - параметр при незалежному факторі;

$x_i$ - значення незалежного фактору для $i$-го спостереження.

Задача полягає у визначенні емпіричних коефіцієнтів регресії, тобто оцінок параметрів економетричної моделі $b_0$ та $b_1$ на основі МНК:

$$b_1 = cov(x,y)/var(x); \quad b_0 = \bar{y} - b_1 \cdot \bar{x}.$$

Допоміжні обчислення наведені у таблиці 1.5.

Отримали значення емпіричних коефіцієнтів регресії, тобто параметрів лінійної залежності:

$$b_1 = 25{,}3739 / 13{,}6389 \approx 1{,}86;$$
$$b_0 = 15{,}51 - 1{,}86 \cdot 6{,}41 \approx 3{,}58.$$

Значення $cov(X,Y) = 25{,}3739$ та $var(x) = 13{,}6389$ можна знайти за допомогою статистичних функцій Excel КОВАР(X;Y) і ДИСП.Г(X) відповідно, де X та Y– діапазон табличних значень вихідних даних.

Таблиця 1.5

| № | $x$ | $y$ | $x^2$ | $(x-\bar{x})^2$ | $(y-\bar{y})^2$ | $(x-\bar{x})\cdot(y-\bar{y})$ | $\hat{y}=y^T$ | $(y^T-y)^2$ |
|---|---|---|---|---|---|---|---|---|
| *1* | *5,4* | *13,7* | *29,16* | *1,0201* | *3,2761* | *1,8281* | *13,631* | *0,0048* |
| *2* | *7,6* | *18,0* | *57,76* | *1,4161* | *6,2001* | *2,9631* | *17,723* | *0,0762* |
| *3* | *2,3* | *6,2* | *5,29* | *16,8921* | *86,676* | *38,2641* | *7,8637* | *2,7680* |
| *4* | *5,9* | *15,5* | *34,81* | *0,2601* | *0,0001* | *0,0051* | *14,561* | *0,8814* |
| *5* | *11,0* | *24,1* | *121,00* | *21,0681* | *73,788* | *39,4281* | *24,049* | *0,0026* |
| *6* | *12,6* | *24,8* | *158,76* | *38,3161* | *86,304* | *57,5051* | *27,025* | *4,9547* |
| *7* | *10,4* | *25,0* | *108,16* | *15,9201* | *90,060* | *37,8651* | *22,933* | *4,2724* |
| *8* | *4,9* | *13,0* | *24,01* | *2,2801* | *6,3001* | *3,7901* | *12,700* | *0,0895* |
| *9* | *2,4* | *8,1* | *5,76* | *16,0801* | *54,908* | *29,7141* | *8,0498* | *0,0025* |
| *10* | *1,6* | *6,7* | *2,56* | *23,1361* | *77,616* | *42,3761* | *6,5614* | *0,0192* |
| **Сума** | ***64,1*** | ***155,1*** | ***547,27*** | ***136,389*** | ***485,12*** | ***253,739*** | ***155,1*** | ***13,0713*** |
| **Сер.зн** | ***6,41*** | ***15,51*** | ***54,727*** | ***13,6389*** | ***48,512*** | ***25,3739*** | ***15,51*** | ***1,3071*** |

Рівняння залежності попиту на товар від його ціни має вигляд:

$$\hat{y} = 3{,}58 + 1{,}86\ x.$$

Пряма регресії має вигляд («Диаграмма» → «Добавить линию тренда», обрати тип «Линейная», встановити прапорці на «Показывать уравнение на диаграмме» та «Поместить на диаграмме величину достоверности аппроксимации (R^2)):

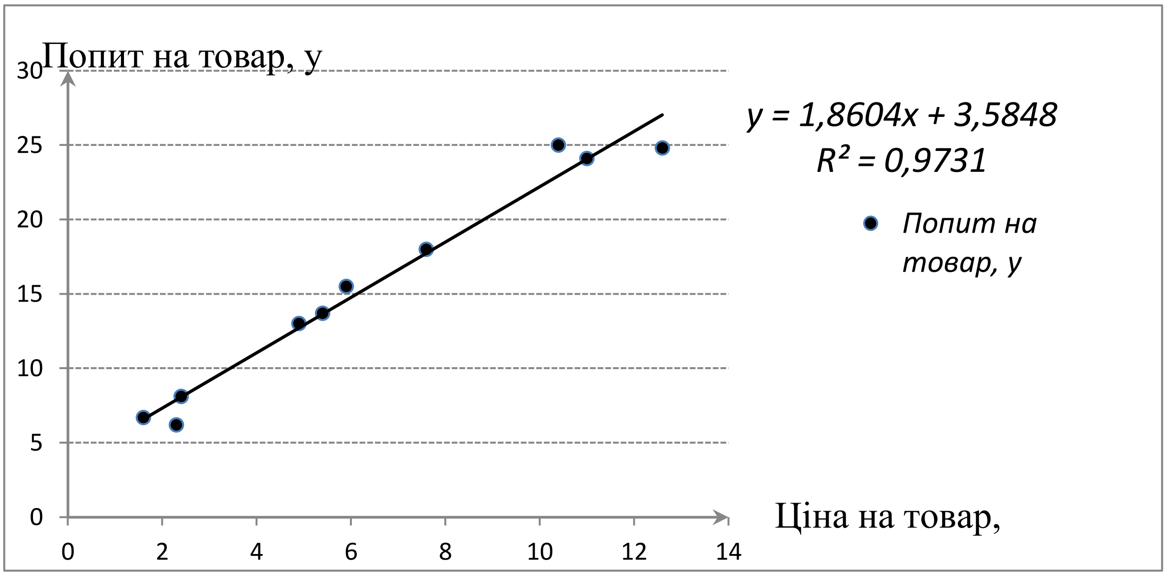



Рис. 1.16

2) Для перевірки гіпотези про те, чи значимо відрізняється від нуля вибірковий коефіцієнт $b_1$, перевіряємо статистичну значущість цього коефіцієнта, використовуючи критерій Стьюдента, тобто перевіряємо гіпотезу

$$H_0\colon b_j = 0 \text{ для } j = 0;\, 1$$

проти відповідної альтернативної гіпотези

$$H_1\colon b_j \neq 0 \text{ для } j = 0;\, 1.$$

Знаходимо розрахункові значення параметрів моделі за критерієм Стьюдента за формулою:

$$t_{b_j} = b_j / \sigma_{b_j},$$

де $b_j$ – значення оцінки $j$-го параметра моделі ($j = 0; 1$);

$\sigma_{b_j}$ – середні квадратичні відхилення $j$-го параметра моделі ($j = 0; 1$), що розраховуються за такими формулами:

$$\sigma_{b_0} = \sqrt{\frac{\sigma^2 \cdot \sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2}}; \qquad \sigma_{b_1} = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}},$$

дисперсія похибок моделі $\sigma^2$ розраховується за формулою (1.19):

$$\sigma^2 = \sum\nolimits_{i=1}^{n} (y_i - \hat{y}_i)^2 / (n-2),$$

$\hat{y}_i$ – теоретичні значення залежного показника; $n$ - кількість спостережень.

Дисперсія похибок моделі:

$$\sigma^2 = 13{,}0713 / (10 - 2) \approx 1{,}634.$$

Середнє квадратичне відхилення похибок моделі $\sigma = 1{,}278$.

Середні квадратичні відхилення параметрів моделі:

$$\sigma_{b_0} = \sqrt{\frac{1{,}634 \cdot 547{,}27}{10 \cdot 136{,}389}} \approx 0{,}81; \quad \sigma_{b_1} = \sqrt{\frac{1{,}634}{136{,}389}} \approx 0{,}109.$$

Оцінка значущості коефіцієнтів регресії за допомогою критерія Стьюдента здійснюється шляхом співставлення їх значень з величиною випадкової помилки:

$$t_{b_0} = b_0/\sigma_{b_0} \approx 3{,}59/0{,}81 \approx 4{,}43;$$
$$t_{b_1} = b_1/\sigma_{b_1} \approx 1{,}86/0{,}109 \approx 17{,}00.$$

За таблицями розподілу Стьюдента знаходимо значення $t_p$ для числа ступенів вільності $k = n - 2 = 10 - 2 = 8$ та рівня значущості $\alpha = 0{,}05$, тобто $t_p\,(0{,}05;8) = 2{,}31$.

Порівнюючи отримані значення $|t_{b_0}| \approx 4{,}43$ та $|t_{b_1}| \approx 17$ з $t_p = 2{,}31$, робимо висновок, що отримані значення коефіцієнтів $b_0$ і $b_1$ є *статистично значущими*, тобто приймається гіпотеза $H_1$: $b_j \neq 0$ для $j = 0; 1$.

Знайдемо інтервальні оцінки для параметрів $\beta_0$ та $\beta_1$ за такими формулами:

$$b_0 - \Delta b_0 \le \beta_0 \le b_0 + \Delta b_0, \quad \Delta b_0 = t_p \cdot \sigma_{b_0};$$
$$b_1 - \Delta b_1 \le \beta_1 \le b_1 + \Delta b_1, \quad \Delta b_1 = t_p \cdot \sigma_{b_1}.$$

Оскільки $\Delta b_0 = 2{,}31 \cdot 0{,}81 = 1{,}87$ і $\Delta b_1 = 2{,}31 \cdot 0{,}109 = 0{,}25$, то для рівня значущості $\alpha = 0{,}05$ з імовірністю 0,95 істинні значення параметрів лежать у таких межах:

$$3{,}59 - 1{,}87 \le \beta_0 \le 3{,}59 + 1{,}87;$$
$$1{,}86 - 0{,}25 \le \beta_1 \le 1{,}86 + 0{,}25,$$

тобто $\quad 1{,}72 \le \beta_0 \le 5{,}46, \qquad 1{,}61 \le \beta_1 \le 2{,}11.$

Для проведення аналізу адекватності побудованої моделі розрахуємо коефіцієнт детермінації і лінійний коефіцієнт парної кореляції за формулами:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^T - y_i\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2};$$

$$r = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Оскільки коефіцієнт кореляції є також вибірковою характеристикою, яка може відхилятися від свого «істинного значення, значущість цього коефіцієнта також перевіряється за допомогою $t$- критерію за формулою (1.30):

$$t_r = r\sqrt{n-2}/\sqrt{1-r^2}.$$

Якщо $|t_r| \geq t_p$, то оцінка $r$ є статистично значущою.

Таким чином, отримуємо такі значення коефіцієнтів:

$$R^2 = 1 - 13{,}0713/\ 485{,}129 \approx 0{,}973\ (97{,}3\%);$$

$$r = 253{,}739\ /\ (136{,}389 \cdot 485{,}129)^{0,5} \approx 0{,}986.$$

У середовищі MS Excel для обчислення коефіцієнта кореляції $r$ можна також використати статистичну функцію КОРРЕЛ(X; Y), де X та Y – відповідні діапазони вихідних даних.

*Висновки*. Коефіцієнт детермінації показує, що *97,3% загальної зміни попиту пояснюється зміною ціни, у той час, як на інші фактори доводиться лише 2,7% зміни*. Коефіцієнт детермінації є мірою тісноти зв'язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є *дуже тісним.* Отримана економетрична модель є *адекватною*.

Перевіримо значущість коефіцієнта кореляції:

$$t_r = 0{,}986 \cdot \sqrt{10-2}/\sqrt{1-0{,}986^2} = 2{,}789/0{,}167 = 16{,}7.$$

Оскільки $|t_r| \geq t_p$ (16,7 > 2,31), то можна зробити висновок про *значущість* коефіцієнта парної кореляції між залежною і пояснювальною змінними.

Перевіримо статистичну значущість моделі парної лінійної регресії в цілому за допомогою критерію Фішера за формулою:

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

Розраховане значення статистики Фішера необхідно порівняти для числа ступенів вільності $k_1 = m;\ k_2 = n - m - 1$, де $m$ – кількість незалежних факторів моделі (для однофакторних моделей $m = 1$) та рівня значущості α = 0,05. Якщо $F_{факт} > F_{кр}\ (\alpha;\ k_1;\ k_2)$, то приймається гіпотеза, що модель є статистично значущою. При $F_{факт} < F_{кр}\ (\alpha;\ k_1;\ k_2)$ приймається протилежна гіпотеза про статистичну незначущість побудованої моделі.

Розраховане фактичне значення статистики Фішера дорівнює:

$$F_{факт} = (0{,}973 / (1 – 0{,}973))\cdot (10 – 2) = 288{,}9129.$$

За таблицею 1 Додатку критичне значення $F_{кр}$ (*0,05; 1; 8*) = 5,32.

Значення $F_{кр}$ (*0,05; 1; 8*) = 5,32 можна отримати за допомогою статистичної функції MS Excel

F.ОБР(вероятность;степени_свободы1; степени_свободы2).

У прикладі: F.ОБР(0,95; 1; 8).

*Висновок*. Оскільки $F_{факт}$ = 288,3 > $F_{кр}$ (*0,05; 1; 8*) = 5,32, то побудована модель є *статистично значущою*, тобто зв'язок між залежною і пояснювальною змінними *істотний*.

3) Побудована проста лінійна економетрична модель є адекватною і статистично значущою як за окремими параметрами, так і в цілому. Отримане рівняння залежності попиту на товар від ціни на нього може бути використано для прогнозу.

Нехай ціна на товар складатиме 15 грош.од., тоді прогнозований попит за рівнянням регресії:

$$\hat{y}_{пр} = 3{,}59 + 1{,}86 \cdot 15 = 31{,}49 \text{ (грош.од.)}.$$

З погляду прийнятих допущень отриманий прогноз є лише *точковою оцінкою* попиту, тому необхідно знайти інтервальні оцінки для отриманого прогнозу, що враховують помилку прогнозу:

$$\hat{y}_{пр} - \Delta\hat{y}_{пр} \le \hat{y}_{пр} \le \hat{y}_{пр} + \Delta\hat{y}_{пр}.$$

Помилку прогнозу $\Delta\hat{y}_{пр}$ знаходимо за формулою (1.38):

$$\Delta\hat{y}_{np} = 2{,}31 \cdot \frac{1{,}278}{\sqrt{10}} \cdot \sqrt{1 + (15 - 6{,}41)^2/13{,}6389} \approx 2{,}36.$$

Тут середнє квадратичне відхилення σ =1,278 знаходиться за формулою (1.31);

значення $var(x)$=13,6389 можна знайти в Excel за допомогою функції ДИСП.Г(X), де X – діапазон табличних значень вихідних даних.

Отримаємо $31{,}49 – 2{,}36 \le \hat{y}_{пр} \le 31{,}49 + 2{,}36.$

Для рівня значущості α = 0,05 одержуємо, що з імовірністю 0,95 істинні прогнозні значення попиту на товар лежать у таких межах:

$$29{,}13 \le \hat{y}_{пр} \le 33{,}85.$$

*Зауваження.* Коефіцієнти регресії $b_0$ та $b_1$ можна також знайти засобами Excel, використовуючи *Мастер функции* $f_x$:

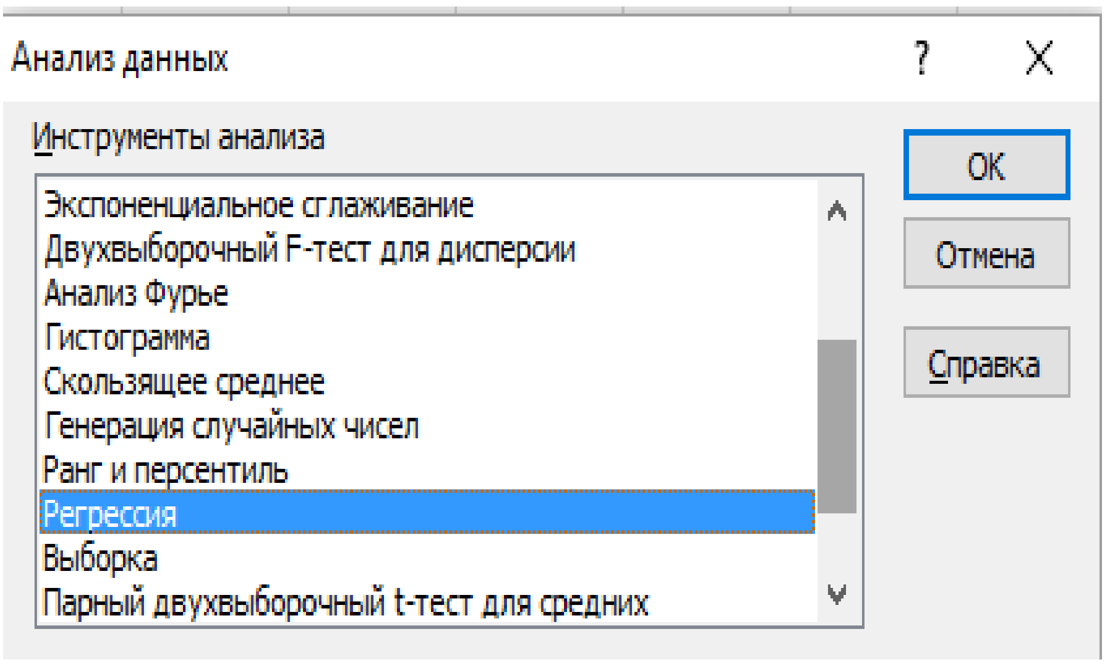
для $b_1$ : вибрати в категорії *Статистическая* функцію *ЛИНЕЙН*

для $b_0$ : вибрати в категорії *Статистическая* функцію *ОТРЕЗОК*. У вікні, що з'являється, ввести відповідні значення змінних $y$ та $x$.

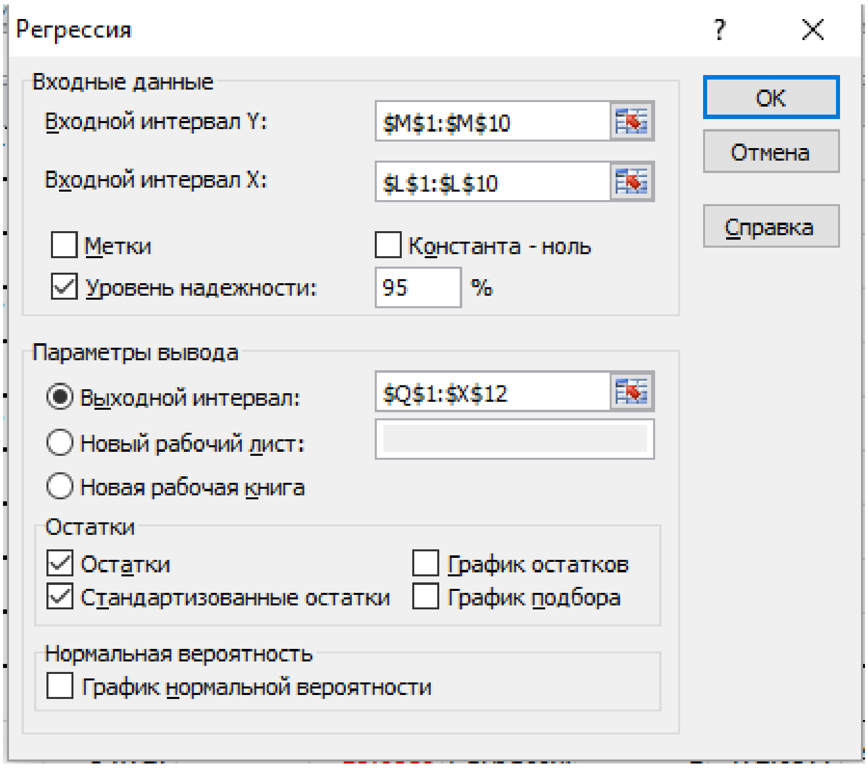
Для розрахунків необхідних величин, параметрів і коефіцієнтів регресії використовують: меню *Сервіс* → вікно *Анализ даннях* → розділ *Регрессия*. Для нових версій MS Excel використовується: вкладинка *Данные* → програма *Анализ данных* (якщо вона не встановлена, то треба спочатку перейти на *Настройку панели быстрого доступа*, а далі виконати такі дії: *Другие команды* → *Надстройки* → *Пакет анализа*.

У нижній частині вікна клацнути на кнопку *Перейти*, після чого у вікні *Надстройки* поставити прапорець біля *Пакет анализа,* підтвердивши свій вибір натисканням ОК).

У вікні *Анализ данных* вибираємо інструмент аналізу *Регрессия*




Після натискання ОК з’являється вікно *Регрессия*, де у відповідних смугах треба вказати вихідні дані та поставити необхідні прапорці:




Результати розрахунків зображені на рис.1.17.

| A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|
| ВЫВОД ИТОГОВ | | | | | | | | |
| | | | | | | | | |
| *Регрессионная статистика* | | | | | | | | |
| Множественный R | 0,986436045 | | | | | | | |
| R-квадрат | 0,97305607 | | | | | | | |
| Нормированный R-квадрат | 0,969688079 | | | | | | | |
| Стандартная ошибка | 1,278244981 | | | | | | | |
| Наблюдения | 10 | | | | | | | |
| | | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | *df* | *SS* | *MS* | *F* | *Значимость F* | | | |
| Регрессия | 1 | 472,0577182 | 472,0577182 | 288,912885 | 1,45692E-07 | | | |
| Остаток | 8 | 13,07128185 | 1,633910231 | | | | | |
| Итого | 9 | 485,129 | | | | | | |
| | | | | | | | | |
| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *жние 95,* | *рхние 95,0%* |
| Y-пересечение | 3,584793495 | 0,80970206 | 4,42729946 | 0,00220493 | 1,717617195 | 5,451969795 | 1,71762 | 5,45197 |
| Переменная X 1 | 1,860406631 | 0,109452182 | 16,99743759 | 1,4569E-07 | 1,608009447 | 2,112803815 | 1,60801 | 2,112804 |

Рис. 1.17

*Інтерпретація статистичних параметрів регресії*
(*Регрессионная статистика*)

*Множественный* $R$=0.9986436045 (це коефіцієнт кореляції $r$).

$R$-*квадрат* = 0.97305607 ($R^2$).

*Нормированный* $R$-*квадрат* = 0.969688079 (застосовується у множинній регресії).

*Стандартная ошибка* = 1,278244981 (осереднене відхилення значень ($x$, $y$) від лінії регресії, σ).

Наблюдения = 10 (розмір вибірки).

*Інтерпретація результатів аналізу змінюваності*
(*Дисперсионный анализ*)

Змінюваність значень функції $y = f(x)$ визначається двома складовими:

- Змінами лінії регресії (залежність результату $y$ від фактору $x$) – *Регрессия*
- Хаотичними (випадковими) змінами від інших факторів – *Остаток*
- Загальна – *Итого*

Поле *df* – це кількості степенем вільності, всього їх 10, з них 1 пов'язана зі змінами лінії регресії, інші 9 (10 – 1 = 9) – з хаотичними змінами величин ($x$, $y$).

Поле *SS* (*sum of squares*) – показники процесу мінімізації відхилень МНК; *SSR* = 472,0577182 – складова лінії регресії, *SSE* = 13,07128185 – хаотичних змін, *SST* = 485,129.

Поле *MS* (*mean squared*) – результати ділення суми квадратичних відхилень на кількість степенів вільності (*SS / df*: 472,0577182/1 = 472,0577182; 3,07128185 / 8 = 1,633910231), тобто це – середні квадратичні відхилення для регресії та для залишків.

Поле *F* (*F*-розподіл) – результати ділення середнього квадратичного відхилення для регресії на середнє квадратичне відхилення для залишків (472,0577182 / 1,633910231 = 288,912885).

Чим більшим є цей показник, тим більша статистична значущість регресії.

Поле *Значимость F* – близьке до нуля значення ($1{,}45692 \cdot 10^{-7}$), яке підтверджує статистичну значущість регресії і, відповідно, правдивість гіпотези про існування лінійної залежності між $y$ та $x$.

*Оцінки статистичних параметрів та їх значущість*

Поле *Коэффициенты* – значення коефіцієнтів $b_0$ = 3,5848 (*Y-пересечение*) та $b_1$ =1,8604 (*X*) рівняння регресії.

Поле *Стандартная ошибка* - стандартні помилки для $b_0$ ($\sigma_{b_0} \approx 0{,}8097$) та $b_1$ ($\sigma_{b_1} \approx 0{,}1095$).

Поле $t$-статистика – результати ділення коефіцієнтів $b_0$ та $b_1$ на їх стандартні помилки: $t_{b_0} \approx 4{,}4273$ і $t_{b_1} \approx 16{,}9974$.

Поле *P-Значение* – двосторонні $p$-оцінки для $t$-статистики (згідно із $t$-розподілом ймовірностей з 9 степенями вільності).

Дві пари полів *Нижние 95%* та *Верхние 95%* – 95%-ві довірчі інтервали для коефіцієнтів $\beta_0$ (1,7176; 5,4520) та $\beta_1$ (1,6080; 2,1128). Ці оцінки для результату $y$ з ймовірністю 95% стверджують, що при підвищенні значення $x$ на 1 значення $y$ зростає у визначеному діапазоні (це – границі шлейфа нижче й вище лінії регресії).

У першій парі полів наведені 95%-ні довірчі інтервали, у другій – довірчі інтервали, що вказуються у вікні "Регрессия" (там було встановлено 95%, тому вони й співпали).

4) Обчислимо коефіцієнт еластичності при $x$ = 12 грош.од. за формулою (1.42):

$$E_{12} = \frac{1{,}86 \cdot 12}{3{,}58 + 1{,}86 \cdot 12} \approx 0{,}86.$$

*Висновок*. У разі збільшення значення основних засобів на 1% валовий регіональний продукт збільшиться наближено на 0,86%. Оскільки $E_{12}$ = 0,86 < 1, то валовий продукт є нееластичним (відносний приріст її спадає).

## 1.10. Завдання для самостійної роботи

### 1.10.1. Тести

*Вибрати правильну відповідь на запитання*

1. Лінійна регресія:
   а) лінія, що відображає зв'язок між незалежною і залежною змінними;
   б) інша назва простої регресії;
   в) лінія, яка завжди має нахил, що дорівнює 1;
   г) графік значень незалежної і залежної змінних;
   д) лінія, яка завжди має нахил, що дорівнює 0.

2. Що з наведеного не є припущенням моделі лінійної регресії:
   а) або $x_i$ є сталими числами, або вони є статистично незалежними від випадкових величин $\varepsilon_i$.
   б) дисперсія випадкової величини $\varepsilon_i$ є сталою;
   в) математичне сподівання випадкової величини $\varepsilon_i$ дорівнює 0;
   г) дисперсія випадкової величини дорівнює 0;
   д) випадкові величини є статистично незалежними.

3. Коефіцієнт парної кореляції змінюється в межах:
   а) від 0 до 1;
   б) від -1 до 0;
   в) від -1 до 1.

4. Коефіцієнт детермінації вимірює:
   а) варіацію незалежної змінної;
   б) перетин лінії регресії;
   в) загальну варіацію залежної змінної, яка пояснюється регресією.

5. Коефіцієнт детермінації:
   а) точка, де лінія регресії перетинає вісь ординат;

б) вимірює придатність лінії регресії;

в) вимірює зв'язок між незалежною і залежною змінними;

г) завжди дорівнює 1;

д) завжди дорівнює 0.

6. Вільний член у рівнянні регресії – це:

а) точка, у якої лінія регресії перетинає вісь ординат;

б) точка, у якої лінія регресії перетинає вісь абсцис;

в) завжди дорівнює 1.

7. Якщо значення коефіцієнта парної кореляції між факторами $x$ і $y$ дорівнює 0,75, то це означає:

а) немає статистично значущого зв'язку між факторами;

б) зі збільшенням фактору $x$ фактор $y$ збільшується;

в) зі збільшенням фактору $x$ фактор $y$ зменшується.

8. З урахуванням співвідношення між заробітною платою (в гривнях) – $y$ і освітою (в роках) – $x$, $y = 12.201 + 525\,x$, особа, що навчалася додатково нуль років, може очікувати на таку додаткову оплату:

а) 12.201; б) 525; в) 24.402; г) 1.050.

9. Якщо регресія має $R^2$=0.8, то регресійна лінія:

а) пояснює на 80% варіації змінної $x$;

б) пояснює на 80% варіації змінної $y$;

в) матиме нахил 0.80;

г) матиме перетин 0.80;

д) не пояснює зв'язку між $x$ і $y$.

10. Якщо $b_1 < 0$, то це означає, що $x$ має дуже малий вплив на $y$:

а) так; б) ні.

11. Для перевірки значущість окремого параметра використовують:

а) $F$-тест;

б) $t$-тест;

в) $\chi^2$–тест;

г) біноміальний розподіл;

д) експоненціальний розподіл.

12. Для перевірки значущості одночасно всіх параметрів використовується:

    а) *F*-тест;

    б) *t*-тест;

    в) $\chi^2$–тест;

    г) біноміальний розподіл;

    д) експоненціальний розподіл.

13. Ступені вільності для *t*-статистики для перевірки значимості параметрів регресії, що складається з 35 спостережень та 3 незалежних змінних такі:

    а) 35; б) 3; в) 32;

    г) 33; д) 31.

14. Ступені вільності чисельника *F*-статистики в регресії, що складається з 50 спостережень та 4 незалежних змінних, такі:

    а) 50; б) 4; в) 3;

    г) 46; д) 45.

15. Ступені вільності знаменника *F*-статистики, що складається з 50 спостережень та 4 незалежних змінних, такі:

    а) 50; б) 4; в) 3;

    г) 46; д) 45.

### 1.10.2. Контрольні запитання

1. Що таке “кореляційна залежність"?
2. Опишіть схему побудови аналітичного групування.
3. Що таке “емпірична лінія регресії"?
4. У чому полягає відмінність між теоретичною і емпіричною лініями регресії?

5. Назвіть основні етапи кореляційно-регресійного аналізу.
6. Що розуміють під специфікацією моделі і як її здійснюють?
7. Визначення узагальненої парної лінійної кореляційно-регресійної моделі.
8. Визначення вибіркової парної лінійної кореляційно-регресійної моделі.
9. Дайте тлумачення випадкової складової економетричної моделі.
10. Назвіть відомі Вам критерії якості оцінювання параметрів економетричних моделей.
10. У чому сутність методу максимальної правдоподібності.
11. У чому сутність методу найменших квадратів.
12. Чим відрізняються метод максимальної правдоподібності від методу найменших квадратів.
13. Дайте економетричну інтерпретацію параметрів парної лінійної кореляційно-регресійної моделі.
14. Основні припущення класичного кореляційно-регресійного аналізу.
15. Як вимірюють силу зв'язку між кореляційно пов'язаними змінними?
16. Які властивості має коефіцієнт кореляції?
17. *У* чому полягає суть статистичної значущості коефіцієнтів регресії?
18. Поясніть суть коефіцієнта детермінації.
19. Які властивості коефіцієнта детермінації?
20. Що таке «зона надійності регресії»?
21. Що показує коефіцієнт еластичності і за якою формулою він обчислюється?
22. Як проводиться економетричний аналіз факторних ознак?
23. Як інтерпретується оцінений коефіцієнт еластичності?
24. Чи може бути $t$-статистика від'ємною?

### 1.10.3. Завдання для самостійної роботи

У таблиці 1.6 наведені щомісячні витрати $x$ в грошових одиницях фірми «Ваш будинок» на рекламу та обсяги її щомісячних продаж $y$ в грошових одиницях.

1. Побудувати лінійну регресію і розрахувати параметри.
2. Дати економічну інтерпретацію параметрів регресії.
3. Обчислити середні квадратичні відхилення параметрів моделі.
4. Побудувати спряжену парну лінійну регресійну модель.
5. Побудувати довірчий інтервал оцінки фактичного значення обсягу продаж.
6. Знайти коефіцієнти кореляції і детермінації. Які висновки можна зробити щодо якості регресійної моделі?
7. Розрахувати $F$-критерій Фішера, знайти за таблицею $F$-розподілу критичне значення. Який висновок щодо адекватності регресійної моделі можна зробити?
8. Побудувати довірчий інтервал для істинних значень параметрів лінійної моделі.
9. Обчислити стандартну вибіркову похибку моделі і побудувати довірчий інтервал для фактичного значення обсягу продаж, яке відповідає значенню витрат на рекламу $7{,}6 + 0{,}1 \cdot N$ (грош. од.)
10. Обчислити стандартну вибіркову похибку та побудувати довірчий інтервал для конкретного значення обсягу продаж у регіоні, в якому величина затрат на рекламу дорівнює $7{,}6 + 0{,}1 \cdot N$ (грош. од.).
11. Обчислити коефіцієнт еластичності при $x = 7{,}6 + 0{,}1 \cdot N$ (грош.од.).

Таблиця 1.6

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Витрати на рекламу | $7{,}0+k$ | $7{,}1+k$ | $7{,}2+k$ | $7{,}3+k$ | $7{,}4+k$ | $7{,}5+k$ | $7{,}7+k$ | $8{,}0+k$ | $8{,}1+k$ |
| Обсяг продаж | $48+N$ | $49+N$ | $50+N$ | $51+N$ | $52+N$ | $53+N$ | $54+N$ | $55+N$ | $56+N$ |

($k = 0{,}1 \cdot N$, $N$ – номер варіанту студента за списком групи).

# РОЗДІЛ 2. ЛІНІЙНА БАГАТОФАКТОРНА МОДЕЛЬ

## 2.1. Приклади використання багатофакторного регресійного аналізу в економічних дослідженнях

У Розділі 1 розглядались економічні задачі, в яких результативна економічна змінна (товарообіг) залежить тільки від однієї незалежної змінної. У парній регресії було показано, що коли значення коефіцієнта детермінації менше одиниці, то на величину товарообігу $Y$, окрім одного фактора $X$, впливають ще й інші фактори. Бажання врахувати вплив не одного, а декількох факторів, що діють й впливають одночасно, привело до необхідності переходу від парної (однофакторної) до множинної (багатофакторної) регресії.

Апарат множинної регресії існує більше ста років. Поняття «множинна регресія» уперше було введено в практику К. Пірсоном у 1908 році і стосувалося досліджень у біології. Методика визначення форми залежності між залежною величиною $Y$ та $n$ – незалежними факторами $x_1, x_2,\ldots, x_n$ у вигляді функції $Y = f\ (x_1,\ldots, x_n)$ дає можливість більш об'єктивно оцінити або визначити значення певного показника залежно від декількох впливаючих факторів.

Наведемо приклади [19] використання множинного регресійного аналізу в економічних дослідженнях.

*Приклад*. На ринку нерухомості ріелтор формує реєстр, де фіксує такі фактори: розмір житлової площі, кількість кімнат, район, транспортне сполучення, рівень сервісу та інші параметри, які мають вплив на вартість житла. Коли ця інформація зібрана для різних будинків, районів та умов проживання, є можливість установити зв'язок між цими параметрами й ціною, за якою житло було продано, у вигляді відповідної функції. Знайдені коефіцієнти для факторів (незалежних змінних) визначають «вагу» відповідних параметрів й з'являється можливість досить об'єктивно й зважено визначати ціну житла, призначеного на продаж. Ця ж функція дозволить з єдиних позицій переглянути вже існуючу статистику, де можуть виявитися суттєві відхилення від сформованих закономірностей («викиди»), які означають занижені чи завищені ціни для відповідних об'єктів.

*Приклад*. У роботі з кадрами множинна регресія дозволяє краще оцінити продуктивність і ефективність роботи працівників, більш об'єктивно визначити форму і розмір оплати праці. Факторами, що необхідно врахувати, можуть бути: рівень відповідальності, кількість виконавців, якими керує фахівець, якість впровадження тощо. Маючи відповідну статистику, менеджер із роботи з кадрами має змогу побудувати відповідне рівняння множинної регресії, за яким можна обчислити суму нарахувань, а також проаналізувати попередній досвід у цій роботі, щоб побачити, де були переплати чи недоплати.

*Приклад*. У реальному житті кожна економічна система зазвичай є сукупністю великої кількості об'єктів, процесів, явищ, що взаємодіють між собою, перебувають у певних причинно-наслідкових відношеннях. Кожна економічна змінна здебільшого 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 використовують на підприємствах, ефективності управлінських рішень тощо.

*Приклад*. До складу доходу консолідованого бюджету України на 2017 рік, входять прямі податки (на доходи) підприємств та домашніх господарств і непрямі податки (ПДВ — податок на додану вартість, акцизні збори). Доход також може залучати стягнення із зарплати (пенсійний фонд, військовий збір) та нефіскальні стягнення (наприклад, доходи від приватизації) і т. ін. Отже, при аналізі та прогнозуванні доходу консолідованого бюджету України необхідно дослідити вплив на його величину податків на додану вартість, податків на доходи підприємств, податків з населення, акцизного збору, пенсійного фонду, військового збору та інших, тобто здійснити багатофакторний аналіз.

*Приклад*. При дослідженні процесу малої приватизації можна бачити, що кількість приватизованих підприємств змінюється не тільки з часом, але й під впливом багатьох факторів, наприклад, індексу реальної продукції, рівня інфляції оптових цін, обсягів кредитів тощо.

*Приклад*. Проблема лібералізації цін залишається досить актуальною для України. На відміну від стабілізації валюти та приватизації, необхідність яких була загальновизнаною, лібералізація цін довго залишалася неузгодженим питанням для прихильників різних поглядів на розвиток економіки України. Для дослідження процесу лібералізації цін на ринках України також важливо виявити та дослідити вплив різноманітних факторів. Прикладами таких факторів є відхилення у відсотках ринкового обмінного курсу валют від встановленого державою (можна вважати цифру відхилення аналогією до відповідного "податку" на операції з обміну валют), від обмеження експорту (розраховується як частка експорту, що здійснюється за ліцензіями та встановленими державою квотами), від адміністративного контролю за внутрішньою торгівлею (розраховується як частка загального випуску продукції, що здійснюється за державними замовленнями і контрактами, а не за рішенням підприємств), від контролю за цінами роздрібної торгівлі (розраховується як  частка загального випуску продукції, що підлягає державному управлінню), від контролю за оптовими цінами, від частки валютних операцій, що здійснюються через неринкові, адміністративні механізми та інші.

*Приклад*. За допомогою множинної регресії можна аналізувати соціальні проблеми. Наприклад, досліджувати, як впливають різні соціальні фактори на кількість зареєстрованих в Україні злочинів останнього десятиліття. Для аналізу та побудови багатофакторної регресійної моделі можна обрати такі фактори, як загальна кількість населення в Україні, забезпеченість населення житлом, продаж алкогольних напоїв у розрахунку на душу населення, кількість людей, що мають вищу та середню освіту в розрахунку на 1000 осіб, реальну середню заробітну плату робітників та службовців, рівень безробіття, рівень культурногорозвидку нації та інші.

Досить актуальною в Україні залишається проблема аналізу народжуваності, а саме впливу різноманітних факторів на кількість народжених. Серед таких факторів виділимо кількість жіночого населення віком від 16 до 47 років, частку пенсіонерів у загальній кількості населелення України, реальні грошові доходи населення, кількість постійних дошкільних закладів, кількість зареєстрованих шлюбів, кількість зареєстрованих розлучень, кількість абортів, кількість хворих з діагнозом на алкоголізм та наркоманію тощо.

Аналогічні приклади можна знайти у навчально-виховній роботі, при оцінюванні громадської думки тощо, але найбільш ефективно цей апарат діє у сфері економіки та бізнесу, де для якісного прогнозу треба враховувати різноманітні фактори різної природи, починаючи від настроїв клієнтів чи погодних умов до світової фінансової чи іншої політики.

Отже, спільний вплив кількох факторів на одну результуючу змінну досліджують за допомогою множинних економетричних моделей. Узагальнюючи розглянуті вище приклади з економічним змістом, зазначимо, що саме множинний регресійний аналіз допомагає знайти явний вигляд залежності досліджуваного показника від численних факторів, що впливають на його зміну, а також кількісно оцінити їхній вплив.

У випадку багатофакторного регресійного аналізу розв'язування проблем форми зв'язку натикається на такі труднощі, яких не було у випадку парної регресії. Так, якщо у виборі рівняння регресії *Y* за *X* можливо було обґрунтувати в деякій мірі емпіричні лінії регресії, то в множинному регресійному аналізі такої можливості немає. Крім того, між існуючими факторами можуть існувати небажані залежності, що роблять неможливим визначення параметрів лінійного рівняння регресії. Складність розрахунків та узагальнення інформації призводять до необхідності широкого використання обчислювальної техніки. У зв'язку з цим побудова та аналіз множинних регресійних моделей використовує сучасні пакети прикладних програм, а також уміння аналізувати отримані результати та робити за ними висновки, вміти оцінити найкращу модель для взаємозв'язаних статистичних даних.

## 2.2. Класична лінійна множинна модель

Розглянемо використання лінійної множинної (багатофакторної) кореляційно-регресійної моделі під час дослідження впливу кількох факторів на результуючу змінну.

*Узагальнена лінійна множинна кореляційно-регресійна модель* описує кореляційну залежність результуючої змінної $y$ від факторних ознак $x_1, x_2, \ldots, x_k$ *для всієї генеральної сукупності*. Вона має вигляд:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon. \qquad (2.1)$$

Тут: $y$ – результуюча (залежна) змінна, або регресант;

$x_1, x_2, \ldots, x_k$ – факторні (незалежні) змінні(ознаки), або регресори;

$\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ – істинні (невідомі) параметри моделі (константи), які треба оцінити; $\beta_0$ - вільний член рівняння регресії, $\beta_1, \beta_2, \ldots, \beta_k$ – коефіцієнти множинної регресії;

$\varepsilon$ – не спостережувана випадкова величина і можна лише зробити припущення відповідно до закону її розподілу;

$k$ - кількість факторів.

На відміну від узагальненої регресійної моделі, *вибіркова модель будується для певної вибірки*. Невідомі параметри вибіркової моделі є випадковими величинами, математичне сподівання яких дорівнює параметрам узагальненої моделі (випадок класичної лінійної регресії), випадкові величини (помилки) можна оцінити, виходячи з вибіркових даних.

Якщо маємо вектори спостережень за результуючою змінною

$$Y = (y_1, y_2, \dots, y_n)$$

та факторними ознаками $x_1, x_2, \dots, x_k$

$$x_1 = (x_{11}, x_{12}, \dots, x_{1n});$$
$$x_2 = (x_{21}, x_{22}, \dots, x_{2n});$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$x_k = (x_{k1}, x_{k2}, \dots, x_{kn}),$$

то на підставі даних цих спостережень можна побудувати вибіркову лінійну множинну кореляційно-регресійну модель:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k, \qquad (2.2)$$

де $\hat{y} = y^{\text{теор}}$ – теоретичне значення результуючої змінної $y$;

$b_0, b_1, b_2, \dots, b_k$ – оцінки невідомих параметрів $\beta_0,\ \beta_1,\ \beta_2, \dots,\ \beta_k$.

Вибіркова лінійна багатофакторна модель має такий вигляд:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + e, \qquad (2.3)$$

де $y$ – залежна змінна;

$x_1, x_2, \dots, x_k$ – незалежні змінні (фактори);

$b_0,\ b_1,\ b_2, \dots, b_k$ – оцінки невідомих параметрів узагальненої моделі (2.1);

$e$ - випадкова величина (помилка, відхилення моделі):

$$e_i = \hat{y}_i - \tilde{y}_i, \quad i = \overline{1, n}.$$

*Лінійною регресійною моделлю називається модель, яка лінійна за своїми параметрами.*

За введеними позначеннями, множинна модель має $k$ незалежних змінних (факторів), які впливають на залежну змінну $Y$, та $(k + 1)$ невідомих параметрів, які потрібно оцінити.

Лінійну теоретичну модель (2.1) можна записати у вигляді системи лінійних алгебраїчних рівнянь:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \dots + \beta_k x_{1k} + \varepsilon_1, \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \dots + \beta_k x_{2k} + \varepsilon_2, \\ y_3 = \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \beta_3 x_{33} + \dots + \beta_k x_{3k} + \varepsilon_3, \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \dots + \beta_k x_{nk} + \varepsilon_{n,} \end{cases} \qquad (2.1.1)$$

або у векторно-матричній формі:

$$Y = X \cdot \beta + \varepsilon, \qquad (2.1.2)$$

де $$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & x_{31} & x_{32} & & x_{3k} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix},$$

$$\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^{\mathrm{T}}, \quad \varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_n)^{\mathrm{T}}.$$

Емпірична модель є статистичним аналогом теоретичної моделі. За її допомогою визначаються статистичні оцінки параметрів (2.1.1). Використовується при цьому статистична обробка вибірки.

Емпірічна модель має такий вигляд:

$$\begin{cases} y_1 = b_0 + b_1 x_{11} + b_2 x_{12} + b_3 x_{13} + \dots + b_k x_{1k} + e_1 \\ y_2 = b_0 + b_1 x_{21} + b_2 x_{22} + b_3 x_{23} + \dots + b_k x_{2k} + e_2 \\ y_3 = b_0 + b_1 x_{31} + b_2 x_{32} + b_3 x_{33} + \dots + b_k x_{3k} + e_3 \\ \vdots \\ y_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + b_3 x_{n3} + \dots + b_k x_{nk} + e_{n,} \end{cases} \qquad (2.3.1)$$

або у векторно-матричній формі:

$$Y = X \cdot b + e,$$

де $b = (b_0, b_1, b_2, \dots, b_k)^{\mathrm{T}}$, $e = (e_1, e_2, e_3, \dots, e_n)^{\mathrm{T}}$.

Вектор $b$ - статистична оцінка теоретичного вектора $\beta$ лінійної множинної регресії (2.1.2), вектор похибок $e$ – статистична оцінка випадкового вектора $\varepsilon$ цієї ж моделі.

У разі узагальненої регресійної моделі, дійсної для всієї генеральної сукупності, випадкова величина $\varepsilon$ є *не спостережуваною величиною, і можна зробити лише деякі* *припущення щодо її поведінки та закону розподілу*. Для класичної множинної регресійної моделі (2.1), яка є узагальненням простої лінійної регресійної моделі (1.3), *всі основні класичні припущення зберігаються, але дещо модифікуються*.

## 2.3. Основні припущення у множинному кореляційно-регресійному аналізі

У класичному множинному кореляційно-регресійному аналізі до узагальненої моделі (2.1) висувають такі припущення.

1°. Математичне сподівання випадкової величини $\varepsilon$ дорівнює нулю:

$$M(\varepsilon) = 0.$$

2°. Випадкові величини $\varepsilon_i$ в різних точках незалежні між собою:

$$cov(\varepsilon_i, \varepsilon_l) = \overline{\varepsilon_\iota \cdot \varepsilon_l} - \overline{\varepsilon_\iota} \cdot \overline{\varepsilon_l} = 0, \qquad i = \overline{1, n}, \qquad l = \overline{1, n}; \; i \neq j.$$

3°. Модель гомоскедастична, тобто має однакову дисперсію для будь-якого спостереження:

$$D(\varepsilon_i) = \sigma^2, \qquad i = \overline{1, n}.$$

4°. Коваріація між випадковою величиною $\varepsilon_i$ та кожною незалежною змінною $x$ дорівнює нулю:

$$cov(x_{ij}, \varepsilon_i) = 0, \quad j = \overline{1, k}, \quad i = \overline{1, n}.$$

Факторні ознаки $x_1, x_2, \ldots, x_k$ є нестохастичними змінними,тобто не залежать від випадкових величин $\varepsilon$.

Зазначимо, що ця властивість виконується автоматично, якщо $x_j$ $(j = \overline{1, k})$ не стохастичні та перше припущення має силу.

5°. Модель повинна були правильно специфікованою.

6°. Випадкова величина $\varepsilon$ має нормальний закон розподілу з математичним сподіванням, що дорівнює нулю, і дисперсією, яка дорівнює $\sigma^2$, тобто щільність розподілу $\varphi(x)$ випадкової величини $\varepsilon$ має вигляд:

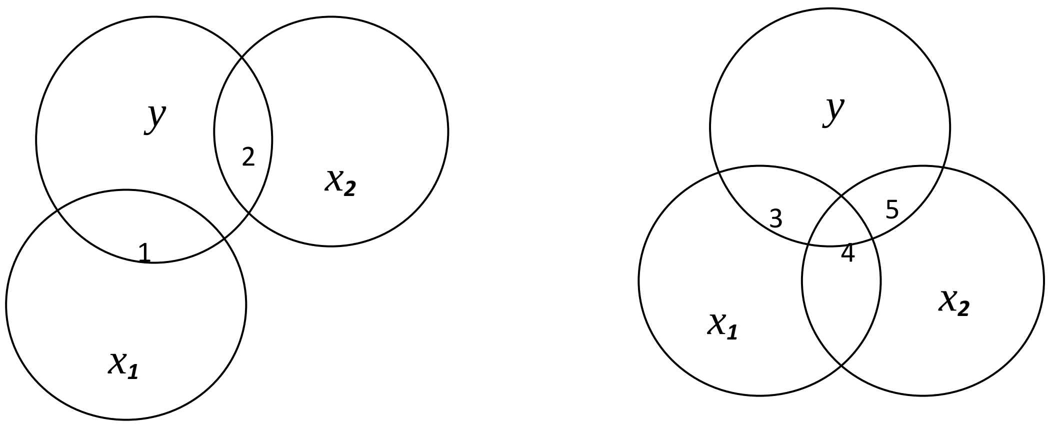
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

7°. Факторні ознаки $x_1, x_2, \ldots, x_k$ незалежні між собою (відсутність мультиколінеарності):

$$cov(x_l, x_j) = 0, \quad l = \overline{1,k}, \quad i = \overline{1,k}.$$

Не повинно бути точного лінійного зв'язку між двома або більше факторами. Припущення 7 не прийнятне для простої парної лінійної регресії, але воно надзвичайно важливе для множинної регресії.

Припустимо, що є лінійна залежність між факторами $x_1$ та $x_2$. Неможливо точно визначити окремий вплив кожного з цих факторів на залежну змінну $y$. Графічно це можна подати, виходячи з кругової діаграми (рис.2.1). На рис. 2.1 *а*) підмножина 1 описує вплив фактора $x_1$ на $y$, підмножина 2 описує вплив фактора $x_2$ на $y$. Підмножини 1 та 2 спільної частини не мають, в цьому випадку фактори $x_1$ і $x_2$ незалежні (неколінеарні). На рис. 2.1 *б*) підмножина 3 описує окремий вплив фактора $x_1$, а підмножина 5 - окремий вплив фактора $x_2$. Підмножина 4 описує як вплив обох факторів $x_1$ та $x_2$ на зміну $y$, так і вплив одного з факторів на інший. Саме підмножина 4 графічно описує ситуацію колінеарності. Коли підмножина 4 - пуста множина, то колінеарності немає, що і показано на рис. 2.1 *а*).




*а*) *б*)

Рис. 2.1. Зв'язок між факторами:

*а*) – відсутність залежності між факторами $x_1$ і $x_2$;

*б*) – наявність такої залежності.

У тому випадку, *коли між факторами існує лінійна статистична залежність*, кажуть, що *для моделі існує явище мультиколінеарності*. На рис. 2.1 *а*) графічно зображений випадок відсутності мультиколінеарності між факторами $x_1$ та $x_2$, а на рис. 2.1 *б*) - її наявності.

## 2.4. Етапи побудови множинної регресійної моделі

Процес побудови множинної регресійної моделі є більш складним, ніж побудова парної лінійної регресії, тому що крім технічних труднощів виникає проблема мультиколінеарності, якої не було у випадку парної регресії.

Можна виділити такі основні етапи побудови вибіркової множинної лінійної кореляційно-регресійної моделі.

1. Вибір та аналіз усіх можливих факторів, які впливають на процес (або показник), що вивчається.
2. Вимір та аналіз знайдених факторів.
3. Математико-статистичний аналіз факторів.
4. Вибір методу та побудова регресійної множинної моделі.
5. Оцінка невідомих параметрів регресійної моделі.
6. Перевірка моделі на адекватність.
7. Розрахунок основних характеристик та побудова інтервалів довіри.
8. Аналіз отриманих результатів, висновки.

Розглянемо детально кожний із етапів побудови множинної регресійної моделі.

На *першому етапі* потрібно глибоко розуміти суть економічного об'єкта, процесу, явища, дослідити його з макроекономічних та мікроекономічних позицій; виявити якомога більше факторів, які в конкретному випадку можуть справити суттєвий або несуттєвий вплив на його зміну. Наприклад, під час дослідження впливу на валовий регіональний продукт можна виділити такі фактори:

величина основних засобів, величина інвестицій в основний капітал, величина оборотних фондів, кількість зайнятого населення в регіоні, кількість упроваджених прогресивних технологічних процесів у виробництві, ефективність управлінських рішень. На цьому етапі можуть знадобитися поради практиків, які працюють у галузі або на підприємстві, що досліджується, і таке інше.

На *другому етапі* проводять кількісний аналіз відібраних факторів: оцінювання можливості їх кількісного вираження, підбирання та розроблення шкал оцінювання якісних факторів, проведення спостережень та вимірювання (збирання статистичних даних), щоб отримати емпіричні дані вибірки. У вищенаведеному прикладі величини основних засобів, інвестицій в основний капітал та оборотних фондів є кількісними факторами, одиницею вимірювання яких є мільйон гривень; трудові ресурси - кількісний фактор, одиниця вимірювання - тисяча осіб; кількість упроваджених технологічних процесів вимірюють в умовних одиницях, а ефективність управлінських рішень - якісний фактор, шкалу оцінювання якого розробити важко.

За результатами проведення другого етапу деякі вибрані фактори, можливо, доведеться відкинути.

Після того, як усі фактори проаналізовано, подано у кількісному вигляді, тобто у вигляді динамічних або варіаційних рядів, переходять до *третього етапу* - етапу математико-статистичного аналізу, який є найважливішим підготовчим етапом для побудови регресійної багатофакторної моделі. Це заключний етап формування необхідної інформаційної бази.

При наявності у динамічних рядах недостатньої інформації за допомогою спеціальних методів проводиться її відтворення. На цьому етапі проводиться перевірка основних припущень класичного регресійного аналізу, крім того, здійснюється найважливіша процедура множинного аналізу - перевірка факторів на мультиколінеарність. Для цього спочатку будується матриця коефіцієнтів парної кореляції ($k$+1)-го порядку, яка є симетричною і має такий вигляд:

$$R = \begin{pmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & & r_{yx_k} \\ r_{x_1y} & r_{x_1x_1} & r_{x_1x_2} & & r_{x_1x_k} \\ r_{x_2y} & r_{x_2x_1} & r_{x_2x_2} & \dots & r_{x_2x_k} \\ \cdot & \cdot & \cdot & & \cdot \\ r_{x_ky} & r_{x_kx_1} & r_{x_kx_2} & & r_{x_kx_k} \end{pmatrix},$$

де $r_{x_ix_j} = r_{x_jx_i}$ $(i = j = \overline{1,k})$ - коефіцієнт парної кореляції між $i$-м та $j$-м факторами;

$r_{yx_j}$- коефіцієнт кореляції між залежною змінною $y$ та $j$-м фактором.

Елементи матриці $R$ обчислюються за формулою:

$$r_{x_ix_j} = \frac{\sum_{l=1}^{n}(x_{li} - \overline{x_i})(x_{lj} - \overline{x_j})}{\sqrt{\sum_{l=1}^{n}(x_{li} - x_i)^2 \sum_{l=1}^{n}(x_{lj} - x_j)^2}} = \frac{\overline{x_ix_j} - \overline{x_i} \cdot \overline{x_j}}{\sigma_{x_i} \cdot \sigma_{x_j}} = \frac{cov(x_i, x_j)}{\sigma_{x_i} \cdot \sigma_{x_j}},$$

Далі аналізуються коефіцієнти парної кореляції між факторами (зазначимо, що $\left|r_{x_ix_j}\right| \leq 1$). Якщо модуль значення деяких з них близький до одиниці, то це вказує на щільний зв'язок між ними, або на мультиколінеарність. У цьому випадку один з факторів треба залишити, а інший вилучити із подальшого розгляду. Постає питання: який саме фактор? Це залежить від конкретної ситуації. Найчастіше той фактор, який з економічної точки зору більш вагомий для аналізу впливу на залежну змінну. Можна також залишити фактор, який має більший коефіцієнт кореляції із залежною змінною $y$. Такий аналіз проводиться для кожної пари залежних між собою факторів. Результатом етапу математико-статистичного аналізу є знаходження множини основних незалежних між собою факторів, які є базою для побудови регресійної моделі.

Метод побудови регресійної множинної моделі (*четвертий етап*) впливає на остаточний вигляд регресійної моделі. Найбільш поширеним методом побудови множинної лінійної кореляційно-регресійної моделі є *МНК або його модифікації.*

Оцінка невідомих параметрів моделі (*п'ятий етап*) $\beta_0,\ \beta_1,\ \beta_2, \dots,\ \beta_k$ здійснюється у лінійних регресійних моделях за МНК, як у випадку простої лінійної регресії.

Після того, як параметри знайдено, проводиться перевірка моделі на адекватність (*шостий етап*) за *F*-критерієм Фішера, а також перевірка значущості знайдених параметрів за *t*-критерієм Ст'юдента. Якщо модель неадекватна, то необхідно повернутися до етапу побудови моделі і, можливо, від лінійної моделі перейти до нелінійної, або ввести додаткові фактори.

У випадку адекватності моделі виконується *сьомий етап* : розраховуються основні характеристики та будуються інтервали довіри. Перевірку множинної лінійної регресійної моделі на точність здійснюють за такою ж схемою, як і для парної лінійної моделі: обчислюють стандартну похибку моделі, коефіцієнти множинної детермінації та кореляції, вибіркову похибку моделі та похибку індивідуального прогнозу, проводять експрес-діагностику моделі. Можна також будувати рівняння часткових регресій і обчислювати коефіцієнти часткової кореляції.

Під час побудови лінійної множинної регресійної моделі керуються двома суперечливими критеріями: з одного боку, чим більше факторних ознак умістить модель, тим вона точніша, з іншого боку, включення в модель великої кількості факторів вимагає більших обсягів емпіричних даних і значно ускладнює процес побудови та дослідження моделі. На *восьмому етапі* здійснюється вибір кінцевої моделі, який полягає у компромісі між цими двома критеріями, а також провести аналіз отриманих результатів (зокрема, зробити висновок про вплив окремих факторів на залежний показник). Це дасть змогу ефективніше керувати економічними системами, прогнозувати результати їх функціонування.

Множинна кореляційно-регресійна модель дає можливість зобразити залежність економічної змінної, яку досліджують, від численних факторів, що впливають на її поведінку, кількісно оцінити їхній вплив, виявити найсуттєвіші фактори.

## 2.5. Знаходження оцінок параметрів лінійного рівняння регресії

Нехай у результаті експерименту або спостережень одержані такі дані (таблиця 2.1), де $k$ – число незалежних між собою факторів, $n$ - число проведених експериментів.

Таблиця 2.1

| | $x_1$ | $x_2$ | … | $x_j$ | … | $x_k$ | $y$ |
|---|---|---|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{12}$ | … | $x_{1j}$ | … | $x_{1k}$ | $y_1$ |
| 2 | $x_{21}$ | $x_{22}$ | … | $x_{2j}$ | … | $x_{2k}$ | $y_2$ |
| … | … | … | … | … | … | … | … |
| $i$ | $x_{i1}$ | $x_{i2}$ | … | $x_{ij}$ | … | $x_{ik}$ | $y_i$ |
| … | … | … | … | … | … | … | … |
| $n$ | $x_{n1}$ | $x_{n2}$ | … | $x_{nj}$ | … | $x_{nk}$ | $y_n$ |

Розглянемо побудову лінійної множинної регресійної моделі за допомогою МНК.

У першому наближенні вважають, що обсяг вибірки, на підставі якої оцінюють параметри моделі, має бути щонайменше у десять разів більший від кількості факторних ознак, до того ж, чим тісніше факторні змінні корелюють між собою, тим більшим повинен бути обсяг вибірки.

Критерій МНК для оцінювання параметрів $\beta_0,\ \beta_1,\ \beta_2, \dots ,\ \beta_k$ узагальненої моделі має такий вигляд:

$$Q(b_0, b_1, b_2, \dots, b_k) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 =$$

$$= \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \dots - b_k x_{ki})^2 \to min.$$

Необхідною умовою мінімуму функції $Q(b_0, b_1, b_2, \dots, b_k)$ є одночасна рівність нулю частинних похідних:

$$\frac{\partial Q}{\partial b_0} = 0, \quad \frac{\partial Q}{\partial b_1} = 0, \quad \frac{\partial Q}{\partial b_2} = 0, \dots, \frac{\partial Q}{\partial b_k} = 0,$$

або

$$\begin{cases} -2\sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_k x_{ik}) = 0 \\ -2\sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_k x_{ik}) x_{i1} = 0 \\ -2\sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_k x_{ik}) x_{i2} = 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ -2\sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_k x_{ik}) x_{ik} = 0. \end{cases}$$

Наведену систему лінійних рівнянь можна записати у вигляді:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^{n} x_{i1} + b_2 \sum_{i=1}^{n} x_{i2} + \dots + b_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} {x_{i1}}^2 + b_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \dots + b_k \sum_{i=1}^{n} x_{i1} x_{ik} = \sum_{i=1}^{n} x_{i1} y_i \\ b_0 \sum_{i=1}^{n} x_{i2} + b_1 \sum_{i=1}^{n} x_{i1} x_{i2} + b_2 \sum_{i=1}^{n} {x_{i2}}^2 + \dots + b_k \sum_{i=1}^{n} x_{i2} x_{ik} = \sum_{i=1}^{n} x_{i2} y_i \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ b_0 \sum_{i=1}^{n} x_{ik} + b_1 \sum_{i=1}^{n} x_{i1} x_{ik} + b_2 \sum_{i=1}^{n} x_{i2} x_{ik} + \dots + b_k \sum_{i=1}^{n} {x_{ik}}^2 = \sum_{i=1}^{n} x_{ik} y_{i.} \end{cases} \quad (2.4)$$

У матричній формі система лінійних рівнянь (2.4) має вигляд:

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} {x_{i1}}^2 & \sum_{i=1}^{n} x_{i2}x_{i1} & & \sum_{i=1}^{n} x_{ik}x_{i1} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} {x_{i2}}^2 & \cdots & \sum_{i=1}^{n} x_{ik}x_{i2} \\ . & . & . & & . \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1}x_{ik} & \sum_{i=1}^{n} x_{i2}x_{ik} & & \sum_{i=1}^{n} {x_{ik}}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \cdots \\ b_k \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & 1 & 1 & & 1 \\ x_{11} & x_{21} & x_{31} & & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \\ . & . & . & & . \\ x_{1k} & x_{2k} & x_{2k} & & x_{nk} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_n \end{pmatrix}. \qquad (2.4.1)$$

Система рівнянь (2.4) - так звана *нормальна система* ($k$ + 1) рівнянь з ($k$ + 1) невідомими. Розв'язавши її, знайдемо значення параметрів моделі $b_0, b_1, b_2, \dots, b_k$.

Можна показати, що функція $Q(b_0, b_1, b_2, \dots, b_k)$ в отриманій точці набуває мінімального значення. Спосіб, яким записують систему нормальних рівнянь (2.4), називається *звичайним*.

Скорочений матричний запис системи (2.4.1) має вигляд:

$$X^T X b = X^T Y. \qquad (2.4.2)$$

Тут транспонована матриця

$$X^T = \begin{pmatrix} 1 & 1 & 1 & & 1 \\ x_{11} & x_{21} & x_{31} & & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \\ . & . & . & & . \\ x_{1k} & x_{2k} & x_{2k} & & x_{nk} \end{pmatrix},$$

$$X^T X = \begin{pmatrix} n & \sum\limits_{i=1}^{n} x_{i1} & \sum\limits_{i=1}^{n} x_{i2} & & \sum\limits_{i=1}^{n} x_{ik} \\ \sum\limits_{i=1}^{n} x_{i1} & \sum\limits_{i=1}^{n} {x_{i1}}^2 & \sum\limits_{i=1}^{n} x_{i2}x_{i1} & & \sum\limits_{i=1}^{n} x_{ik}x_{i1} \\ \sum\limits_{i=1}^{n} x_{i2} & \sum\limits_{i=1}^{n} x_{i1}x_{i2} & \sum\limits_{i=1}^{n} {x_{i2}}^2 & \cdots & \sum\limits_{i=1}^{n} x_{ik}x_{i2} \\ . & . & . & & . \\ \sum\limits_{i=1}^{n} x_{ik} & \sum\limits_{i=1}^{n} x_{i1}x_{ik} & \sum\limits_{i=1}^{n} x_{i2}x_{ik} & & \sum\limits_{i=1}^{n} {x_{ik}}^2 \end{pmatrix}$$

Оскільки виконується припущення про відсутність мультиколінеарності, то звідси випливає, що детермінант матриці коефіцієнтів при невідомих не дорівнює нулю. Отже, система лінійних алгебраїчних рівнянь (2.4) має єдиний розв'язок $b = (b_0, b_1, b_2, \dots, b_k)^{\mathrm{T}}$, координати якого є оцінками параметрів $\beta_0,\ \beta_1,\ \beta_2, \dots,\ \beta_k$. Завдяки тому, що матриця $X^T X$ має обернену, тобто обернена матриця $(X^T X)^{-1}$ до $X^T X$ існує, то помноживши обидві частини рівності (2.4.2) на матрицю $(X^T X)^{-1}$, отримаємо:

$$b = (X^T X)^{-1} X^T Y. \qquad (2.5)$$

Формула (2.5) є фундаментальним результатом для визначення невідомих параметрів лінійної множинної регресії.

*Інтерпретація емпіричного коефіцієнта $b_i$: зміна величини i-го регресора* (*незалежної змінної $x_i$*) *на одиницю при інших рівних умовах викликає зміну величини $\hat{y}$ на $b_i$ одиниць.*

*Властивості оцінок параметрів $b$* рівнянь регресії:

- незміщеність;
- обґрунтованість;
- ефективність;
- інваріантність.

**Приклад 2.1** ([11]). На підставі статистичних даних про основні соціально-економічні показники (валовий регіональний продукт, величину основних засобів, величину інвестицій в основний капітал, кількість впроваджених прогресивних технологічних процесів у промисловості та кількість зайнятого населення) у регіонах України за 2004 рік (таблиця 2.2) побудувати лінійну багатофакторну кореляційно-регресійну модель, яка описує залежність величини валового регіонального продукту від величини основних засобів, величини інвестицій в основний капітал, кількості впроваджених прогресивних технологічних процесів у промисловості та кількості зайнятого населення.

Таблиця 2.2. Вибіркова сукупність даних для побудови множинної лінійної кореляційно-регресійної моделі

| № з/п | Область | $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|---|---|
| 1 | АР Крим | *9901* | *43758* | *2740* | *90* | *899,7* |
| 2 | Вінницька | *8123* | *25993* | *1155* | *33* | *720,8* |
| 3 | Волинська | *4994* | *14962* | *1064* | *12* | *423,9* |
| 4 | Дніпропетровська | *30040* | *128686* | *5906* | *63* | *1554,7* |
| 5 | Донецька | *45617* | *129043* | *7239* | *115* | *2086,0* |
| 6 | Житомирська | *5947* | *20207* | *854* | *31* | *561,6* |
| 7 | Закарпатська | *5297* | *15182* | *1113* | *21* | *537,8* |
| 8 | Запорізька | *15255* | *59225* | *2745* | *210* | *824,8* |
| 9 | Івано-Франківська | *7311* | *26811* | *1589* | *16* | *513,5* |
| 10 | Київська | *11883* | *42289* | *3547* | *45* | *764,4* |
| 11 | Кіровоградська | *5594* | *19729* | *1357* | *40* | *444,5* |
| 12 | Луганська | *14672* | *51946* | *2941* | *43* | *1019,8* |
| 13 | Львівська | *13992* | *51890* | *3634* | *73* | *1057,0* |
| 14 | Миколаївська | *7934* | *26451* | *1963* | *13* | *534,7* |
| 15 | Одеська | *17029* | *57173* | *5137* | *29* | *1039,4* |
| 16 | Полтавська | *13983* | *52493* | *2887* | *28* | *670,3* |
| 17 | Рівненська | *5599* | *18964* | *1937* | *9* | *437,4* |
| 18 | Сумська | *6275* | *26542* | *1102* | *57* | *534,4* |
| 19 | Тернопільська | *3948* | *13913* | *632* | *13* | *387,7* |
| 20 | Харківська | *20524* | *75594* | *5055* | *108* | *1285,7* |
| 21 | Херсонська | *5200* | *18219* | *886* | *12* | *477,6* |
| 22 | Хмельницька | *6344* | *22399* | *1745* | *12* | *573,6* |
| 23 | Черкаська | *6623* | *23238* | *2408* | *13* | *561,1* |
| 24 | Чернівецька | *3277* | *12893* | *656* | *8* | *353,4* |
| 25 | Чернігівська | *6181* | *27430* | *1140* | *35* | *506,1* |
| 26 | м.Київ | *61357* | *129325* | *13859* | *596* | *1348,9* |
| 27 | м. Севастополь | *2213* | *6714* | *423* | *1* | *176,9* |
| **Разом** | | **345113** | **1141069** | **75714** | **1725** | **20295** |

***Розв'язання***

Результуючою змінною $y$ є величина валового регіонального продукту (млн. грн.), а решта показників соціально-економічного розвитку регіонів, є факторними ознаками, зокрема:

$x_1$ - величина основних засобів у фактичних цінах, млн. грн.;

$x_2$ - інвестиції в основний капітал у фактичних цінах, млн. грн.;

$x_3$ - кількість впроваджених прогресивних технологічних процесів у промисловості, од.;

$x_4$ - кількість зайнятого населення, тис. осіб.

Припустимо, що між величиною $y$ величиною валового регіонального продукту і вказаними показниками соціально-економічного розвитку регіону існує лінійний зв'язок. Лінійна багатофакторна економетрична модель має вигляд:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4.$$

У [11, с. 122 - 123] проведені розрахунки для знаходження невідомих параметрів моделі $b_0, b_1, b_2, b_3, b_4$ . Система нормальних рівнянь для знаходження цих параметрів моделі на підставі заданої вибіркової сукупності має вигляд:

$$\begin{cases} 27b_0 + 1141069\, b_1 + 75714b_2 + 1725b_3 + 20295{,}7b_4 = 345113 \\ 1141069b_0 + 81218618219\, b_1 + 5517582767b_2 + 142756681b_3 + 1227226322b_4 = 26221566544 \\ 75714b_0 + 5517582767b_1 + 418668764b_2 + 12011643b_3 + 81202227{,}3b_4 = 1917195999 \\ 1725b_0 + 142756681\, b_1 + 12011643b_2 + 456883b_3 + 1908171{,}3b_4 = 55057091 \\ 20295{,}7b_0 + 1227226322\, b_1 + 81202227{,}3b_2 + 1908171{,}3b_3 + 19964523{,}3b_4 = 384903829{,}7 \end{cases}$$

Знайдемо розв'язок системи лінійних алгебраїчних рівнянь за допомогою методу оберненої матриці. Отримаємо:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 27 | 1141069 | 75714 | 1725 | 20295,7 | | 345113 |
| | 1141069 | 81218618219 | 5517582767 | 142756681 | 1227226322 | | 26221566544 |
| $X^TX=$ | 75714 | 5517582767 | 418668764 | 12011643 | 81202227,3 | $X^TY=$ | 1917195999 |
| | 1725 | 142756681 | 12011643 | 456883 | 1908171,3 | | 55057091 |
| | 20295,7 | 1227226322 | 81202227,3 | 1908171,3 | 19964523,3 | | 384903829,7 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0,249693 | 6,22975E-06 | 2,43482E-06 | -0,000420115 | -0,00060653 | | -2391,62363 |
| | 6,23E-06 | 5,32861E-10 | -2,14282E-09 | -1,13686E-08 | -2,9286E-08 | | 0,115957329 |
| $(X^TX)^{-1}=$ | 2,43E-06 | -2,14282E-09 | 5,71815E-08 | -6,84758E-07 | -3,7883E-08 | $B=$ | 1,998150945 |
| | -0,00042 | -1,13686E-08 | -6,84758E-07 | 1,49716E-05 | 2,48009E-06 | | 22,98498002 |
| | -0,00061 | -2,92862E-08 | -3,78834E-08 | 2,48009E-06 | 2,38395E-06 | | 4,258755959 |

Оцінки $b_0, b_1, b_2, b_3, b_4$ параметрів $\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4$:

$$b_0 = -2391{,}62363, \quad b_1 = 0{,}115957329, \quad b_2 = 1{,}998150945,$$

$$b_3 = 22{,}98498002, \quad b_4 = 4{,}258755959$$

Вибіркова лінійна множинна кореляційно-регресійна модель має такий вигляд:

$$y = -2391{,}6235 + 0{,}1160\, x_1 + 1{,}9982\, x_2 + 22{,}9850\, x_3 + 4{,}2588\, x_4 .$$

## 2.6. Економетричний зміст параметрів багатофакторної моделі

Коефіцієнт $b_1$ *парної* лінійної регресійної моделі (коефіцієнт парної регресії) відображає *умовний вплив* факторної ознаки на результуючу змінну. Умовність впливу факторної ознаки полягає у тому, що цей вплив є результатом дії ознаки спільно з іншими факторами, кореляційно пов'язаними з нею. Отже, коефіцієнт $b_1$ показує вплив факторної ознаки на результуючу змінну на "фоні" решти факторів, що кореляційно пов'язані із факторною ознакою парної моделі. Цей "фон" може змінюватись зі зміною факторної ознаки. Наявність такого "фону" призводить до виникнення явища, яке при дослідженні кореляційного зв'язку між змінними за допомогою кількох парних моделей називають *подвійним рахунком.*

На підставі даних прикладу 2.1 пояснимо явище подвійного рахунку. Для цього побудуємо чотири парні лінійні кореляційно-регресійні моделі (розрахунки провести самостійно), в яких результуючою змінною є величина валового регіонального продукту, а факторною ознакою - почергово величина основних засобів ($x_1$), величина інвестицій в основний капітал ($x_2$), кількість упроваджених прогресивних технологічних процесів у промисловості ($x_3$) та кількість працівників ($x_4$):

$$(A) \quad \hat{y} = 2122{,}6756 + 0{,}3527\, x_1;$$

$$(B) \quad \hat{y} = 120{,}3538 + 4{,}6010\, x_2;$$

$$(C) \quad \hat{y} = 6698{,}8605 + 95{,}2138\, x_3;$$

$$(D) \quad \hat{y} = -7251{,}6331 + 26{,}6513\, x_4 .$$

*Модель* (*A*) описує залежність валового регіонального продукту від величини основних засобів. Коефіцієнт парної детермінації $R^2$ = 0,8769 показує, що 87,69% загальної зміни величини валового регіонального продукту пояснюється зміною основних засобів.

*Модель* (*B*) відображає залежність валового регіонального продукту від величини інвестицій в основний капітал. Коефіцієнт парної детермінації $R^2$ = 0,9334 для цієї моделі показує, що 93,34% загальної зміни величини валового регіонального продукту пояснюється зміною величини інвестицій в основний капітал.

*Модель* (*C*) описує залежність валового регіонального продукту від кількості впроваджених прогресивних технологічних процесів у промисловості. Коефіцієнт парної детермінації $R^2$ = 0,6716 для цієї моделі показує, що 67,16% загальної зміни величини валового регіонального продукту пояснюється зміною кількості впроваджених прогресивних технологічних процесів у промисловості.

*Модель* (*D*) показує залежність валового регіонального продукту від кількості працівників. Коефіцієнт парної детермінації $R^2$ = 0,7146 для моделі показує, що 71,46% загальної зміни величини валового регіонального продукту пояснюється зміною кількості працівників.

Сума часток впливу усіх незалежних між собою факторних ознак на зміну результуючої змінної повинна дорівнювати 100%. Оскільки величина інвестицій в основний капітал, кількість упроваджених технологій та кількість працівників, в свою чергу, залежать від величини основних засобів, а також корелюють між собою (так, кількість упроваджених технологій залежить від величини інвестицій, кількість працівників - від кількості впроваджених технологій), то маємо явище *подвійного рахунку*: сума коефіцієнтів парної детермінації дорівнює

$$0,8769 + 0,9334 + 0,6716 + 0,7146 = 3,1965 \quad (\text{або } 319,65\%).$$

Коефіцієнт множинної регресії $b_j$ ($j = \overline{1,k}$), відображає *чистий вплив* фактора $x_j$ на результуючу змінну (якщо лінійна множинна кореляційно-регресійна модель містить усі фактори, які впливають на результуючу змінну). У цьому разі сумісний вплив усіх факторних ознак розподіляється між ними відповідно до вагових коефіцієнтів, що оцінюють ступінь дії кожного фактора на результуючу змінну, а "фон" впливу кожного фактора залишається фіксованим. Однак реально кількість усіх факторних ознак, які впливають на результуючу змінну, настільки велика, що врахувати їх у множинній лінійній регресійній моделі фактично неможливо. Тому під час побудови моделі серед усіх факторів вибирають 3-5 найважливіших, вплив інших є неістотним, але знову ж таки створює деякий "фон", якщо вони кореляційно пов'язані з вибраними факторами. У такому разі вважають, що коефіцієнт регресії $b_j$ ($j = \overline{1,k}$) відображає *умовно чистий вплив факторної ознаки* $x_j$ *на результуючу змінну*.

Якщо у багатофакторну кореляційно-регресійну модель включити ще один фактор, то ступінь умовності коефіцієнтів множинної регресії *зменшується* і, якщо факторні ознаки пов'язані між собою кореляційною залежністю, *чисельні значення коефіцієнтів регресії також зменшуються*.

*Вільний член* лінійної множинної кореляційно-регресійної моделі $b_0$, *оцінює середнє значення результуючої змінної при нульових значеннях факторних ознак*, а також указує на те, що для *середніх значень усіх змінних*, які містить модель, *вона є точною*:

$$b_0 = \bar{y} - b_1\overline{x_1} - b_2\overline{x_2} - \cdots - b_k\overline{x_k}.$$

*Економетрична інтерпретація параметрів регресії* (приклад 2.1, с. 108-111). Коефіцієнти множинної регресії показують, на скільки одиниць зміниться результуюча змінна під час зростання тієї чи іншої факторної ознаки на одну одиницю. Так, величина валового регіонального продукту у середньому зросте:

- на 0,116 млн. грн. при зростанні величини основних засобів на 1 млн. грн. ,

- на 1,9982 млн. грн. при зростанні величини інвестицій в основний капітал на 1 млн. грн.,
- на 22,985 млн. грн. при упровадженні одного додаткового прогресивного технологічного процесу;
- на 4,2588 млн. грн. при зростанні кількості працівників на підприємствах регіону на 1 тисячу осіб.

При порівнянні лінійної множинної кореляційно-регресійної моделі з парними моделями ($A$) – ($D$) можна зауважити, що *значення коефіцієнтів множинної регресії є меншими, ніж відповідні коефіцієнти парної регресії*.

Наприклад, для факторної ознаки $x_1$ коефіцієнт множинної регресії становить 0,116, а коефіцієнт парної регресії відповідно 0,3527. Це свідчить про *наявність кореляційного зв'язку між факторними ознаками*.

Вільний член лінійної множинної кореляційно-регресійної моделі показує середнє значення величини валового регіонального продукту при нульових значеннях факторних ознак - величини основних засобів, величини інвестицій в основний капітал, кількості впроваджених прогресивних технологічних процесів у промисловості та кількості працівників у регіонах України. Проте значення $b_0$ = -2391,6235 є *умовним*, оскільки нульові значення факторних ознак не потрапляють в область існування регресійної моделі по кожній із цих змінних.

## 2.7. Порівняння моделей. Значущість моделі

**Приклад 2.2.** Досліджується залежність прибутку підприємства $y$ (млн.грн.) від таких чинників: $x_1$ – витрати на рекламу (сотні тис. грн.); $x_2$ – обсяг інвестицій (сотні тис. грн.). Початкові дані (умовні) для 10 підприємств подані у таблиці 2.3.

1) Побудувати моделі, що виражають:
   - залежність прибутку підприємства від витрат на рекламу та інвестицій:

$$(A) \quad y = b_0+b_1 x_1+b_2 x_2+e;$$

- залежність прибутку підприємства від витрат на рекламу:

$$(B) \quad y = b_0 + b_1 x_1 + e;$$

- залежність прибутку підприємства від інвестицій:

$$(C) \quad y = b_0 + b_2 x_2 +e.$$

2) Для кожної з моделей оцінити дисперсію залишків $\hat{\sigma}_e^2$, середній квадрат модельної помилки (*MSE - model error square*), середню похибку апроксимації.

3) Зробити порівняльну характеристику моделей і визначити найбільш впливовий чинник.

4) Перевірити моделі на значущість.

Таблиця 2.3

| $i$ | $y$ | $x_1$ | $x_2$ |
|---|---|---|---|
| 1 | 7 | 9 | 5 |
| 2 | 10 | 12 | 8 |
| 3 | 10 | 12 | 8 |
| 4 | 7 | 9 | 5 |
| 5 | 6 | 8 | 7 |
| 6 | 6 | 8 | 8 |
| 7 | 5 | 9 | 6 |
| 8 | 5 | 9 | 4 |
| 9 | 4 | 8 | 5 |
| 10 | 9 | 12 | 7 |

*Побудова і дослідження моделі* ($A$)

Для оцінки параметрів регресії $b_0,b_1,b_2$ моделі ($A$) використаємо формулу (2.5). Тут матриця $X$ і вектор $Y$ відповідно мають вигляд:

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 9 & 12 & 12 & 9 & 8 & 8 & 9 & 9 & 8 & 12 \\ 5 & 8 & 8 & 5 & 7 & 8 & 6 & 4 & 5 & 7 \end{pmatrix}^T$$

$$Y = (7, 10, 10, 7, 6, 6, 5, 5, 4, 9)^T$$

Таким чином,

$$B = (X^T X)^{-1} X^T Y =$$

$$= \left[ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 9 & 12 & 12 & 9 & 8 & 8 & 9 & 9 & 8 & 12 \\ 5 & 8 & 8 & 5 & 7 & 8 & 6 & 4 & 5 & 7 \end{pmatrix} \times \begin{pmatrix} 1 & 9 & 5 \\ 1 & 12 & 8 \\ 1 & 12 & 8 \\ 1 & 9 & 5 \\ 1 & 8 & 7 \\ 1 & 8 & 8 \\ 1 & 9 & 6 \\ 1 & 9 & 4 \\ 1 & 8 & 5 \\ 1 & 12 & 7 \end{pmatrix} \right]^{-1} \times$$

$$\times \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 9 & 12 & 12 & 9 & 8 & 8 & 9 & 9 & 8 & 12 \\ 5 & 8 & 8 & 5 & 7 & 8 & 6 & 4 & 5 & 7 \end{pmatrix} \times \begin{pmatrix} 7 \\ 10 \\ 10 \\ 7 \\ 6 \\ 6 \\ 5 \\ 5 \\ 4 \\ 9 \end{pmatrix} = \begin{pmatrix} -4.697 \\ 0.962 \\ 0.374 \end{pmatrix}.$$

Модель залежності прибутку підприємства від витрат на рекламу та інвестицій має вигляд:

$$(A) \qquad \hat{y} = -4{,}697 + 0{,}962 x_1 + 0{,}374 x_2 .$$

Оцінені коефіцієнти $b_1 = 0{,}962$ і $b_2 = 0{,}374$ мають відповідно таке тлумачення:

$b_1$: *при збільшенні* (*з меншенні*) *витрат на рекламу на 100 тис. грн. і при незмінності обсягу інвестицій, прибуток підприємства збільшиться* (*зменшиться*) *на* 0,962 *млн. грн.*;

$b_2$: *при збільшенні* (*зменшенні*) *обсягу інвестицій на 100 тис. грн., і при незмінності витрат на рекламу, прибуток підприємства збільшиться* (*зменшиться*) *на 0,374 млн. грн.*

За допомогою формули (2.2) визначимо $\hat{y}$:

$$\hat{y} = Xb = \begin{pmatrix} 1 & 9 & 5 \\ 1 & 12 & 8 \\ 1 & 12 & 8 \\ 1 & 9 & 5 \\ 1 & 8 & 7 \\ 1 & 8 & 8 \\ 1 & 9 & 6 \\ 1 & 9 & 4 \\ 1 & 8 & 5 \\ 1 & 12 & 7 \end{pmatrix} \times \begin{pmatrix} -4,697 \\ 0,962 \\ 0,374 \end{pmatrix} = \begin{pmatrix} 5,836 \\ 9,846 \\ 9,846 \\ 5,836 \\ 5,622 \\ 5,996 \\ 6,210 \\ 5,462 \\ 4,874 \\ 9,472 \end{pmatrix}.$$

Результати обчислень запишемо у таблицю 2.4:

Таблиця 2.4

| $i$ | $y_i$ | $x_{i1}$ | $x_{i2}$ | $\hat{y}_i$ | $e_i$ | $e_i^2$ | $\lvert e_i / y_i \rvert$ | $(y_i - \bar{y})^2$ | $(\hat{y}_i - \bar{y})^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 9 | 5 | 5,8361 | 1,1639 | 1,3546 | 0,1663 | 0,01 | 1,1318 |
| 2 | 10 | 12 | 8 | 9,8460 | 0,1540 | 0,0237 | 0,0154 | 9,61 | 8,6789 |
| 3 | 10 | 12 | 8 | 9,8460 | 0,1540 | 0,0237 | 0,0154 | 9,61 | 8,6789 |
| 4 | 7 | 9 | 5 | 5,8361 | 1,1639 | 1,3546 | 0,1663 | 0,01 | 1,1318 |
| 5 | 6 | 8 | 7 | 5,6219 | 0,3781 | 0,1429 | 0,063 | 0,81 | 1,6335 |
| 6 | 6 | 8 | 8 | 5,9961 | 0,0039 | 2E-05 | 0,0007 | 0,81 | 0,8171 |
| 7 | 5 | 9 | 6 | 6,2103 | -1,2103 | 1,4647 | 0,2421 | 3,61 | 0,4757 |
| 8 | 5 | 9 | 4 | 5,4620 | -0,4620 | 0,2134 | 0,0924 | 3,61 | 2,0679 |
| 9 | 4 | 8 | 5 | 4,8736 | -0,8736 | 0,7633 | 0,2184 | 8,41 | 4,1061 |
| 10 | 9 | 12 | 7 | 9,4719 | -0,4719 | 0,2227 | 0,0524 | 4,41 | 6,6145 |
| Сума | 69 | 96 | 63 | 69,0000 | 0,0000 | **5,56367** | **1,0323** | **40,90** | **35,3363** |
| Сер.зн | 6,9 | 9,6 | 6,3 | 6,9 | 0 | 0,5564 | 0,1032 | 4,09 | 3,5336 |

Математичне сподівання випадкової величини $e$ дорівнює:

$$M(e) = (\sum_{i=1}^{n} e_i)/n = (\sum_{i=1}^{10} e_i)/10 \approx -5{,}66 \cdot 10^{-18},$$

тобто воно практично дорівнює нулеві, а це означає, що виконується перше припущення багатофакторного кореляційно-регресійного аналізу (п. 2.3).

Дисперсія залишків дорівнює

$$\hat{\sigma}_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1} = \frac{\sum_{i=1}^{10} e_i^2}{10-2-1} \approx \frac{5{,}564}{7} \approx 0{,}795.$$

Середній квадрат модельної помилки *MSE* дорівнює

$$MSE = \frac{1}{n} \sum_{i=1}^{n} e_i^2 = \frac{1}{n} \sum_{i=1}^{10} e_i^2 \approx \frac{5{,}564}{10} \approx 0{,}556.$$

Середня похибка апроксимації або середня абсолютна процентна помилка дорівнює

$$MAPE = \frac{1}{n}\left(\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right|\right) \cdot 100\% \approx 0{,}1032 \cdot 100\% = 10{,}32\%.$$

*Висновок*. Апроксимація моделлю (*A*) вихідних даних є *доброю*, оскільки 10 < *MAPE* < 20.

*Побудова і дослідження моделі* (*B*)

Побудуємо модель (*B*), що виражає залежність прибутку підприємства від витрат на рекламу:

$$(B): \quad \hat{y} = b_0 + b_1 x_1 + e.$$

Модель (*B*) є *однофакторною регресією*. Для оцінки параметрів моделі $\beta_0$ і $\beta_1$ скористаємося засобами Excel. Спочатку треба виділити масив комірок (5×2), потім скористатися командою *Мастер функции* $\boldsymbol{f_x}$, вибрати в категорії *Статистичекая* функцію ЛИНЕЙН.

У першому рядку таблиці розміщені оцінки $b_0$ і $b_1$ параметрів моделі $\beta_0$ і $\beta_1$ :

| | |
|---|---|
| **1,121212** | **-3,86364** |
| 0,191091 | 1,860562 |
| 0,81144 | 0,981843 |
| 34,42672 | 8 |
| 33,18788 | 7,712121 |

Модель ($B$) має такий вигляд:

$$\hat{y} = -\,3{,}864 + 1{,}121\,x_1.$$

Аналогічно, як і при побудові і дослідженні моделі ($A$), отримаємо:

$$M(e) = -6{,}66 \cdot 10^{-18},\ \hat{\sigma}_e^2 \approx 0{,}964,\ \ MSE \approx 0{,}771,\ MAPE \approx 14{,}3\%.$$

*Висновок*. Апроксимація вихідних даних моделлю ($B$) є *доброю* , оскільки 10 < $MAPE$ < 20.

*Побудова і дослідження моделі* ($C$)

Модель ($C$) виражає залежність прибутку підприємства від інвестицій:

$$(C): \qquad \hat{y} = b_0 + b_2 x_2 + e.$$

Результати розрахунку економетричної моделі (С) на основі стандартної програми ЛИНЕЙН:

| | |
|---|---|
| **0,910448** | **1,164179** |
| 0,388251 | 2,507152 |
| 0,407364 | 1,740647 |
| 5,499015 | 8 |
| 16,66119 | 24,23881 |

Модель (С) має такий вигляд:

$$\hat{y} = 1{,}164 + 0{,}910\,x_2.$$

Аналогічно попередньому, отримаємо оцінки:

$$b = (1{,}164;\ 0{,}910)^T,\ M(e) \approx 4{,}98 \cdot 10^{-18},\ \hat{\sigma}_e^2 = 3{,}030,$$

$$MSE \approx 2{,}424,\ \ MAPE \approx 22{,}97\%.$$

*Висновок*. Апроксимація вихідних даних моделлю ($C$) є *задовільною,* оскільки $20 < MAPE < 50$.

Отримані результати запишимо у таблицю 2.5.

Таблиця 2.5

| | Модель | $\hat{\sigma}_e^2$ | $MSE$ | $MAPE$ | $F_{ст}$ | $F_{кр}$ | $M(e)$ |
|---|---|---|---|---|---|---|---|
| $A$ | $\hat{y} = -4.697 + 0.962x_1 + 0.374x_2$ | 0,795 | 0,556 | 10,3% | 321,715 | 4,74 | $-5{,}66 \cdot 10^{-18}$ |
| $B$ | $\hat{y} = -3.864 + 1.121x_1$ | 0,964 | 0,771 | 14,3% | 528,299 | 5,32 | $-6{,}66 \cdot 10^{-18}$ |
| $C$ | $\hat{y} = 1.164 + 0.910x_2$ | 3,030 | 2,424 | 22,9% | 162,64 | 5,32 | $4{,}98 \cdot 10^{-18}$ |

*Висновки*

1. Перша умова ($M(e) = 0$) виконується у всіх трьох моделях.
2. За показниками $\hat{\sigma}_e^2$ і $MSE$ найкращою є модель ($A$), далі модель ($B$) , а на останок, модель ($C$).
3. Оскільки модель ($B$) краща за модель ($C$), то *чинник* $x_1$ (*витрати на рекламу*) *є більш впливовим* на прибуток підприємства, *ніж чинник* $x_2$ (*обсяг інвестицій*).
4. Середня похибка апроксимації по різному характеризує моделі: у моделях ($A$) і ($B$) *добра* апроксимація, а у моделі ($C$) – *задовільна*.
5. Усі три моделі є з*начущими* при рівні значущості $\alpha = 0{,}05$.

## 2.8. Стандартизовані коефіцієнти регресії

Змінні $x_j$ ( $j = \overline{1,k}$ ), вимірюються різними одиницями виміру, тоді величини елементів вектору $b$ не дають можливість визначити важливість кожного з чинників лінійної регресійної моделі. Однак, це можливо зробити, якщо від звичайних коефіцієнтів $b_j$ перейти до *стандартизованих* (або *бета-коефіцієнтів*). Оцінені значення стандартизованих коефіцієнтів обчислюються за такою формулою:

$$b_j^s = b_j \frac{\hat{\sigma}_{x_j}}{\hat{\sigma}_y} \tag{2.6}$$

де $\hat{\sigma}_{x_j}$ і $\hat{\sigma}_y$ - стандартні відхилення:

$$\hat{\sigma}_{x_j} = \sqrt{\frac{\sum_{i=1}^{n}(x_{ji}-\bar{x}_j)^2}{n-1}}, \quad \hat{\sigma}_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y}\ \ )^2}{n-1}} \qquad \left(j=\overline{1,k}\right). \qquad (2.7)$$

Добуток $b_j \cdot \hat{\sigma}_{x_j}$ виражає типовий ефект впливу $j$-го регресора на регресанд. Величина цього ефекту визначається формулою:

$$b_j^s = b_j \frac{\hat{\sigma}_{x_j}}{\hat{\sigma}_y}.$$

*Чим більшим є за абсолютною величиною стандартизований коефіцієнт* $b_j^s$*, тим більш впливовим є* $j$*-ий регресор* (*чинник*).

Якщо дані спостережень стандартизувати, наприклад, $x_{ji}^s = \frac{x_{j_i}-\bar{x}_j}{\hat{\sigma}_{x_j}}$, то замість звичайних коефіцієнтів $b_j$ отримаємо стандартизовані $b_j^s$.

У прикладі 2.2 для моделі (*A*):

$$\hat{y} = -4.697 + 0{,}962\, x_1 + 0{,}374\, x_2$$

розрахуємо стандартизовані коефіцієнти і визначимо чинник, який є найбільш впливовим на прибуток підприємства.

Очевидно, що $b_0^s = 0$. Обчислимо $b_1^s, b_2^s$. Спочатку розрахуємо $\hat{\sigma}_y$, $\hat{\sigma}_{x_1}, \hat{\sigma}_{x_2}$. Маємо:

$$\bar{x}_1 = \frac{\sum_{i=1}^{10} x_{1i}}{10} = 9{,}6, \qquad \bar{x}_2 = \frac{\sum_{i=1}^{10} x_{2i}}{10} = 6{,}3, \quad \bar{y} = \frac{\sum_{i=1}^{10} y_i}{10} = 6{,}9.$$

Використовуючи формули (2.7), отримаємо:

$$\hat{\sigma}_{x_1} = \sqrt{\frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}{10-1}} = 1{,}713; \qquad \hat{\sigma}_{x_2} = \sqrt{\frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}{10-1}} = 1{,}494;$$

$$\hat{\sigma}_y = \sqrt{\frac{\sum_{i=1}^{10}(y_i-\bar{y}\ \ )^2}{10-1}} = 2{,}132.$$

Згідно з формулою (2.1), одержимо:

$$b_1^s = b_1 \frac{\hat{\sigma}_{x_1}}{\hat{\sigma}_y} = 0{,}773, \qquad b_2^s = b_2 \frac{\hat{\sigma}_{x_2}}{\hat{\sigma}_y} = 0{,}262.$$

Стандартизоване регресійне рівняння для моделі ($A$) має вигляд:

$$\hat{y}^s = 0{,}773 + 0{,}262x_2.$$

*Висновок*. Оскільки $b_1^s > b_2^s$, то *чинник* $x_1$ (*витрати на рекламу*) *є більш вагомим на прибуток підприємства, ніж чинник* $x_2$ (*обсяг інвестицій*). Такий самий висновок був отриманий у п. 2.7, коли порівнювали моделі ($B$) і ($C$).

## 2.9. Коефіцієнти еластичності

При інтерпретації регресійних коефіцієнтів приймаються до уваги одиниці виміру регресорів і регресанда. Для визначення міри впливу регресора на регресанд без урахування одиниць їх виміру використовують коефіцієнти еластичності.

Оцінена еластичність регресанда $y$ відносно регресора $x_j$ обчислюється за формулою:

$$E_j = b_j \frac{x_j^*}{y^*} \quad (y^* \neq 0, \quad j = 2,3, \dots k) \tag{2.8}$$

де $x_j^*, y^*$ - значення $j$-го регресора і регресанда, що визначають точку регресійної функції, для якої розраховується коефіцієнт еластичності. Частіше за все використовують значення арифметичних середніх $\bar{x}_j$ , $\bar{y}$ у базовому часовому ряду. При прогнозуванні за допомогою регресійного рівняння доцільно вибирати такі значення $y^*$ і $x_j^*$, які є суттєвими у прогнозному часовому просторі.

*Статистичний сенс коефіцієнтів еластичності* $E_j$ *полягає в тому, що вони показують, на скільки відсотків наближено зросте регресант, якщо* $j$*-й регресор збільшиться на один відсоток*.

У прикладі 2.2 розрахуємо коефіцієнти еластичності $E_1$ і $E_2$ для моделі ($A$) і дамо інтерпретацію отриманих результатів.

Коефіцієнти еластичності розрахуємо при $x_1^* = \bar{x}_1$ , $x_2^* = \bar{x}_2$ , $y^* = \bar{y}$ (тобто у середніх точках). Використовуючи формулу (2.8) і попередні результати, одержимо:

$$E_1 = b_1 \frac{\bar{x}_1}{\bar{y}} = 1{,}339, \quad E_2 = b_2 \frac{\bar{x}_2}{\bar{y}} = 0{,}342.$$

*Висновок*. Якщо витрати на рекламу збільшити на один відсоток, то прибуток підприємства збільшиться наближено на 1,339 %. Якщо обсяг інвестицій збільшити на один відсоток, то прибуток підприємства збільшиться наближено на 0,342 %.

## 2.10. Дисперсійно-коваріаційна матриця

У класичній регресійній моделі вектор залишків $u$ і залежний від нього вектор $Y$ являються випадковими змінними. Оскільки вектор $Y$ у свою чергу входить до функції оцінювання коефіцієнтів регресії за допомогою МНК, то ці коефіцієнти також є випадковими. Для характеристики випадкових змінних $b_j$ $(j = \overline{1,k})$ разом з математичним сподіванням використовується дисперсія ${\sigma_{b_j}}^2$ і коваріація $cov(b_j, b_l) = \sigma_{b_j b_l}$ $(j, l = \overline{1,k},\ j \neq l)$. Значення цих параметрів класичної регресійної моделі утворюють квадратну ($k$ + 1)-го порядку дисперсійно-коваріаційну матрицю $\Sigma_b$. Оцінена методом МНК дисперсійно-коваріаційна матриця має такий вигляд

$$\Sigma_b = \hat{\sigma}_e^2 (X^T X)^{-1} = \begin{pmatrix} \hat{\sigma}_{b_0}^2 & \hat{\sigma}_{b_0 b_1} & \dots & \hat{\sigma}_{b_0 b_k} \\ \hat{\sigma}_{b_1 b_0} & \hat{\sigma}_{b_1}^2 & \dots & \hat{\sigma}_{b_1 b_k} \\ \dots & \dots & \dots & \dots \\ \hat{\sigma}_{b_k b_0} & \hat{\sigma}_{b_k b_1} & \dots & \hat{\sigma}_{b_k}^2 \end{pmatrix}. \qquad (2.9)$$

Елементи цієї матриці використовуються при тестуванні гіпотез за окремими регресійними коефіцієнтами і для розрахунку інтервалу довіри.

На головній діагоналі матриці $\Sigma_b$ елемент $\hat{\sigma}_{b_j}^2$ є МНК-оцінником дисперсії $\hat{\sigma}_{b_j b_j}$ $j$-го оціненого регресійного коефіцієнта $b_j$, а ($j$ ,$l$) -й елемент $\hat{\sigma}_{b_j b_l}$ є МНК- оцінником коваріації між $b_j$ і $b_l$ .

*Чим більші оцінені дисперсії і коваріації, тим ширші інтервали довіри для оцінених коефіцієнтів регресії і залежні від них інтервали прогнозу, що є небажаним.*

Якщо оцінена коваріація додатня ($\hat{\sigma}_{b_j b_l} > 0$ ), то зв′язок між $b_j$ і $b_l$ *прямий*, тобто зі збільшенням (зменшенням) $b_j$ збільшується (зменшується) $b_l$ і навпаки; якщо оцінена коваріація від'ємна ( $\hat{\sigma}_{b_j b_l} < 0$), то зв′язок між між $b_j$ і $b_l$ *обернений*, тобто зі збільшенням (зменшенням) $b_j$ зменшується (збільшується) $b_l$ і навпаки.

Коваріація має недолік: *вона необмежена зверху і знизу*, тобто $-\infty < \hat{\sigma}_{b_j b_l} < +\infty$. Тому до більш конкретного висновку про зв′язок між оціненими коефіцієнтами розраховуються коефіцієнти кореляції:

$$\hat{r}_{b_j b_l} = \frac{\hat{\sigma}_{b_j b_l}}{\hat{\sigma}_{b_j} \cdot \hat{\sigma}_{b_l}}, \tag{2.10}$$

де $\hat{\sigma}_{b_l} = \sqrt{\hat{\sigma}_{b_l}^2}$ - стандартне відхилення.

Коефіцієнти кореляції задовольняють умові: $\left|\hat{r}_{b_j b_l}\right| \leq 1$.

Розрахуємо оцінену дисперсійно-коваріаційну матрицю для моделі ($A$) приклада 2.2. Для цього скористаємося формулою (2.9):

$$\Sigma_b = \hat{\sigma}_e^2 (X^T X)^{-1} = \begin{pmatrix} 3.111 & -0.240 & -0.115 \\ -0.240 & 3.943 \cdot 10^{-2} & -2.197 \cdot 10^{-2} \\ -0.115 & -2.197 \cdot 10^{-2} & 5.178 \cdot 10^{-2} \end{pmatrix}.$$

Одержимо такі значення:

$$\hat{\sigma}_{b_0}^2 = 3.111, \quad \hat{\sigma}_{b_1}^2 = 3.943 \cdot 10^{-2}, \quad \hat{\sigma}_{b_2}^2 = 5.178 \cdot 10^{-2},$$

$$\hat{\sigma}_{b_0 b_1} = \hat{\sigma}_{b_1 b_0} = -2{,}401, \; \hat{\sigma}_{b_0 b_2} = \hat{\sigma}_{b_2 b_0} = -0{,}115, \; \hat{\sigma}_{b_1 b_2} = \hat{\sigma}_{b_2 b_1} = -2{,}197 \cdot 10^{-2}$$

Усі оцінені коваріації від′ємні, тобто зв′язок між коефіцієнтами обернений: зі збільшенням (зменшенням) одного коефіцієнта регресії зменшується (збільшується) інший і навпаки.

Оцінені коефіцієнти кореляції розрахуємо за формулою (2.10):

$$\hat{r}_{b_0b_1} = \frac{\hat{\sigma}_{b_0b_1}}{\hat{\sigma}_{b_0} \cdot \hat{\sigma}_{b_1}} = \frac{\hat{\sigma}_{b_0b_1}}{\sqrt{\hat{\sigma}^2_{b_0}} \cdot \sqrt{\hat{\sigma}^2_{b_1}}} = \frac{-0.241}{\sqrt{3{,}111} \cdot \sqrt{3{,}943 \cdot 10^{-2}}} = -0{,}686,$$

$$\hat{r}_{b_0b_2} = \frac{\hat{\sigma}_{b_0b_2}}{\hat{\sigma}_{b_0} \cdot \hat{\sigma}_{b_2}} = \frac{\hat{\sigma}_{b_0b_2}}{\sqrt{\hat{\sigma}^2_{b_0}} \cdot \sqrt{\hat{\sigma}^2_{b_2}}} = \frac{-0.115}{\sqrt{3{,}111} \cdot \sqrt{5{,}178 \cdot 10^{-2}}} = -0{,}287,$$

$$\hat{r}_{b_1b_2} = \frac{\hat{\sigma}_{b_1b_2}}{\hat{\sigma}_{b_1} \cdot \hat{\sigma}_{b_2}} = \frac{\hat{\sigma}_{b_1b_2}}{\sqrt{\hat{\sigma}^2_{b_1}} \cdot \sqrt{\hat{\sigma}^2_{b_2}}} = \frac{-2.197 \cdot 10^{-2}}{\sqrt{3{,}943 \cdot 10^{-2}} \cdot \sqrt{5{,}178 \cdot 10^{-2}}} = -0{,}486.$$

Зазначимо, що

$$\hat{r}_{b_lb_l} = \frac{\hat{\sigma}_{b_lb_l}}{\hat{\sigma}_{b_l} \cdot \hat{\sigma}_{b_l}} = \frac{\hat{\sigma}^2_{b_l}}{\hat{\sigma}^2_{b_l}} = 1,$$

тобто *діагональними елементами кореляційної матриці є одиниці*.

Таким чином, оцінена кореляційна матриця $\hat{r}$ має такий вигляд:

$$\hat{r} = \begin{pmatrix} 1 & -0.686 & -0.287 \\ -0.686 & 1 & -0.486 \\ -0.287 & -0.486 & 1 \end{pmatrix}.$$

## 2.11. Значущість та інтервали довіри коефіцієнтів регресії

Після побудови регресійної моделі постає питання: *чи суттєво впливає кожний з регресорів на регресанд, тобто, чи відрізняється істинне* (*але невідоме*) *значення коефіцієнта регресії від нуля*?

Для відповіді на це питання скористаємося $t$-тестом Ст'юдента:

1) формулюємо гіпотези $H_0$: $b_j = 0$ і $H_1$: $b_j \neq 0$ $(j = \overline{1,k})$;
2) вибираємо рівень значущості $\alpha$;

3) розраховуємо статистичні значення за формулою $t_{ст} = b_j / \hat{\sigma}_{b_j}$;

4) за таблицею 2 Додатку знаходимо критичні значення статистики Ст'юдента $t_{кр} = t_{\alpha;\, n-k}$;

5) робимо висновок: якщо $|t_{ст}| > t_{кр}$, то гіпотеза $H_0$ відхиляється і навпаки;

6) даємо інтерпретацію результатів.

Оцінки коефіцієнтів регресії називають *точковими*, оскільки на числовій осі *можливе значення цього коефіцієнта являє одна точка*. Наскільки ці точкові оцінки відрізняються від відповідних їм істинних значень? Відповідь можна отримати, якщо побудувати інтервали довіри для коефіцієнтів регресії.

Інтервал довіри (інтервальна оцінка) для величини окремого регресійного коефіцієнта дає цінну інформацію про надійність точкової оцінки.

Таким чином, інтервал довіри для регресійного коефіцієнта $\beta_k$ при рівні довіри $(1-\alpha)$ є інтервалом з випадковими межами і включає істинне значення $j$-го регресійного коефіцієнта з надійністю $(1-\alpha)$ 100%. Він має такі межі:

нижня межа: $b_j - t_{\alpha;\, n-k}\, \hat{\sigma}_{b_j}$,

верхня межа: $b_j + t_{\alpha;\, n-k}\, \hat{\sigma}_{b_j}$ .

Знання інтервалів довіри для значення $j$-го коефіцієнта регресії $\beta_j$ дасть можливість відповісти на питання про значущість цього коефіцієнта: *якщо нуль належить інтервалу довіри, то коефіцієнт незначимо відрізняється від нуля з надійністю* $(1-\alpha)$100%.

Якщо деякі коефіцієнти моделі незначимо відрізняються від нуля, то цю модель не можна використовувати для прогнозу, а тільки для дослідження. І не варто моделлю користуватися, якщо всі коефіцієнти незначимо відрізняються від нуля.

Визначимо значущість і інтервали довіри коефіцієнтів моделі (*A*) прикладу 2.2:

$$\hat{y} = -4{,}697 + 0{,}962x_1 + 0{,}374x_2 .$$

Для значущості коефіцієнтів моделі визначимо критичне значення статистики Ст'юдента

$$t_{кр} = t_{\alpha;\, n-k} = t_{0,05;\, 7} = 2{,}365.$$

Отримаємо

$$b_0: \quad t_{cm} = b_0 / \hat{\sigma}_{b_0} = -4.697 / \sqrt{3.111} \approx -2{,}663,$$

$$b_1: \quad t_{cm} = b_1 / \hat{\sigma}_{b_1} = 0.962 / \sqrt{3.944 \cdot 10^{-2}} \approx 4{,}847,$$

$$b_2: \quad t_{cm} = b_2 / \hat{\sigma}_{b_2} = 0.374 / \sqrt{5.178 \cdot 10^{-2}} = 1{,}644.$$

Числа під знаком арифметичного кореня (дисперсії відповідних коефіцієнтів регресії) і є діагональними елементами дисперсійно-коваріаційної матриці $\Sigma_b$.

*Висновок*. Оскільки для коефіцієнтів $b_0$ і $b_1$ виконується нерівність $|t_{ст}| > t_{кр}$, то це означає, що гіпотеза $H_0$ *відхиляється*, тобто коефіцієнти $b_0$ і $b_1$ *значимо відрізняються від нуля з надійністю* 95%. Оскільки для коефіцієнта $b_2$ виконується нерівність $|t_{ст}| < t_{кр}$, то це означає, що гіпотеза $H_0$ приймається, тобто коефіцієнт $b_2$ *незначимо відрізняється від нуля з надійністю* 95%.

Визначимо інтервали довіри для коефіцієнтів моделі (*A*).

$b_0$: нижня межа: $b_0 - t_{\alpha;n-k}\, \hat{\sigma}_{b_0} = -4{,}697 - 2{,}365 \cdot \sqrt{3.111} = -8.868$,

верхня межа: $b_0 + t_{\alpha;n-k}\, \hat{\sigma}_{b_0} = -4{,}697 + 2{,}365 \cdot \sqrt{3.111} = -0.526$.

$$-8{,}868 \le b_0 \le -0{,}526.$$

*Висновок*. Істинне значення $b_0$ з надійністю 95% належить проміжку [-8,868; -0,526]. Оскільки $0 \notin [-8{,}868; -0{,}526]$, то з надійністю 95% коефіцієнт $b_0$ *значимо відрізняється від нуля*, що співпадає з раніше отриманим результатом.

$b_1$: нижня межа: $b_1 - t_{\alpha;n-k}\, \hat{\sigma}_{b_1} = 0{,}962 - 2{,}365 \cdot \sqrt{3.944 \cdot 10^{-2}} = 0.493$, верхня межа: $b_1 + t_{\alpha;n-k}\ \hat{\sigma}_{b_1} = 0{,}962 + 2{,}365 \cdot \sqrt{3.944 \cdot 10^{-2}} = 1.432$.

$$0{,}493 \le b_1 \le 1{,}432.$$

Істинне значення $b_1 \in [0.493; 1.432]$, *значимо відрізняється від нуля*.

$b_2$: нижня межа: $b_2 - t_{\alpha;n-k}\ \hat{\sigma}_{b_2} = 0{,}374 - 2{,}365 \cdot \sqrt{5.178 \cdot 10^{-2}} = -0.164$

верхня межа: $b_2 + t_{\alpha;n-k}\ \hat{\sigma}_{b_2} = 0{,}374 + 2{,}365 \cdot \sqrt{5.178 \cdot 10^{-2}} = 0.912$.

$$-0.164 \le b_2 \le 0.912.$$

*Висновок*. Оскільки нуль належить проміжку [-0,164; 0,912], то коефіцієнт $b_2$ *незначимо відрізняється від нуля з надійністю* 95%.

## 2.12. Точкові та інтервальні прогнози регресанда

*Прогнозуванням* називається наукове передбачення ймовірнісних шляхів розвитку процесів і явищ для більш-менш віддаленого майбутнього.

Прогноз може бути точковим або інтервальним. Точковий прогноз показника отримаємо підстановкою відомих значень фактора $X_0$ в отримане рівняння економетричної моделі

$$\hat{Y}_0 = X_0 b. \tag{2.11}$$

При цьому дійсне значення показника $Y$ для прогнозного періоду дорівнює $Y_0 = X_0 b + e$, де $e$ – вектор залишків. Помилка прогнозу (різниця між дійсним значенням показника $Y$ для прогнозного періоду і точковим прогнозом показника) обчислюється за формулою

$$v = Y_0 - \hat{Y}_0.$$

Дисперсія прогнозу дорівнює

$$\sigma_{np}^2 = \sigma_e^2 X_0^T (X^T X)^{-1} X_0,$$

звідки

$$\sigma_{np} = \sigma_e \sqrt{X_0^T (X^T X)^{-1} X_0} = \sqrt{X_0^T \Sigma_b X_0}. \tag{2.12}$$

В економічних дослідженнях важливе значення має не точкова оцінка прогнозу регресанда, а інтервал прогнозу.

Інтервальний прогноз отримаємо для математичного сподівання залежної змінної *Y* та для індивідуального значення *Y*.

Довірчий інтервал для математичного сподівання залежної змінної *Y* має такий вигляд:

$$\hat{Y}_0 - t_\alpha \hat{\sigma}_e \sqrt{X_0^T (X^T X)^{-1} X_0} \le M(Y_0) \le \hat{Y}_0 + t_\alpha \hat{\sigma}_e \sqrt{X_0^T (X^T X)^{-1} X_0},$$

де $t_\alpha$ – критичне значення $t$-критерію Ст'юдента при $n - k - 1$ ступенях вільності та рівні значущості α; $\hat{\sigma}_e^2$ – скоригована дисперсія залишків, яка обчислюється за формулою:

$$\hat{\sigma}_e^2 = \frac{\sum_{i=1}^{n}(Y - \hat{Y})^2}{n - k - 1} = \frac{e^T e}{n - k - 1}.$$

Інтервальний прогноз індивідуального значення має такий вигляд

$$\hat{Y}_0 - t_\alpha \hat{\sigma}_e \sqrt{1 + X_0^T (X^T X)^{-1} X_0} \le Y_0 \le \hat{Y}_0 + t_\alpha \hat{\sigma}_e \sqrt{1 + X_0^T (X^T X)^{-1} X_0}.$$

Вибіркове середнє квадратичне відхилення для інтервального прогнозу індивідуального значення обчислюється за формулою:

$$\sigma_{np}^* = \hat{\sigma}_e \sqrt{1 + X_0^T (X^T X)^{-1} X_0} = \sqrt{\hat{\sigma}_e^2 + X_0^T \Sigma_b X_0}. \qquad (2.13)$$

Оскільки дисперсія залишку додатна ($\hat{\sigma}_e{}^2 > 0$), то прогнозний інтервал індивідуального значення регресанда завжди буде більшим, ніж прогнозний інтервал величини математичного сподівання регресанда. Обидва прогнозні інтервали будуть найменшими для $x_0 = \bar{x}$ ($\overline{X}$ - вектор середніх значень $x_i$).

Розрахуємо точковий і інтервальні прогнози для моделі (*А*) при вектор-рядку

$$x_0 = \bar{x} = \left(1; \sum_{i=1}^{n} x_{1i} \Big/ n; \sum_{i=1}^{n} x_{2i}/n\right) = (1;\ \ 9{,}6;\ \ 6{,}3)\ .$$

Використавши формулу (2.11), отримаємо:

$\hat{y}_{np}$ = - 4,697+ 0,962 $x_1$ + 0,374 $x_2$ = -4,697+ 0,962·9,6 + 0,374·6,3 = 6,9.

Прогнозне значення $\hat{y}_{np.}$ = 6,9 (млн.грн.) має такі інтерпретації:

• середній прибуток підприємств, що витрачають на рекламу 9,6 (сотен тис. грн.) і на інвестиції 6,3 (сотен тис. грн.), дорівнює 6,9 (млн. грн.);

• прибуток підприємства, яке витрачає на рекламу 9,6 (сотен тис. грн.) і на інвестиції 6,3 (сотен тис. грн.), дорівнює 6,9 (млн.грн.).

Знайдемо довірчий інтервал для математичного сподівання залежної змінної $Y$. За формулою (2.12) обчислимо вибіркове середнє квадратичне відхилення прогнозу $\sigma_{np}$:

$$\sigma_{np} = \sqrt{(1; 9.6; 6.3)\begin{pmatrix} 3.111 & -2.401 & -0.115 \\ -2.401 & 3.9\cdot 10^{-2} & -2.197\cdot 10^{-2} \\ -0.115 & -2.197\cdot 10^{-2} & 5.178\cdot 10^{-2} \end{pmatrix}\begin{pmatrix} 1 \\ 9.6 \\ 6.3 \end{pmatrix}} = 0.282.$$

Знайдемо за таблицею 2 Додатку $t_{\alpha}$ – критичне значення $t$-критерію Ст'юдента при ступінях вільності $n - (k + 1) = 10 - 3 = 7$ ($n = 10$ – обсяг вібіркової сукупності, $k + 1 = 3$ – кількість параметрів регресії $b_0, b_1, b_2$) та рівні значущості $\alpha = 0{,}05$: $t_{1-\alpha,\, n-k-1} = t_{0{,}95;7} = 2{,}365$.

Нижня межа: $\hat{y}(t) - t_{\alpha;\, n-k}\, \sigma_{np} = 6{,}9 - 2{,}365 \cdot 0{,}282 = 6{,}233$,

верхня межа: $\hat{y}(t) + t_{\alpha;\, n-k}\, \sigma_{np} = 6{,}9 + 2{,}365 \cdot 0{,}282 = 7{,}567$,

$$6{,}233 \le M(Y_0) \le 7{,}567.$$

*Висновок*. Середній прибуток підприємств, що витрачають на рекламу 9,6 (сотен тис. грн.) і на інвестиції 6,3 (сотен тис. грн.), з надійністю 95% буде коливатись у межах [6,233; 7,567] (млн.грн.).

Знайдемо довірчий інтервал для індивідуального значення залежної змінної $Y$. За формулою (2.13) обчислимо вибіркове середнє квадратичне відхилення прогнозу $\sigma^*_{np}$:

$$\sigma^*_{np} = \sqrt{0.795 + (1;9.6;6.3)\begin{pmatrix} 3.111 & -2.401 & -0.115 \\ -2.401 & 3.943\cdot 10^{-2} & -2.197\cdot 10^{-2} \\ -0.115 & -2.197\cdot 10^{-2} & 5.178\cdot 10^{-2} \end{pmatrix}\begin{pmatrix} 1 \\ 9.6 \\ 6.3 \end{pmatrix}} = 0.935$$

Нижня межа: $\hat{y}(t) - t_{\alpha;\, n-k-1}\, \sigma^*_{np} = 6{,}9 - 2{,}365 \cdot 0{,}935 = 4{,}689$,

верхня межа: $\hat{y}(t) + t_{\alpha;\, n-k-1}\, \sigma^*_{np} = 6{,}9 + 2{,}365 \cdot 0{,}935 = 9{,}112$,

$$4{,}689 \le Y_0 \le 9{,}112.$$

*Висновок*. Прибуток підприємства, що витрачає на рекламу 9,6 (сотен тис. грн.) і на інвестиції 6,3 (сотен тис. грн.), з надійністю 95% буде коливатись у межах [4,689; 9,112] (млн.грн.).

## 2.13. Коефіцієнт детермінації

При побудові регресійної моделі намагаються отримати найбільш адекватну модель. *Для оцінки коефіцієнтів регресії та для порівняння декількох моделей використовують суму квадратів залишків*. Однак, *її не можна використовувати як показник адекватності, так як вона необмежена зверху*. Відсутність верхньої межі у сумі квадратів залишків, як недолік, усувається за допомогою *коефіцієнта детермінації* $R^2$, який розраховується за формулою:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{2.14}$$

де $\bar{y} = \left(\sum_{l=1}^{n} y_l\right)/n$ (зауважимо, що $\bar{y} = \bar{\hat{y}}$).

Коефіцієнт детермінації можна визначити також за формулою:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} {e_i}^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{2.15}$$

Коефіцієнт детермінації показує, *яка частина варіації залежної змінної обумовлена варіацією пояснюючої змінної*.

Має місце така рівність сум квадратів:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(\hat{y}_i - y_i)^2 . \tag{2.16}$$

З рівності (2.16) маємо:

$$0 \le \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \le \sum_{i=1}^{n}(y_i - \bar{y})^2,$$

а це означає, що $0 \le R^2 \le 1$.

Чим ближче значення коефіцієнта детермінації до одиниці, тим краще регресійна модель апроксимує емпіричні дані. Рівність $\sum_{i=1}^{n} {e_i}^2 = 0$ означає, що $R^2 = 1$. У цьому випадку $y(t) = \hat{y}(t)$ , тобто всі емпіричні дані розташовані на регресійній гіперплощині. Випадок $R^2 = 0$ можливий при $\hat{y} = \bar{y}$. З двох регресійних рівнянь, які відрізняються лише регресорами (кількість їх однакова), *кращим вважається рівняння з найбільшим коефіцієнтом детермінації*.

Коефіцієнт детермінації можна використати для визначення значущості регресійного рівняння.

Для цього розраховується фактичне статистичне значення

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

і якщо $F_{факт} > F_{кр} = F(\alpha; k-1; n-k)$, то *рівняння регресії достатньо якісно з надійністю* $(1-\alpha)100\%$ *моделює динаміку зміни залежної змінної Y*.

## 2.14. Скоригований за Тейлом коефіцієнт детермінації

*Недоліком коефіцієнта детермінації є те, що він не зменшується при збільшенні пояснюючих змінних*, тобто:

$$R^2(k+1) \geq R^2 \cdot k,$$

де $k$ - кількість регресорів у моделі. Таким чином, *якщо до моделі додати не дуже важливий чинник, то коефіцієнт детермінації не зменшується, а тільки може збільшитись. В той самий час не обов'язково модель стане кращою*. Тому краще використовувати скоригований за Тейлом коефіцієнт детермінації, який *враховує кількість чинників, що входять до моделі, тобто зменшується вплив кількості чинників на значення коефіцієнта детермінації.*

Скоригований за Тейлом коефіцієнт детермінації обчислюється за формулою:

$$R_T^2 = 1 - \frac{\sum_{l=1}^{n} e_i/(n-k)}{\sum_{l=1}^{n}(y_i - \bar{y})^2/(n-1)} \quad . \qquad (2.17)$$

На відміну від коефіцієнта детермінації $R^2$ скоригований за Тейлом коефіцієнт детермінації $R_T^2$ коригується з урахуванням ступенів вільності, суми квадратів похибок і загальної суми квадратів. Для отримання формули (2.17), необхідно у формулі (2.15) чисельник і знаменник дробу поділити на відповідні ступіні вільності.

Має місце такий зв′язок скоригованого за Тейлом коефіцієнта детермінації $R_T^2$ з коефіцієнтом детермінації $R^2$:

$$\bar{R}_T^2 = (1 - R^2)\frac{n-1}{n-k} \qquad (2.18)$$

Відзначимо деякі *властивості скоригованого за Тейлом коефіцієнта детермінації*:

а) $R_T^2 \leq R^2$ при $k > 1$ ;

б) якщо кількість чинників зростає, то скоригований за Тейлом коефіцієнт детермінації $R_T^2$ зростає повільніше, ніж коефіцієнт $R^2$, тобто зменшується вплив кількості чинників на величину коефіцієнта детермінації;

в) $R_T^2$ може бути і від′ємним, тоді як $R^2$ завжди невід′ємний;

г) якщо до регресії додається новий чинник і його $|t_{cm}|>1$, то $R_T^2$ збільшується ( і тільки у цьому випадку).

Отже, збільшення $R_T^2$ при введені нового чинника не обов′язково означає що його коефіцієнт значимо відрізняється від нуля. Тому *не потрібно вважати, що збільшення $R_T^2$ означає покращення специфікації моделі*. Це і є однією з причин, що заважає широкому використанню скоригованого за Тейлом коефіцієнта детермінації. Та й взагалі увага до коефіцієнта детермінації $R^2$ зменшилась, тому що у погано специфікованої моделі може бути високим коефіцієнт детермінації (наприклад, при наявності мультиколінеарності). Сучасні дослідники розглядають коефіцієнт детермінації $R^2$ як *один з діагностичних показників, що розраховуються при побудові моделі*.

## 2.15. Частковий коефіцієнт детермінації

Дослідимо питання: *як зміниться коефіцієнт детермінації $R^2$, якщо з моделі виключити $j$-ий регресор*? Для цього необхідно розрахувати коефіцієнти детермінації для основної моделі і моделі зі зменшеною кількістю регресорів і порівняти їх:

$$\Delta R_j^2 = R^2(x_{1,} \dots, x_k) - R^2(x_{1,} \dots, x_{j-1}, x_{j+1}, x_k).$$

Але можна піти іншим шляхом: без значних обчислювальних витрат розрахувати коефіцієнт $\Delta R_j^2$, який називається *частковим коефіцієнтом детермінації*:

$$\Delta R_j^2 = \frac{1-R^2}{n-k}\left(\frac{b_j}{\hat{\sigma}_{b_j}}\right)^2 \qquad (2.19)$$

*Частковий коефіцієнт детермінації* $\Delta R_j^2$ *показує, на яку величину зменшиться коефіцієнт детермінації* $R^2$*, якщо j-ий регресор буде виключено з моделі*. Враховуючи суть коефіцієнта детермінації, можна зробити такий висновок: *чим більшим є* $\Delta R_j^2$*, тим більш важливим є у моделі j-ий регресор*.

У прикладі 2.2 для моделей (*A*), (*B*) і (*C*) розрахуємо коефіцієнти детермінації $R^2$ і $R_T^2$, а для моделі (*A*) - часткові коефіцієнти детермінації $\Delta R_1^2$, $\Delta R_2^2$, а також значущість моделей.

Для моделі (A) обчислимо коефіцієнт детермінації, скориставшись сумами двох останніх стовпців таблиці 2.4:

$$R_A^2 = \frac{\sum_{i=1}^{10}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{10}(y_i - \bar{y})^2} = \frac{35{,}3363}{40{,}9} \approx 0{,}864.$$

Обчислимо скоригований за Тейлом коефіцієнт детермінації:

$$R_{T\,A}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} = 0{,}825$$

і часткові коефіцієнти детермінації $\Delta R_1^2$, $\Delta R_2^2$:

$$\Delta R_1^2 = \frac{1-R^2}{n-k}\left(\frac{b_1}{\hat{\sigma}_{b_1}}\right)^2 = \frac{1-0{,}864}{10-3}\left(\frac{0{,}962}{\sqrt{3{,}948 \cdot 10^{-2}}}\right)^2 = 0{,}457,$$

$$\Delta R_2^2 = \frac{1-R^2}{n-k}\left(\frac{b_2}{\hat{\sigma}_{b_2}}\right)^2 = \frac{1-0{,}864}{10-3}\left(\frac{0{,}374}{\sqrt{5{,}178 \cdot 10^{-2}}}\right)^2 = 0{,}053.$$

*Висновок*. Якщо з моделі ($A$) вивести чинник $x_1$, то коефіцієнт детермінації $R^2$ моделі зменшиться на 0,457, а якщо вивести чинник $x_2$, то $R^2$ зменшиться на 0,053. Оскільки $\Delta R_1^2 > \Delta R_2^2$, то *чинник* $x_1$ (*витрати на рекламу*), *є більш впливовим в моделі, ніж чинник* $x_2$ (*інвестиції*).

Для визначення значущості моделі ($A$) розрахуємо фактичне статистичне значення критерію Фішера:

$$F_{\text{факт}} = \frac{R^2}{1-R^2}\frac{n-k}{k-1} = \frac{0.864}{1-0.864}\frac{10-3}{3-1} = 22.24.$$

За таблицею 1 Додатку

$$F_{\text{кр}} = F(\alpha; k-1; n-k) = F(0{,}05; 2; 7) = 4{,}74.$$

*Висновок*. Оскільки $F_{факт} > F_{кр}$, то модель ($A$) достатньо якісно з надійністю 95% моделює динаміку зміни залежної змінної $\hat{y}$**.**

Аналогічно розрахуємо коефіцієнти детермінації $R^2$ і $\bar{R}_n^2$ для моделей ($B$) і ($C$). Отримаємо відповідно такі значення:

$R_B^2$=0,811 , $R_{T\,B}^2 = 0{,}788$, $F_{\text{ст}.B} = 34{,}32$, $F_{\text{кр}B} = 5{,}32$;
$R_C^2 = 0{,}407$, $R_{T\,C}^2 = 0{,}333$, $F_{\text{ст}\,C} = 5{,}49$, $F_{\text{кр}\,C} = 5{,}32$.

Отримані результати запишимо у таблицю 2.6:

Таблиця 2.6

| | Модель | $R^2$ | $\bar{R}_{\text{T}}^2$ | $F_{ст}$ | $F_{кр}$ |
|---|---|---|---|---|---|
| $A$ | $\hat{y} = -4.697 + 0.962x_1 + 0.374x_2$ | 0,864 | 0,825 | 22,24 | 4,74 |
| $B$ | $\hat{y} = -3.864 + 1.121x_1$ | 0,811 | 0,788 | 34,32 | 5,32 |
| $C$ | $\hat{y} = 1.164 + 0.910x_2$ | 0,407 | 0,333 | 5,49 | 5,32 |

*Висновки*

1) Значення $F_{\text{ст}}$ і $F_{\text{кр}}$ вказують на те, що *всі моделі значимо відрізняються від нуля з надійністю* 95%.
2) Значення коефіцієнтів $R^2$ і $R_T^2$ вказують на те, що найкращою є модель ($A$), потім - модель ($B$) і, наостанок, - модель ($C$) (що отримували і раніше).

3) Чинник $x_1$ (витрати на рекламу) є більш впливовим на прибуток підприємства, ніж чинник $x_2$ (інвестиції).

4) Коефіцієнти детермінації $R^2$ і $R_T^2$ для моделей $(A)$ і $(B)$ є досить великими.

**Приклад 2.3.** Інвестиційна компанія «Інвест» розглядає інвестиційний проект, пов'язаний з купівлею пакета акцій підприємства «Будівельник». Менеджери компанії зібрали дані про аналогічні угоди (таблиця 2.7) і для оцінювання вартості пакета акцій $y$ (млн.. грн.) взяли такі параметри:

вартість основних фондів підприємства $x_1$, млн. грн.;

обсяг річного товарообороту підприємства $x_2$, млн. грн.;

кредиторська заборгованість підприємства $x_3$, млн. грн.;

дебіторська заборгованість підприємства $x_4$, млн. грн.;

заборгованість підприємства із заробітної плати $x_5$, млн. грн.;

вартість виготовленої нереалізованої продукції та вихідної сировини $x_6$, млн. грн.

Таблиця 2.7

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $Y$ |
|---|---|---|---|---|---|---|
| 122,30 | 665,70 | 88,50 | 77,70 | 123,70 | 77,90 | 654,28 |
| 103,96 | 515,92 | 67,26 | 56,72 | 84,73 | 52,19 | 507,63 |
| 116,43 | 608,78 | 80,17 | 68,97 | 106,09 | 65,97 | 598,04 |
| 101,29 | 490,07 | 63,50 | 52,83 | 77,13 | 47,10 | 482,27 |
| 111,42 | 563,58 | 73,66 | 62,34 | 93,32 | 57,47 | 553,69 |
| 100,28 | 479,04 | 61,87 | 51,12 | 73,72 | 44,82 | 471,46 |
| 114,32 | 579,64 | 75,73 | 64,00 | 95,40 | 58,63 | 569,10 |
| 100,60 | 475,31 | 61,19 | 50,18 | 71,36 | 43,15 | 467,69 |
| 114,69 | 575,12 | 74,90 | 62,82 | 92,34 | 56,44 | 564,45 |
| 105,51 | 506,11 | 65,31 | 53,78 | 76,83 | 46,51 | 497,47 |
| 113,95 | 566,84 | 73,67 | 61,52 | 89,73 | 54,69 | 556,33 |
| 121,93 | 626,36 | 81,92 | 69,27 | 102,92 | 63,12 | 614,05 |

Для підприємства «Будівельник» складові вартості мають такі значення:

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|
| 131,683 | 701,521 | 92,408 | 79,248 | 120,214 | 74,225 |

Необхідно визначити можливість купівлі пакета акцій підприємства «Будівельник». Вартість пакета становить 800 млн. грн.

***Розв'язання.***

Складаємо таблицю початкових даних і відкриваємо вікно «*Анализ данных*» (меню *Сервис* → вікно *Анализ даннях* (для Excel 1997-2003) або *Данные* → *Анализ данных* (якщо треба, то встановити *Пакет анализа*) для пізніших версій) (Рис. 2.2), де беремо розділ «*Регрессия*», в якому зазначимо вхідний інтервал *y* та вихідний інтервал *x* (Рис.2.3).

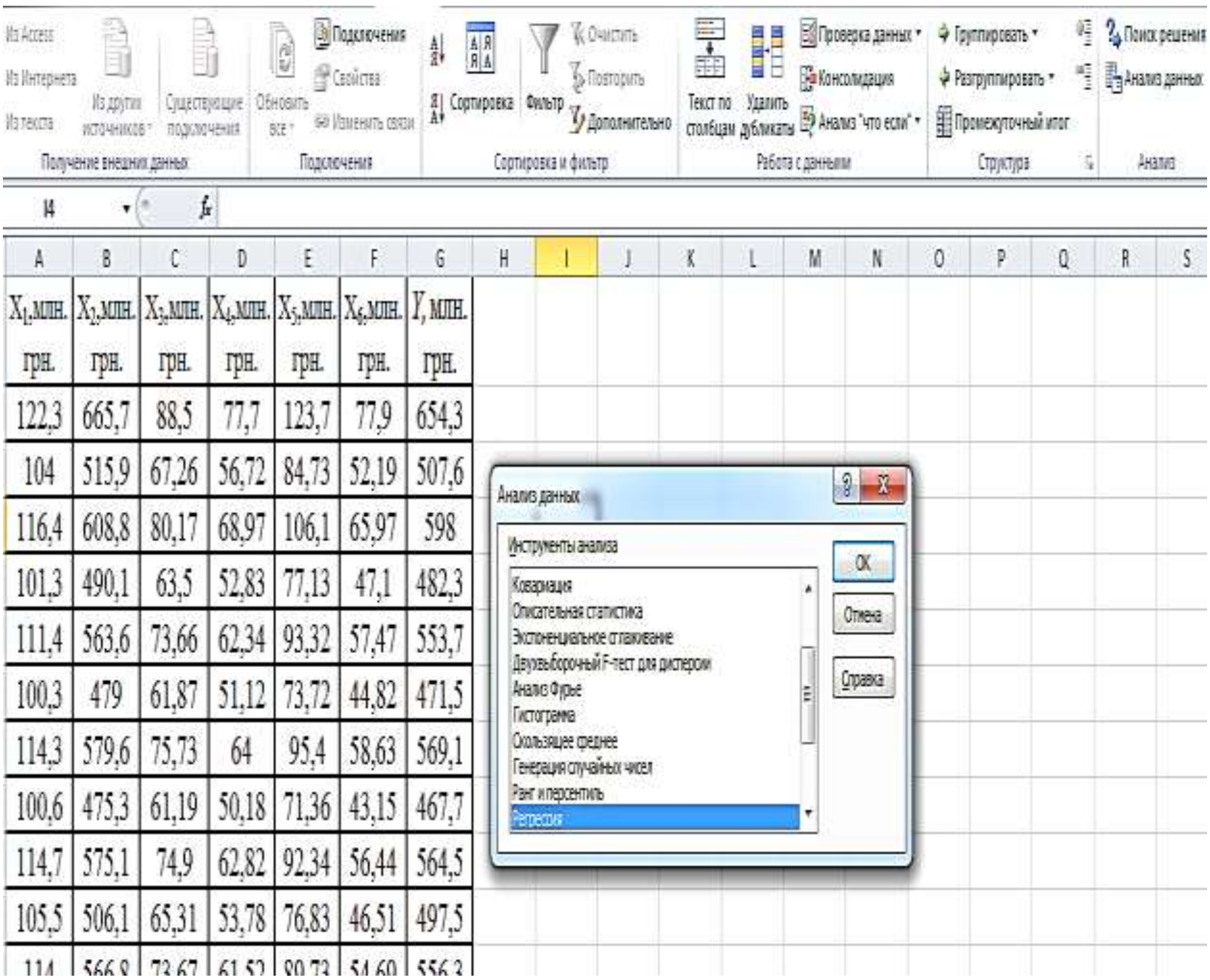


Рис. 2.2. Вікно «Анализ данных»

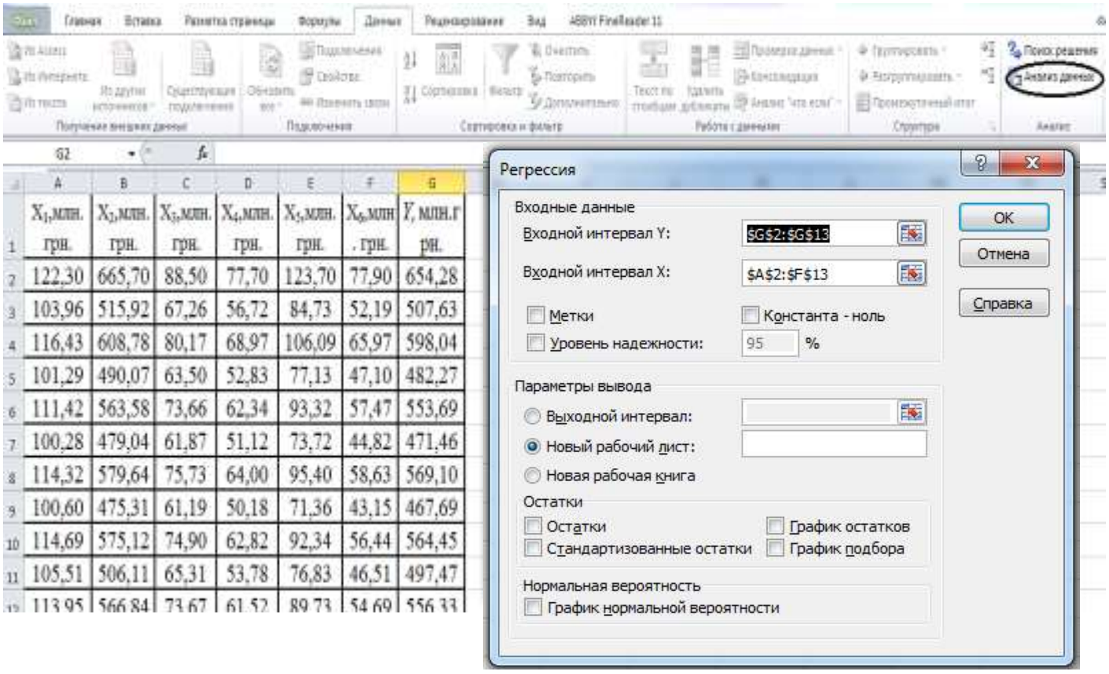



Рис. 2.3. Вікно «Регрессия» для встановлення вхідних інтервалів функції *y* та вихідних інтервалів параметра *x*

Після цього в окремому вікні (Рис.2.4) буде подано результати розрахунків

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ВЫВОД ИТОГОВ | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | *Регрессионная статистика* | | | | | | | | |
| 4 | Множественный R | 0,999999998 | | | | | | | |
| 5 | R-квадрат | 0,999999995 | | | | | | | |
| 6 | Нормированный R-квадрат | 0,999999989 | | | | | | | |
| 7 | Стандартная ошибка | 0,006268127 | | | | | | | |
| 8 | Наблюдения | 12 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | Дисперсионный анализ | | | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *ачимость F* | | | |
| 12 | Регрессия | 6 | 39660,8539 | 6610,142 | 1,68E+08 | 1,36E-20 | | | |
| 13 | Остаток | 5 | 0,000196447 | 3,93E-05 | | | | | |
| 14 | Итого | 11 | 39660,8541 | | | | | | |
| 15 | | | | | | | | | |
| 16 | | *Коэффициенты* | *ндартная оши* | *татисти* | *Р-Значени* | *ижние 95%* | *ерхние 95%* | *жние 95,0* | *рхние 95,0%* |
| 17 | Y-пересечение | 0,503359417 | 1,913847879 | 0,263009 | 0,803037 | -4,41634 | 5,423062 | -4,41634 | 5,423062 |
| 18 | Переменная X 1 | 0,368042955 | 0,208206423 | 1,767683 | 0,137357 | -0,16717 | 0,903255 | -0,16717 | 0,903255 |
| 19 | Переменная X 2 | 0,969849224 | 0,188590247 | 5,142627 | 0,003637 | 0,485063 | 1,454636 | 0,485063 | 1,454636 |
| 20 | Переменная X 3 | -0,283404209 | 1,015900409 | -0,27897 | 0,791439 | -2,89486 | 2,328051 | -2,89486 | 2,328051 |
| 21 | Переменная X 4 | -1,201803524 | 1,000257666 | -1,20149 | 0,283361 | -3,77305 | 1,369441 | -3,77305 | 1,369441 |
| 22 | Переменная X 5 | 0,575614578 | 0,481856766 | 1,194576 | 0,285824 | -0,66304 | 1,814267 | -0,66304 | 1,814267 |
| 23 | Переменная X 6 | 0,133424333 | 0,434535178 | 0,307051 | 0,771181 | -0,98358 | 1,250433 | -0,98358 | 1,250433 |

Рис 2.4. Вікно з результатам розрахунків

Розраховані значення коефіцієнтів рівняння регресії дають змогу записати рівняння у такому вигляді:

$$y = 0{,}5 + 0{,}37x_1 + 0{,}97\,x_2 - 0{,}28\,x_3 - 1{,}20\,x_4 + 0{,}58\,x_5 + 0{,}13\,x_6.$$

Підставимо значення параметрів для підприємства «Будівельник» у рівняння і розраховуємо вартість пакета акцій: $y$ = 687,009 (млн. грн.).

*Висновок.* Вартість запропонованого для продажу пакета акцій (800 млн. грн.) *завищена*, а тому за цією вартістю купувати пакет акцій *недоцільно*.

## 2.16. Завдання для самостійної роботи

### 2.16.1. Тести

*Вибрати правильну відповідь на запитання*

1. Під частковою кореляцією розуміють:
   а) зележність між результативною і однією з факторних ознак при фіксованих значеннях інших факторних ознак;
   б) залежність результативної ознаки і двох або більше факторних ознак, що включені у дослідження;
   в) зв'язок між двома ознаками (результативною і факторною, або двома факторними);
   г) залежність між якісними ознаками.
2. Скоригований за Тейлом коефіцієнт детермінації для множинної регресії:
   а) більше, ніж звичайний коефіцієнт детермінації;
   б) менше або дорівнює звичайному коефіцієнту детермінації;
   в) менше, ніж звичайний коефіцієнт детермінації.
3. Зі збільшенням числа пояснюючих факторів скоригований за Тейлом коефіцієнт детермінації:
   а) не змінюється;
   б) збільшується швидше, ніж значення звичайного коефіцієнту детермінації;

в) зменшується швидше, ніж значення звичайного коефіцієнту детермінації.

4. Коефіцієнт еластичності показує:

а) на скілько одиниць змінюється фактор при зміні результату на одну одиницю;

б) на скільки одиниць зміниться результат при зміні фактора на одну одиницю;

в) наближено на скільки процентів зміниться результат при зміні фактора на 1%;

г) наближено на скільки процентів зміниться фактор при зміні результата на 1%;

д) в скільки разів зміниться результат при зміні фактора на одну одиницю.

5. Критичні значення критерія Ст'юдента визначаються за:

а) рівнєм значущості та числом ступенів вільності;

б) рівнем значущості;

в) числом ступенів вільності;

г) двома ступенями вільності.

6. У множинній регресії:

а) більш, ніж одна залежна змінна і тільки одна незалежна змінна;

б) більш, ніж одна незалежна змінна і тільки одна залежна змінна;

в) більш, ніж одна залежна змінна і більш, ніж одна незалежна змінна;

г) тільки одна залежна змінна і тільки одна незалежна змінна;

д) більш, ніж дві залежні змінні і більш, ніж одна незалежна змінна.

7. Однією з проблем, що може виникнути у множинній регресії і ніколи не буває в простій регресії, є:

а) кореляція між величинами помилок;

б) нерівна дисперсія помилок;
в) кореляція між помилками та незалежними змінними;
г) кореляція між незалежними змінними;
д) помилка без нульової середньої.

8. У множинній регресії кожен параметр показує:
   а) загальний вплив усіх незалежних змінних на залежну змінну;
   б) вплив незалежної змінної на залежну за умови, що всі інші незалежні залишаються незмінними;
   в) де площина регресії перетинає вісь ординат*;*
   г) як частковий, так і загальний вплив незалежних змінних;
   д) точку, що є значенням перетину.

9. Мультиколінеарність виникає тоді, коли:
   а) помилка не має нульового середнього значення;
   б) помилка залежить від незалежної змінної;
   в) дві помилки корелюють між собою;
   г) незалежні змінні корелюють між собою;
   д) дисперсія помилок не є постійною.

10. Для перевірки значущість окремого параметра використовують:
    а) $F$-тест Фішера;
    б) $t$-тест Ст'юдента;
    в) $\chi^2$–тест Пірсона;
    г) біноміальний розподіл;
    д) експоненціальний розподіл.

11. Для перевірки значущості одночасно всіх параметрів регресії використовується:
    а) $F$-тест Фішера;
    б) $t$-тест Ст'юдента;
    в) $\chi^2$–тест Пірсона;
    г) біноміальний розподіл;

д) експоненціальний розподіл.

12. Ступені вільності для $t$-статистики Ст'юдента для перевірки значущості параметрів регресії, що складається з 35 спостережень та 3 незалежних змінних, такі:

а) 35; б) 3; в) 32; г) 33; д) 31.

13. За інших рівних умов, якщо збільшується кількість незалежних змінних у множинній регресії, то коефіцієнт детермінації:

а) збільшується;

б) зменшується;

в) може або збільшитись, або зменшитись;

г) немає ніякого ефекту;

д) імовірність мультиколінеарності зменшується.

14. Скоригований за Тейлом коефіцієнт детермінації по відношенню до коефіцієнта детермінації:

а) завжди більший;

б) завжди менший;

в) непорівнянний;

г) значно більший.

### 2.16.2. Контрольні запитання

1. Які припущення висувають до багатофакторної лінійної кореляційно-регресійній моделі?
2. У чому суть явища мультиколінеарності?
3. Як проводять економетричний аналіз факторних ознак?
4. Назвіть етапи побудови множинної лінійної кореляційно-регресійній моделі.
5. Які вимоги задовольняють оцінки параметрів множинної лінійно-кореляційної моделі, що отримані МНК при виконанні основних припущень?
6. Які особливості має лінійна множинна регресійна модель,

записана у стандартизованому вигляді?

7. Що таке «явище подвійного рахунку»?
8. Що таке стандартна похибка множинно лінійної кореляційно-регресійної моделі?
9. Як обчислюється середній квадрат модельних помилок ($MSE$)?
10. Запишіть формулу середньої похибки апроксимації (середньої абсолютної процентної помилки, $MAPE$).
11. Вкажіть порогові значення середньої похибки апроксимації.
12. Як отримують стандартизовані коефіцієнти регресії?
13. Як обчислюється коефіцієнт еластичності?
14. У чому полягає статистичний сенс коефіцієнтів еластичності?
15. Запишіть дисперсійно-коваріаційну матрицю.
16. У чому полягають точкові оцінки коефіцієнтів регресії?
17. Що таке інтервал довіри (інтервальна оцінка для величини окремого регресійного коефіцієнта)?
18. Як здійснюється точковий прогноз регресанда?
19. Запишіть інтервал прогнозу виличини математичного сподівання регресанда.
20. Запишіть інтервал прогнозу індивідуального значення регресанда.
21. За якою формулою обчислюється коефіцієнт детермінації?
22. У чому полягає економетричний сенс коефіцієнта детермінації?
23. У чому полягає недоліки коефіцієнта детермінації?
24. Запишіть формулу обчислення скоригованого за Тейлом коефіцієнта детермінації.
25. У чому перевага скоригованого за Тейлом коефіцієнта детермінації?
26. Сформулюйте властивості скоригованого за Тейлом коефіцієнта детермінації.

27. Запишіть формулу часткового коефіцієна детермінації?
28. У чому полягає економетричний зміст часткового коефіцієнта детермінації?
29. Чи може бути *t*-статистика Ст'юдента від'ємною?
30. Запишіть систему нормальних рівнянь для визначення невідомих параметрів множинної лінійної кореляційно-регресійної моделі.

### 2.16.3. Завдання для контрольної роботи

***Завдання 1*** .

Залежність між продуктивністю праці $Y$ та фондозабезпеченістю (тис.грн.) $X_1$, стажем роботи в роках $X_2$, плинністю кадрів (в долях) $X_3$, рівнем оплати праці (тис.грн./рік) $X_4$ наведена в таблиці 2.7 (тут $N$ – номер варіанту студента за списком групи) .

Таблиця 2.7

| $i$ | $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|---|
| 1 | 14,85 | $40 + 0{,}2\cdot N$ | 30 | 0,15 | 5,0 |
| 2 | 11,94 | $38 + 0{,}2\cdot N$ | 19 | 0,02 | 3,1 |
| 3 | 8,03 | $29 + 0{,}2\cdot N$ | 8 | 0,14 | 4,7 |
| 4 | 7,11 | $18 + 0{,}2\cdot N$ | 18 | 0,11 | 2,5 |
| 5 | 9,5 | $35 + 0{,}2\cdot N$ | 9 | 0,12 | 4,9 |
| 6 | 9,4 | $27 + 0{,}2\cdot N$ | 23 | 0,10 | 2,6 |
| 7 | 11,6 | $37 + 0{,}2\cdot N$ | 15 | 0,13 | 4,6 |
| 8 | 8,14 | $17 + 0{,}2\cdot N$ | 17 | 0,09 | 3,4 |
| 9 | 15,62 | $39 + 0{,}2\cdot N$ | 28 | 0,07 | 4,8 |
| 10 | 11,12 | $36 + 0{,}2\cdot N$ | 16 | 0,12 | 4,9 |
| 11 | 7,34 | $25 + 0{,}2\cdot N$ | 7 | 0,08 | 3,2 |
| 12 | 10,58 | $34 + 0{,}2\cdot N$ | 15 | 0,11 | 4,7 |
| 13 | 7,37 | $13 + 0{,}2\cdot N$ | 25 | 0,15 | 2,7 |
| 14 | 10,63 | $35 + 0{,}2\cdot N$ | 8 | 0,13 | 5,0 |
| 15 | 10,15 | $28 + 0{,}2\cdot N$ | 24 | 0,17 | 2,9 |

1) Записати вибіркову лінійну множинну регресійну модель. Дати економетричну інтерпретацію параметрів регресії.
2) Побудувати та дослідити модель, що виражає залежність продуктивності праці від стажу роботи і рівнем оплати праці.
3) Обчислити стандартизовані коефіцієнти регресії для моделі залежності продуктивності праці від стажу роботи і рівнем оплати праці.
4) Обчислити коефіцієнти еластичності для моделі залежності продуктивності праці від стажу роботи і рівнем оплати праці.
5) Побудувати точкові та інтервальні прогнози для моделі залежності продуктивності праці від стажу роботи і рівнем оплати праці.
6) Обчислити коефіцієнт множинної детермінації для моделі залежності продуктивності праці від стажу роботи і рівнем оплати праці.

***Завдання 2.***

Досліджується залежність розміру заробітної плати $Y$ фахівців деякого підприємства від віку, освіти (професійно-технічної на базі повної загальної середньої, вищої і післядипломної) та стажу праці за даним фахом. Випадково вибрани 25 фахівців.Результати обстеження наведені в таблиці 2.8 (тут $N$ – номер варіанту студента за списком групи), де $Y$ - зарабітна плата (в умовних одиницях), $t_1$ - кількість років навчання, $t_2$ – вік (роки) і $t_3$ - стаж праці за даним фахом (роки).

1) Записати вибіркову лінійну множинну регресійну модель. Дати економетричну інтерпретацію параметрів регресії.
2) Побудувати моделі, що виражають залежність заробітної плати від кожного з трьох вищезазначених чинників.
3) Зробити порівняльну характеристику кожної з моделей і визначити найбільш впливовий чинник.
4) Перевірити побудовані моделі на значущість.

Таблиця 2.8

| $i$ | $Y$ | $t_1$ | $t_2$ | $t_3$ |
|---|---|---|---|---|
| 1 | 126 + N | 4 | 27+0,5N | 9 |
| 2 | 160+ N | 6 | 30+0,5N | 12 |
| 3 | 104+ N | 4 | 42+0,5N | 10 |
| 4 | 162+ N | 7 | 32+0,5N | 13 |
| 5 | 212+ N | 5 | 44+0,5N | 16 |
| 6 | 130+ N | 2 | 25+0,5N | 6 |
| 7 | 102+ N | 0 | 21+0,5N | 3 |
| 8 | 245+ N | 8 | 53+0,5N | 30 |
| 9 | 183+ N | 6 | 47+0,5N | 26 |
| 10 | 149+ N | 3 | 34+0,5N | 12 |
| 11 | 116+ N | 2 | 27+0,5N | 10 |
| 12 | 158+ N | 5 | 29+0,5N | 10 |
| 13 | 178+ N | 7 | 49+0,5N | 25 |
| 14 | 196+ N | 8 | 54+0,5N | 32 |
| 15 | 142+ N | 4 | 32+0,5N | 14 |
| 16 | 156+ N | 8 | 39+0,5N | 20 |
| 17 | 171+ N | 8 | 49+0,5N | 26 |
| 18 | 140+ N | 3 | 26+0,5N | 14 |
| 19 | 156+ N | 7 | 34+0,5N | 12 |
| 20 | 115+ N | 3 | 28+0,5N | 10 |
| 21 | 142+ N | 9 | 35+0,5N | 12 |
| 22 | 134+ N | 3 | 29+0,5N | 11 |
| 23 | 147+ N | 6 | 36+0,5N | 16 |
| 24 | 173+ N | 7 | 41+0,5N | 21 |
| 25 | 144+ N | 4 | 30+0,5N | 12 |

# РОЗДІЛ 3. НЕЛІНІЙНІ МОДЕЛІ РЕГРСІЇ

Поєднання економічної теорії з практичними результатами є наріжним каменем економетрії. Економічна теорія виявила і дослідила значну кількість сталих і стабільних зв'язків між різними показниками. Апарат оцінювання параметрів моделей лінійної регресії (лінійними за змінними та параметрами) розроблений досить добре. Використання лінійних моделей для моделювання економічних залежностей в багатьох випадках дають цілком задовільні результати, які можуть бути використані для аналізу й прогнозу досліджуваних економічних систем (процесів).

Внаслідок багатогранності й складності за своєю структурою економічних процесів обмежуватися розглядом лише лінійних моделей стає неможливим, оскільки в більшості випадків економічні залежності не можуть бути описані лінійними рівняннями. Так, наприклад, нелінійними є виробничі функції (залежності між обсягом виробленої продукції і основними факторами виробництва – працею, капіталом), функції попиту (залежність між попитом на товари або послуги і їх цінами або прибутком) та інші.

При аналізі залежності еластичності попиту від ціни доцільно використати степеневу модель

$$Y = \beta_0 X^{\beta_1},$$

при аналізі залежності витрат $Y$ від обсягу виробництва $X$ найбільш обґрунтованою буде поліноміальна модель (степені $p$)

$$Y = \sum_{i=0}^{p} \beta_i X^i = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \cdots + \beta_p X^p.$$

Для дослідження виробничих функцій набула чинності така степенева (мультиплікативна) модель:

$$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2}.$$

Широко використовуються в економетрії також гіперболічна (обернена)

$$Y = \beta_0 + \beta_1(1/X)$$

та показникова (експоненціальна) моделі

$$Y = \beta_0 e^{\beta_1 X}.$$

Нелінійність регресії може бути обумовленою двома причинами:

1) нелінійність за пояснювальними змінними (але лінійні за параметрами, що оцінюються);
2) нелінійність за коефіцієнтами (параметрами) регресії.

При оцінюванні економетричних нелінійних моделей регресії виникають певні труднощі, які відсутні при оцінюванні лінійних економетричних моделей. Тому, по можливості, нелінійні специфікації за допомогою відповідних перетворень зводять до лінійних, тобто виконується їх *лінеаризація*.

## 3.1. Нелінійні за змінними моделі

В нелінійних за змінними моделях виконуються такі дії:

1) регресори, що мають степінь, відмінну від першої, змінюються на інші незалежні змінні;
2) до нової системи змінних застосовується МНК;
3) після того, як отримано рівняння регресії з оціненими параметрами, нові незалежні змінні, що були введені у нього, замінюються на вихідні.

Для перетворення рівняння поліноміальної регресії, наприклад, другої степені

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2$$

до лінійного вигляду, використовується така заміна:

$$Z_1 = X_1, \quad Z_2 = X_2, \quad Z_3 = X_1^2, \quad Z_4 = X_2^2.$$

Маючи на увазі, що $Z$ – матриця регресорів, використовуємо МНК і отримаємо таку оцінку вектора параметрів:

$$\hat{\beta} = (Z^T Z)^{-1} Z^T Y.$$

Для даної регресії можна побудувати точністні оцінки точкових оцінок параметрів та інтервальні оцінки. Випадкове збурення ε початкового рівняння не зазнавало ніяких перетворень, тому, якщо передумови Гаусса-Маркова виконуються для початкової специфікації, то вони виконуються і для перетвореної.

Поліноміальні рівняння регресії степені $p$

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \cdots + \beta_p X^p + \varepsilon \quad \left(Z_i = X^i, \quad i = \overline{1,p}\right)$$

використовуються в мікроекономічному аналізі. У даній специфікації параметри добре піддаються змістовному поясненню:

$\beta_0$ – значення ендогенної змінної при нульовому значенні регресора, тобто початковий рівень;

$\beta_1$ – приріст ендогенної змінної при збільшенні значення регресора на одиницю (швидкість зростання);

$\beta_2$ – швидкість зміни швидкості (прискорення зростання);

$\beta_3$ - зміна прискорення і т.д.

Поліноміальна специфікація при $p = 3$ добре описує залежність загальних витрат $TC$ (total costs) від обсягу випуску продукції $Q$:

$$TC = \beta_0 + \beta_1 Q + \beta_2 Q^2 + \beta_3 Q^3 + \varepsilon \ .$$

Параболічна регресія ($p = 2$) може використовуватися при моделюванні:

- залежності середніх $AC$ (*average costs*), або граничних $MC$ (*marginal costs*) витрат від обсягу випуску продукції $Q$

$$AC = \beta_0 + \beta_1 Q + \beta_2 Q^2 + \varepsilon;$$

- залежність прибутку фірми π від витрат на рекламу $Y$

$$\pi = \beta_0 + \beta_1 Y + \beta_2 Y^2 + \varepsilon \text{ і т.ін.}$$

Для перетворення *гіперболічної регресії*

$$Y = \beta_0 + \beta_1 \cdot (1/X) + \varepsilon \tag{3.1}$$

до лінійного вигляду використовується заміна:

$$\tilde{X} = 1/X.$$

В економічних дослідженнях гіперболічна регресія використовується при моделюванні залежності, де при необмеженому збільшенні регресора ендогенна змінна асимптотично наближується до деякого граничного значення. Так, для моделі (3.1) при необмеженому збільшенні регресора значення $Y$ наближається до значення параметра $\beta_0$.

У зв'язку з цим економічна інтерпретація параметра $\beta_0$ – *рівень ендогенної змінної, яка усталюється при великих значеннях регресора, а параметр $\beta_1$ характеризує швидкість наближення до даного рівня*.

Прикладом використання специфікації (3.1) є моделі, що описують залежність попиту від цін ($\beta_1 < 0$) або прибутку ($\beta_1 > 0$) - криві Енгеля; або моделі попиту на товари першої потреби і/або відносної розкоші залежно від рівня прибутку споживача (функції Торнквіста).

Специфікація (3.1) також використовується при моделюванні залежності між рівнем безробіття у відсотках та процентною зміною заробітної плати (крива Філліпса).

Гіперболічна регресія вигляду

$$Y = \frac{1}{\beta_0 + \beta_1 X + \varepsilon}$$

зводиться до лінійної регресії заміною

$$\tilde{Y} = 1/Y.$$

Разом із тим, рівностороння гіпербола - не єдина можлива функція для опису кривої Енгеля. У 1943 році економіст і статистик Е.Уоркінг і у 1964 році відомий англійський економетрист С.Лізер для цієї ж мети використовували напівлогарифмічну криву

$$y = \beta_0 + \beta_1 lnx + \varepsilon.$$

Зробивши заміну змінних: $z$ = $lnx$, знову одержуємо лінійне за параметрами і нелінійне за пояснювальною змінною $x$ рівняння: $y = \beta_0 + \beta_1 z + \varepsilon$ . Оцінки параметрів $\beta_0$ і $\beta_1$ можуть бути знайдені МНК. Система нормальних рівнянь при цьому має вигляд:

$$\begin{cases} n\beta_0 + \beta_1 \sum \ln x = \sum y \\ \beta_0 \sum \ln x + \beta_1 \sum (\ln x)^2 = \sum y \ln x \,. \end{cases}$$

При дослідженні врожайності, трудомісткості та інших показників сільськогосподарського виробництва використовується така лінійна за параметрами та нелінійна за пояснювальною змінною $x$ модель:

$$y = \beta_0 + \beta_1 \sqrt{x} + \varepsilon.$$

Система нормальних рівнянь для оцінки її параметрів має вигляд:

$$\begin{cases} n\beta_0 + \beta_1 \sum \sqrt{x} = \sum y \\ \beta_0 \sum \sqrt{x} + \beta_1 \sum x = \sum y\sqrt{x} \,. \end{cases}$$

### 3.1.1. Поліноміальна модель

Загальний запис цієї моделі такий:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_p x_i^p \quad + \varepsilon_i, \;\; i = \overline{1,n} \,, \qquad (3.2)$$

де збурення $\varepsilon_i$ задовольняють умовам використання МНК.

Зі зміною індексу $i = \overline{1,n}$ отримаємо систему лінійних алгебраїчних рівнянь, яку у векторно-матричній формі можна записати так:

$$y = X \cdot \beta + \varepsilon, \qquad (3.3)$$

де

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix}; \;\; X = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p \\ 1 & x_2 & x_2^2 & \dots & x_2^p \\ 1 & x_3 & x_3^2 & \dots & x_3^p \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^p \end{pmatrix}; \;\; \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{pmatrix}; \;\; \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Статистичним образом моделі (3.3) є модель

$$y_i = \beta_0^* + \beta_1^* \cdot x_i + \beta_2^* \cdot x_i^2 + \beta_3^* \cdot x_i^3 + \cdots + \beta_p^* \cdot x_i^p + e_i, \;\; i = \overline{1,n},$$

де $\beta_0^*$, $\beta_1^*$, $\beta_2^*$, $\beta_3^*$, … , $\beta_p^*$ - точкові незміщені статистичні оцінки для теоретичних параметрів $\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_p$ відповідно з рівняння (3.2).

З урахуванням всіх значень індексу $i = \overline{1,n}$ одержана система лінійних рівнянь у векторно-матричній формі запису має вигляд

$$\vec{y} = X \cdot \vec{\beta}^* + \vec{e}, \tag{3.4}$$

де

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_n \end{pmatrix}; \; X = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^p \\ 1 & x_2 & x_2^2 & \ldots & x_2^p \\ 1 & x_3 & x_3^2 & \ldots & x_3^p \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_n & x_n^2 & \ldots & x_n^p \end{pmatrix}; \; \vec{\beta}^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \\ \ldots \\ \beta_p^* \end{pmatrix}; \; \vec{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \ldots \\ e_n \end{pmatrix}.$$

Використовуючи МНК, отримуємо

$$\vec{\beta}^* = (X^T X)^{-1} X^T \cdot \vec{y}; \tag{3.5}$$

$$cov\left(\vec{\beta}^* (\beta^*)^T\right) = S_\varepsilon^2 (X^T X)^{-1},$$

де

$$S_\varepsilon^2 = \frac{\sum_{i=1}^n e_i^2}{n-p-1}.$$

*Квадратичні функції*

Квадратичні функції використовуються для опису дуже широкого спектру різноманітних економічних процесів, завдяки їх універсальним властивостям. У загальному випадку квадратична функція має вигляд:

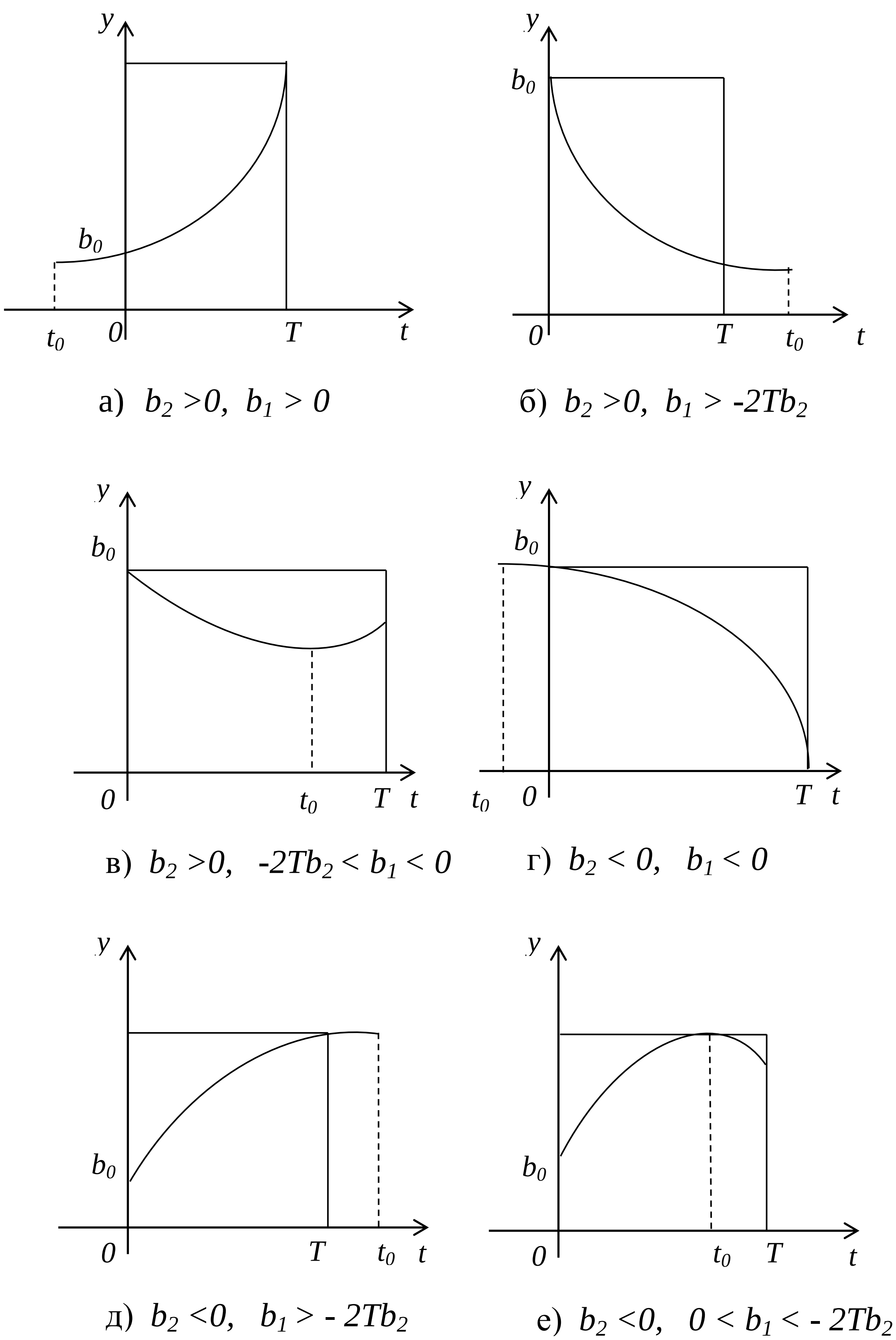
$$y = b_0 + b_1 x + b_2 x^2.$$

Якщо фактор $x$ інтерпретувати як зміну часу $t$, то цю модель можна записати у вигляді:

$$y = b_0 + b_1 t + b_2 t^2.$$

Виходячи із значень параметрів $b_0$, $b_1$, $b_2$, крива може відображати еволюцію – дуже різну на інтервалах часу від 0 до $T$ ([19]).

Якщо $b_2 > 0$, маємо параболу, яка містить мінімум; при $b_2 < 0$ – максимум. Екстремум функції досягається у точці $t = - b_1 / (2b_2)$.




а) $b_2 > 0,\ \ b_1 > 0$

б) $b_2 > 0,\ \ b_1 > -2Tb_2$

в) $b_2 > 0,\ \ -2Tb_2 < b_1 < 0$

г) $b_2 < 0,\ \ b_1 < 0$

д) $b_2 < 0,\ \ b_1 > -2Tb_2$

е) $b_2 < 0,\ \ 0 < b_1 < -2Tb_2$

Рис. 3.1  Квадратична функція

Можливі такі випадки поведінки квадратичної функції.

1. Параметри $b_2 > 0$, $b_1 > 0$.

Квадратична функція описує поліпшене зростання (абсциса вершини параболи $t_0$ розташована перед часом $T$; рис. 3.1 *а*)).

2. Параметри $b_2 > 0$, $b_1 > - 2 \cdot T \cdot b_2$.

Квадратична функція описує уповільнений спад (абсциса вершини параболи $t_0$ розташована після часу $T$; рис. 3.1 *б*)).

3. Параметри $b_2 > 0$, $- 2 \cdot T \cdot b_2 < b_1 < 0$.

Класичний випадок – парабола у точці з абсцисою $t_0$ ($t_0 \in [0; T]$) досягає мінімуму; рис. 3.1 *в*).

4. Параметри $b_2 < 0$, $b_1 < 0$.

Квадратична функція описує прискорений спад (абсциса вершини параболи $t_0$ розташована перед 0; рис. 3.1 *г*)).

5. Параметри $b_2 < 0$, $b_1 > - 2 \cdot T \cdot b_2$.

Квадратична функція описує уповільнене зростання (абсциса вершини параболи $t_0$ розташована після часу $T$; рис. 3.1 *д*)).

6. Параметри $b_2 < 0$, $0 < b_1 < - 2 \cdot T \cdot b_2$.

Класичний випадок – парабола у точці з абсцисою $t_0$ ($t_0 \in [0; T]$) досягає максимуму; рис. 3.1 *е*).

Квадратична функція зводиться до багатофакторної регресії заміною змінних: $x_1 = t$, $x_2 = t^2$. Невідомі параметри розраховуються як для випадку багатофакторної регресії.

### 3.1.2. Гіперболічна модель

Гіперболічна модель в загальному випадку має такий вигляд:

$$y_i = \beta_0 + \beta_1 \cdot (1/x_i) + \varepsilon_i, \;\; i = \overline{1,n} \tag{3.6}$$

Для всіх значень індексу $i = \overline{1,n}$ рівняння (3.6) у векторно-матричній формі набуде такого вигляду:

$$y = X \cdot \beta + \varepsilon, \tag{3.7}$$

де введені такі матриці та вектори:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix}; \; X = \begin{pmatrix} 1 & 1/x_1 \\ 1 & 1/x_2 \\ 1 & 1/x_3 \\ \dots & \dots \\ 1 & 1/x_n \end{pmatrix}; \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Оскільки випадкові збурення задовольняють вимогам МНК і, враховуючи, що статистичний образ моделі (3.7) має вигляд

$$y = X \cdot B + e \tag{3.8}$$

за аналогією з моделлю (3.4) отримаємо формули:

$$B = (X^T X)^{-1} X^T \cdot \vec{y}; \tag{3.9}$$

$$cov(BB^T) = S_\varepsilon^2 (X^T X)^{-1},$$

$$S_\varepsilon^2 = \sum_{i=1}^{n} e_i^2 \Big/ (n-2).$$

Графіки гіперболічних моделей (3.8) визначаються знаками параметрів $b_0$ та $b_1$ . Розглянемо три варіанти:

I. При $b_0 < 0$, $b_1 > 0$ крива залежності між змінними $Y$ та $X$ має вигляд:

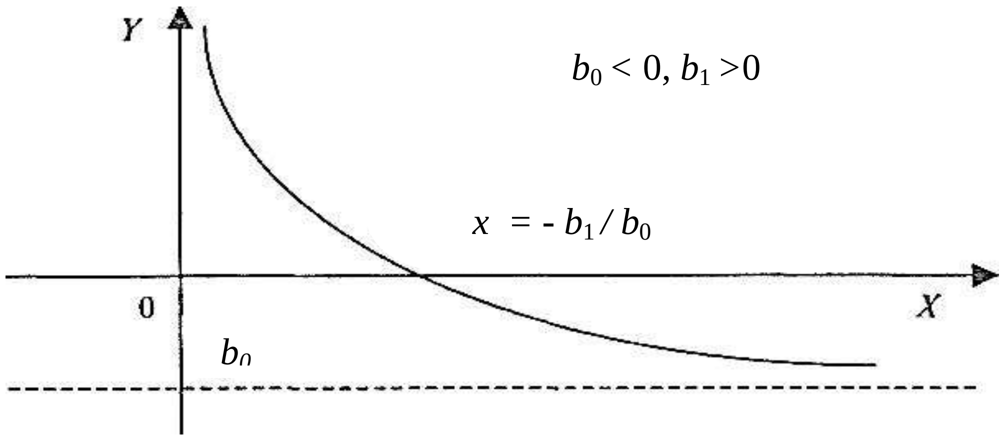



Рис. 3.2

На рис.3.2 зображена так звана крива Філліпса (О.Філліпс (1914 – 1975)-новозеландський економіст; у 1949 році винайшов гідравлічний комп'ютер *MONIAC*). Модель (3.6) використовується для аналізу залежності між змінною у відсотках заробітної плати $Y$ від рівня безробіття $X$ у відсотках.

II. При $b_0 > 0$, $b_1 < 0$ крива залежності між змінними $Y$ та $X$ буде мати вигляд:

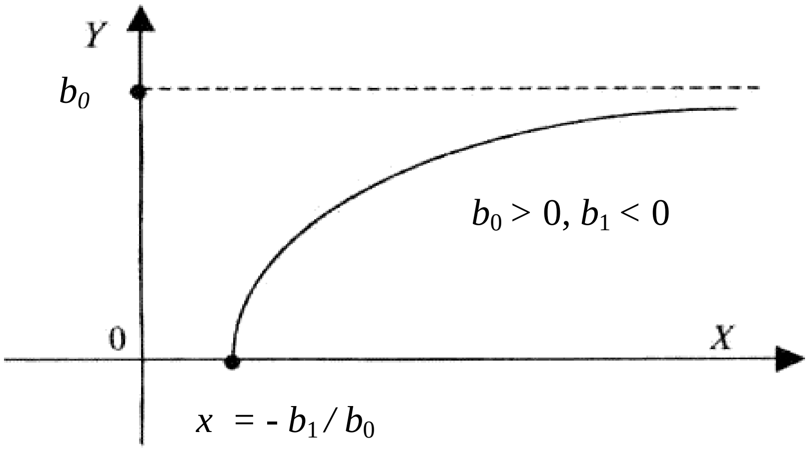



Рис. 3.3

На рис. 3.3 зображена функція Торнквіста. За допомогою цієї функції можна описати залежність між попитом $Y$ на товари першої необхідності й прибутком $X$.

III. При $b_0 > 0$, $b_1 > 0$ крива залежності між факторними ознаками $Y$ та $X$ набуде вигляду:

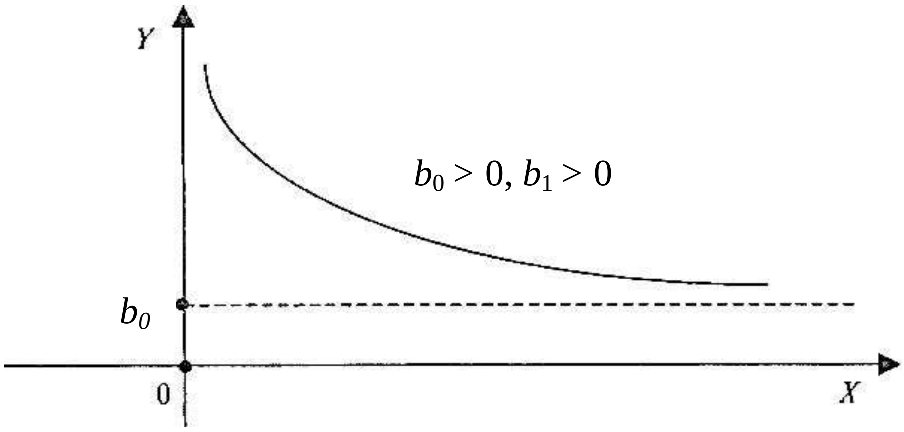



Рис. 3.4

Така залежність має місце при дослідженні зв'язку між середніми фіксованими витратами $Y$ і обсягами випуску продукції $X$.

В окремих випадках може використовуватися обернена модель такого вигляду:

$$y = \frac{1}{\beta_0 + \beta_1 x + \varepsilon}.$$

Це - нелінійна модель, яка є різновидністю гіперболи. При $\beta_1 > 0$ ($\beta_1 < 0$) рівняння характеризує обернену (пряму) пропорційну залежність результативної ознаки від фактора. Модель доцільно використовувати при дуже швидкому (повільному) збільшенні рівнів результативної ознаки із зростанням значень фактора. Якщо у рівнобічній гіперболі $y = \beta_0 + \frac{\beta_1}{x} + \varepsilon$ перетворенню підлягає пояснювальна змінна $1/x = z$, то для набування лінійної форми залежності в оберненої моделі перетворюється $y$, а саме: $1 / y = z$. В наслідок цього обернена модель стає внутрішньо нелінійною і вимога МНК виконується не для фактичних значень ознаки $y$, а для їх обернених величин $1/y$, а саме: $\sum(z - \hat{z}_x)^2 \rightarrow min$. Відповідно:

$$\sum\left(\frac{1}{y}\right) = \sum \hat{z}_{x,} \quad \text{але} \quad \sum y \neq \sum \hat{y}_x.$$

### 3.2. Нелінійні моделі за коефіцієнтами рівняння регресії

До нелінійних за коефіцієнтами (параметрами) рівняння регресії відносять рівняння, в яких залежна змінна нелінійним чином залежить від коефіцієнтів регресії. Прикладами таких нелінійних регресійних моделей є такі функції:

- степенева $\quad Y = \beta_0 X^{\beta_1} \cdot \varepsilon;$ (3.10)
- показникова $\quad Y = \beta_0 {\beta_1}^X \cdot \varepsilon;$ (3.11)
- експоненціальна $\quad Y = \beta_0 e^{\beta_1 X} \cdot \varepsilon.$ (3.12)

Для обчислення параметрів нелінійних рівнянь регресій можливі два підходи.

*Перший підхід* полягає у використанні деякого (найчастіше нелінійного) перетворення, що приводить до лінійної регресії, але вже відносно нових коефіцієнтів і/або нових змінних. Цей підхід проілюструємо степеневою регресією, яка використовується в економетричних дослідженнях, наприклад, при вивченні залежності попиту товару від його ціни. Після логарифмування рівняння (3.10)

$$ln\ Y = ln\ \beta_0 + \beta_1\ ln\ X + ln\ \varepsilon$$

і заміни

$$Y' = lnY,\ \ \beta_0' = ln\beta_0,\ \ X' = lnX,\ \ \varepsilon' = ln\varepsilon$$

маємо лінійну регресійну модель

$$Y' = \beta_0' + \beta_1 \ X' + \varepsilon', \tag{3.13}$$

якій відповідає рівняння парної лінійної регресії:

$$\hat{y}' = b_0' + b_1 x'. \tag{3.14}$$

Параметри регресії $b_0', b_1$ обчислюються за допомогою МНК. Виконуючи зворотне перетворення $b_0 = e^{b_0'}$, отримаємо шукані оцінки $b_0$, $b_1$ для параметрів регресії (3.10).

*Зауваження***.** Ефективність оцінок, що одержані за допомогою МНК, засновується на припущенні про те, що збурення $\varepsilon_i$ не корельовані між собою і підлягають нормальному розподілу $N(0,\sigma^2)$, тобто мають однакову дисперсію $\sigma^2$. Але виконання нелінійних перетворень приводить до порушення цього припущення. Для того, щоб проілюструвати це, розглянемо перетворене рівняння регресії (3.14). Коефіцієнти цього рівняння є ефективні оцінки для $\beta_0', \beta_1'$, якщо $\varepsilon' = ln\, \varepsilon \sim N(0, \sigma^2)$, тобто збурення $\varepsilon_i$ початкової моделі (3.10) повинні мати логарифмічно нормальний розподіл, що дуже рідко зустрічається на практиці.

Порушення властивості гомоскедастичності приводить до того, що обчислені на основі МНК коефіцієнти будуть *незміщеними, обґрунтованими* оцінками для відповідних коефіцієнтів регресійної моделі, але вони *неефективні*, тобто можна обчислити (використовуючи інші алгоритми) оцінки із меншою дисперсією.

*Другий підхід* використовується у випадках, коли неможливо підібрати перетворення для переходу до нової лінійної регресії.

Для ілюстрації розглянемо приклад регресійної моделі

$$Y = \beta_0 X^{\beta_1} + \varepsilon. \tag{3.15}$$

Логарифмування цього рівняння

$$lnY = ln\left(\beta_0 X^{\beta_1} + \varepsilon\right)$$

не приводить до лінійної регресійної моделі.

У цих випадках оцінка для коефіцієнтів регресійної моделі обчислюється на основі мінімізації деякого функціонала, наприклад, функціонала МНК. Так, для моделі (3.15) рівняння регресії має вигляд:

$$\hat{y}(x) = b_0 x^{b_1}, \qquad (3.16)$$

а функціонал МНК, що мінімізується, має вигляд:

$$F(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_0 x_i^{b_1})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

Існує багато алгоритмів мінімізації різних функціоналів. Деякі алгоритми реалізовані у табличному процесорі Excel (команда *Поиск решения* пункта меню *Сервис* та ін.) детально описані у Розділі 5.

### 3.2.1. Показникова (експоненціальна) модель

Модель вигляду

$$Y = \beta_0 {\beta_1}^{x} \qquad (3.17)$$

називається *показниковою*.

Випадкові збурення $\varepsilon_i$ $(i = \overline{1, n})$ можна ввести в модель (3.17) у вигляді

$$y_i = \beta_0 {\beta_1}^{x_i} e^{\varepsilon_i}, \qquad (3.18)$$

або

$$y_i = \beta_0 {\beta_1}^{x_i} \varepsilon_i. \qquad (3.19)$$

При логарифмуванні моделей (3.18) і (3.19) одержимо відповідно

$$\ln y_i = \ln\beta_0 + x_i \ln \beta_1 + \varepsilon_i \qquad (3.20)$$

і

$$\ln y_i = \ln\beta_0 + x_i \ln \beta_1 + \ln \varepsilon_i\,. \qquad (3.21)$$

У формі запису (3.21) порушується структура вимог до випадкових збурень ($M(\varepsilon_i) = 0$, $\varepsilon_i \sim N(0; \varepsilon_i)$), а саме $\varepsilon_i$ в цьому випадку як випадкова величина має логарифмічний закон розподілу ймовірностей, тому надалі використовуватимемо модель (3.18), (3.20).

Експоненціальна функція може набирати різних чисельних еквівалентних форм:

$y = \beta_0 \cdot \beta_1^{x}$ (основна форма $\beta_1 > 0$);

$y = \beta_0 \cdot e^{b_1 x}$ ($\beta_1$ замінюємо на $e^{b_1}$);

$y = \beta_0(1-r)^x$, ($\beta_1$ замінюємо на $1 - r$);
$\mathrm{y} = \mathrm{e}^{\mathrm{b_0+b_1x}}$ ($\beta_0$ замінюється на $\mathrm{e}^{\mathrm{b_0}}$, $\beta_1$ - на $\mathrm{e}^{\mathrm{b_1}}$),
$y$= $10^{\mathrm{b_0+b_1x}}$, ($\beta_0$ замінюється на $10^{\mathrm{b_0}}$, $\beta_1$ – на $10^{\mathrm{b_1}}$).

Усі ці форми використовуються на практиці для опису різних економічних процесів, наприклад, форму $y = \beta_0(1-r)^x$ найчастіше використовують у фінансах. У цьому разі $r$ інтерпретується як норма річного відсотка. Розглянемо декілька прикладів застосування експоненціальної функції у бізнесі та фінансах.

**Приклад 3.1**. «*Правило 70*». Припустимо, що капітал $C_0$ знаходиться у банку протягом $t$ років із річним відсотком $r$. Він змінюється відповідно до функції $y = \beta_0(1-r)^x$, тобто через $t$ років капітал дорівнюватиме:

$$C_t = C_0 \cdot (1 + r)^t.$$

Із експоненціальної форми $y = \beta_0 \cdot e^{b_1 x}$ можна отримати так зване «*правило 70»,* поширене в фінансових розрахунках. «Правило 70» дає значення часу $t$**,** через яке змінна подвоїть своє значення відносно початкового часу. Нехай у початковій період часу має місце капітал $C_0$, а в період часу $t$ - капітал $C_t$ . Подвоєння капіталу можна знайти, якщо перейти від форми $y = \beta_0(1-r)^x$ до форми $y = \beta_0 e^{b_1 x}$:

$$2C_0 = C_0 e^{b_1 t}, 2 = e^{b_1 t}, \qquad t = \frac{ln\,2}{b_1} \approx \frac{0{,}6931}{b_1} \approx \frac{0{,}70}{b_1}, \qquad t \approx \frac{0{,}70}{b_1}.$$

*Висновок*. Капітал подвоїться через $0{,}70/b_1$ років. Правило має назву «70», тому що $ln\,2 \approx 0{,}70$.

**Приклад 3.2.** *Приведення витрат до поточного часу*.

При порівнянні різних інвестиційних проектів в першу чергу порівнюють кошти, які на них втрачаються, та майбутні прибутки, які вони будуть приносити. Найпростіше це зробити, оцінюючи майбутні витрати та прибутки у вартості поточного року. Наприклад, заощаджуючи в банку *А* грн. у поточному році при ставці процента $r$, через $t$ років клієнт отримає суму:

$$B = A \cdot e^{rt}.$$

Відповідно *В* гривень, отриманих у $t$ році при ставці відсотка $r$, у поточному році коштуватимуть:

$$A = B \cdot e^{-rt}.$$

Величина *А* називається *приведеними витратами В* гривень через $t$ років до поточного часу (при нормі відсотка $r$).

Приведені витрати можуть також визначатись для потоку платежів. При нормі відсотка $r$ приведені витрати платежу $B_1$, який мали б через $t_1$ років, платежу $B_2$, через $t_2$ років, …, платежу $B_n$, відповідно через $t_n$ років мають вигляд:

$$A = B_1 \cdot e^{-rt_1} + B_2 \cdot e^{-rt_2} + \dots + B_n \cdot e^{-rt_n}.$$

Експоненціальні криві використовуються для опису швидко зростаючих або спадаючих економічних процесів. При цьому, якщо

$\beta_1 > 0$ ($b_1 > 0$) – функція зростає до нескінченості,

$\beta_1 < 0$ ($b_1 < 0$) – функція спадає до 0.

Параметр $b_1$ можна інтерпретувати як коефіцієнт зростання у часі.

Виконуючи логарифмічні перетворення, можна звести експоненціальну модель у будь-якій формі до лінійної функції, що дає змогу розрахувати параметри за допомогою МНК та використовувати подальший аналіз моделі, як і в разі простої лінійної регресії. Таким чином, маємо:

$$y = \beta_0 \cdot \beta_1^x, \qquad \ln y = \ln \beta_0 + x \cdot \ln\beta_1;$$

$$y = b_0 e^{b_1 x}, \qquad \ln y = \ln b_0 + b_1 \cdot x;$$

$$y = \beta_0 (1 - r)^x, \qquad \ln y = \ln \beta_0 + x \ln(1 - r);$$

$$y = e^{b_0 + b_1 x}, \qquad \ln y = b_0 + b_1 \cdot x;$$

$$y = 10^{b_0 + b_1 x}, \qquad \lg y = b_0 + b_1 \cdot x.$$

Зробивши відповідну заміну змінних, можна перейти до моделі лінійної регресії, розрахунок параметрів та всі необхідні дослідження якої детально розроблені.

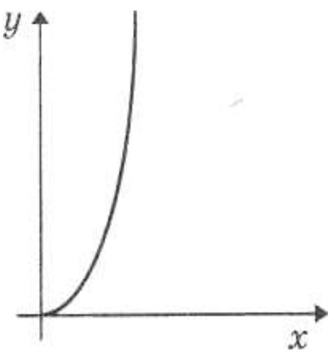
### 3.2.2. Степенева модель

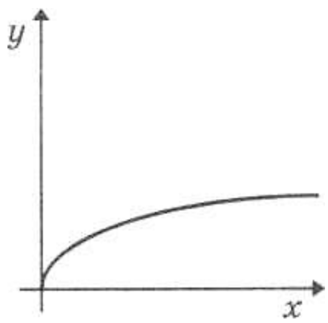
Степенева функція

$$y = \beta_0 \cdot x^{\beta_1}$$

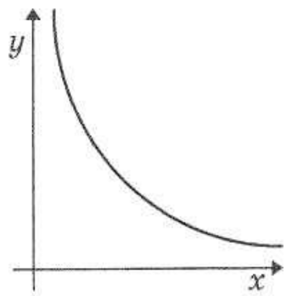
описує дуже широкий спектр економічних процесів.

Розглянемо типовий для економічних процесів випадок, коли параметр $\beta_0 > 0$. Якщо значення параметра $\beta_1$ — не ціле число, то розглядають лише випадок, коли $x \geq 0$. При цьому залежно від знака параметра $\beta_1$ степенева функція описуватиме різні економічні процеси: прискорене зростання, уповільнене зростання та спад. Якщо $\beta_1 = 1$, степенева функція перетворюється на лінійну. Ці різні ситуації зображені на рис.3.5 (а, б, в, г). Якщо параметр $\beta_1$ степеневої функції - ціле число, то залежно від того, парне чи непарне його значення, графік функції має різний вигляд. Якщо $\beta_1$ - парне, тобто його можна записати у вигляді: $\beta_1 = 2k,\ k \in Z_+$, тоді $y \in [0,+\infty)$, а графік функції симетричний відносно осі ординат (рис. 3.6, а)).

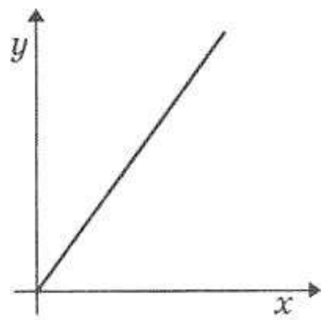



а) $\beta_1 > 1$ (прискорене зростання) б) $0 < \beta_1 < 1$ (уповільнене зростання)

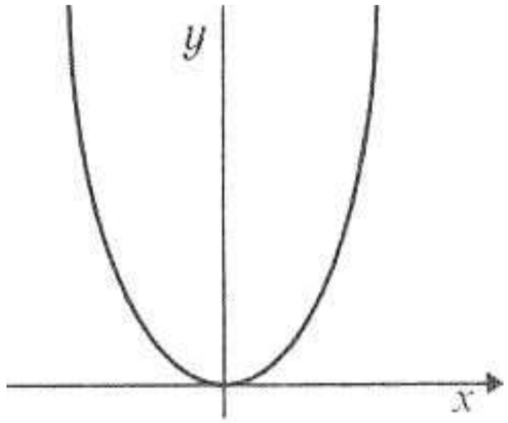



*в)* $\beta_1 < 0$

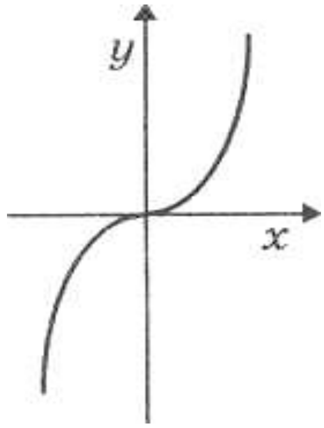



*г)* $\beta_1 = 1$

Рис. 3.5. Вигляд степеневої функції, коли $\beta_1$ – не ціле число, $x \geq 0$.




б) $\beta_1$ - парне




в) $\beta_1$ - непарне

Рис. 3.6. Графік степеневої функції, коли $\beta_1$ – ціле число.

Для того, щоб розрахувати невідомі параметри степеневої (мультиплікативної) функції, зведемо її до лінійної функції

$$z = b_0 + b_1 x_1,$$

використовуючи логарифмічне перетворення

$$\ln y = \ln \beta_0 + \beta_1 \ln x,$$

і заміну $\ln y = z$, $\ln \beta_0 = b_0$, $\beta_1 = b_1$, $\ln x = x_1$. Це дає змогу обчислити параметри за допомогою МНК.

На практиці степеневі функції використовуються для опису різних економічних процесів: кривих байдужості, попиту на товари різних категорій і таке інше. Прикладом багатофакторної степеневої моделі є, наприклад, така виробнича функція Кобба –Дугласа:

$$Y = aL^{\alpha_L}C^{\alpha_C},$$

де $Y$ – величина валового регіонального продукту; $L$ – зайнятість населення (*labor*), $C$ – основні засоби у фактичних цінах (*capital*).

### 3.2.3. Кореляція для нелінійної регресії

Для оцінки тісноти зв’язку нелінійної кореляційної залежності використовується *індекс кореляції* $i_r$ (кореляційне відношення, п.1.6.2, формула (1.27)):

$$i_r = \sqrt{1 - \frac{\sigma^2_{\text{зал}}}{\sigma^2_y}},$$

де $\sigma^2_y$ – загальна дисперсія результативної ознаки $Y$;

$\sigma^2_{\text{зал}}$ – дисперсія випадкових відхилень (помилок), або залишкова (непояснена) дисперсія.

Оскільки $\sigma^2_y = \frac{1}{n}\sum(y_i - \bar{y})^2$ і $\sigma^2_{\text{зал}} = \frac{1}{n}\sum(y_i - \hat{y}_x)^2$, то індекс кореляції можна записати так:

$$i_r = \sqrt{1 - \frac{\sum(y_i - \hat{y}_x)^2}{\sum(y_i - \bar{y})^2}}.$$

Величина даного показника знаходиться у проміжку: $0 \leq i_r \leq 1$. Чим ближче величина індексу кореляції до одиниці, тим тісніше зв'язок розгляданих ознак, тим надійніше рівняння регресії.

Якщо нелінійне відносно пояснюючої змінної рівняння регресії при лінеаризації приймає форму лінійного рівняння парної регресії, то для оцінювання тісноти зв'язку може використовуватися лінійний коефіцієнт кореляції, величина якого у цьому випадку збігається із індексом кореляції.

Зовсім інша справа у випадку, коли перетворення рівняння у лінійну форму пов'язане із залежною змінною. У цьому випадку лінійний коефіцієнт кореляції за перетвореними значеннями ознак дає лише приблизну оцінку тісноти зв'язку і чисельно не збігається із індексом кореляції. В посібнику [26] на прикладі квадратичної, напівлогарифмічної та гіперболічної функцій показано, що значення лінійного коефіцієнту кореляції і індексу кореляції збігаються (на відміну від степеневої та експоненціальної залежностей).

Внаслідок близькості результатів і простоти розрахунків з використанням комп'ютерних програм для характеристики тісноти зв'язку у нелінійних функціях широко використовується лінійний коефіцієнт кореляції. Незважаючи на близькість значень лінійного коефіцієнта кореляції і індексу кореляції в нелінійних функціях із перетворенням значень ознаки *Y* треба пам'ятати, що при лінійній залежності ознак один й той же лінійний коефіцієнт кореляції характеризує регресію як *Y* на *X* , так і *X* на *Y*. Індекси кореляції нелінійних регресій *Y* на *X* та *X* на *Y* не рівні між собою. Коефіцієнт детермінації $R^2$ (п. 1.6.2) для нелінійних зв'язків називається *індексом детермінації* і має той же сенс, що і коефіцієнт детермінації.

Якщо після перетворення рівняння регресії (нелінійне за пояснювальними змінними) приймає форму парного лінійного рівняня регресії, то для оцінки тісноти зв'язку може бути використан лінійний коефіцієнт кореляції для перетвореної величини фактора. Якщо перетворення в лінійну форму пов'язані з результативним фактором (нелінійність за параметрами), то лінійний коефіцієнт кореляції за перетвореними значеннями факторів дає лише аближену оцінку тісноти зв'язку. Він не дорівнює індексу кореляції, оскільки лінійний коефіцієнт кореляції тепер розраховується між отриманими при лінеаризації логарифмами, а коефіцієнт детермінації використовує суми квадратів відхилень ознаки *Y*, а не її логарифма.

Індекс кореляції використовується для перевірки істотності у цілому рівняння нелінійної регресії за *F*- критерієм Фішера:

$$F_{\text{факт}} = \frac{i_r^2}{1 - i_r^2} \cdot \frac{n - m - 1}{m},$$

де $i_r$ – індекс кореляції, *n* – число спостережень, $m$ – число факторів. Величина $m$ характеризує ступінь вільності для факторної суми квадратів, а ($n$ – $m$ – 1) – ступінь вільності для залишкової суми квадратів. Індекс детермінації нелінійної залежності можна порівняти із коефіцієнтом детермінації для обґрунтування можливості використання лінійного рівняння регресії. Чим більша кривизна лінії регресії, тим величина коефіцієнту детермінації менша за індекс детермінації нелінійної залежності. Близькість цих показників означає, що немає необхідності ускладнювати форму рівняння регресії і можна використовувати лінійне рівняння регресії. На практиці, якщо величина різниця цих величин не перевищує 0.1, то припущення про лінійну форму зв'язку вважається вірним. У супротивному випадку проводиться оцінювання істотності їх різниці, що обчислені за одними й тими ж вихідними даними, через *t*- критерій Ст'юдента:

$$t = \frac{i_r^2 - r^2}{2\sqrt{\left((i_r^2 - r^2) - (i_r^2 - r^2)^2 \cdot (2 - \left(i_r^2 + r^2\right))\right)/n}}.$$

Знаменник дробу є помилкою різниці між індексом детермінації та коефіцієнтом детермінації.

Якщо $t_{факт} > t_{таб}$ , то різниця між розглянутими показниками кореляції істотні і заміна нелінійної регресії лінійним рівнянням неможлива. Якщо величина $t < 2$, то різниця між індексом кореляції і лінійним коефіцієнтом кореляції несуттєва, і, отже, можливо використання лінійної регресії, навіть якщо є припущення про деяку нелінійність розгляданих співвідношень ознак фактора і результату.

Якщо індекс детермінації і коефіцієнт детермінації наближено рівні, використовують стандартну процедуру, яка відома під назвою теста Бокса – Кокса [7].

## 3.3. Приклади застосування нелінійних економетричних моделей в економічних дослідженнях

**Приклад 3.3** [9]. Досліджується залежність між прибутком банків *Y* (млн. грн.) та величиною залучених коштів (депозити фізичних і юридичних осіб) *X* (млн. грн.) на основі вибіркових даних, що наведені в таблиці 3.1.

1) Визначити форму регресійної залежності між прибутком банків *Y* та величиною залучених коштів *X*.
2) Знайти статистичні оцінки для параметрів визначеної регресійної залежності.

3) Перевірити адекватність моделі за коефіцієнтом детермінації та індексом кореляції.

4) Перевірити статистичну значущість параметрів моделі.

5) Обчислити коефіцієнт еластичності при $x_i = 55{,}5$ (млн..грн.).

Таблиця 3.1

| *i* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* |
|---|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 0,8 | 0,9 | 0,9 | 1,1 | 1,3 | 1,7 | 2,2 | 2,2 | 2,6 | 8,1 |
| $x_i$ | 20,4 | 30,6 | 32,0 | 34,0 | 33,4 | 35,2 | 37,3 | 39,7 | 35,8 | 49,5 |
| *i* | *11* | *12* | *13* | *14* | *15* | *16* | *17* | *18* | *19* | *20* |
| $y_i$ | 8,3 | 8,5 | 8,8 | 9,6 | 10,5 | 11,6 | 22,1 | 29,7 | 34,2 | 49,6 |
| $x_i$ | 51,6 | 52,1 | 53,3 | 54,1 | 54,3 | 55,5 | 58,3 | 59,8 | 61,5 | 64,7 |

***Розв'язання***

1) *Визначення форми регресійної залежності між прибутком банків Y та величиною залучених коштів X.*

Ідентифікуємо змінні: *Y* – прибуток банків (залежна змінна); *X* – величина залучених коштів (незалежна змінна).

Побудуємо графік залежності (рис. 3.7) для вибіркових даних. На основі візуального аналізу розглядуваної залежності можна висунути припущення про поліноміальний степені $p = 3$ характер зв'язку між факторами: $\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3$.

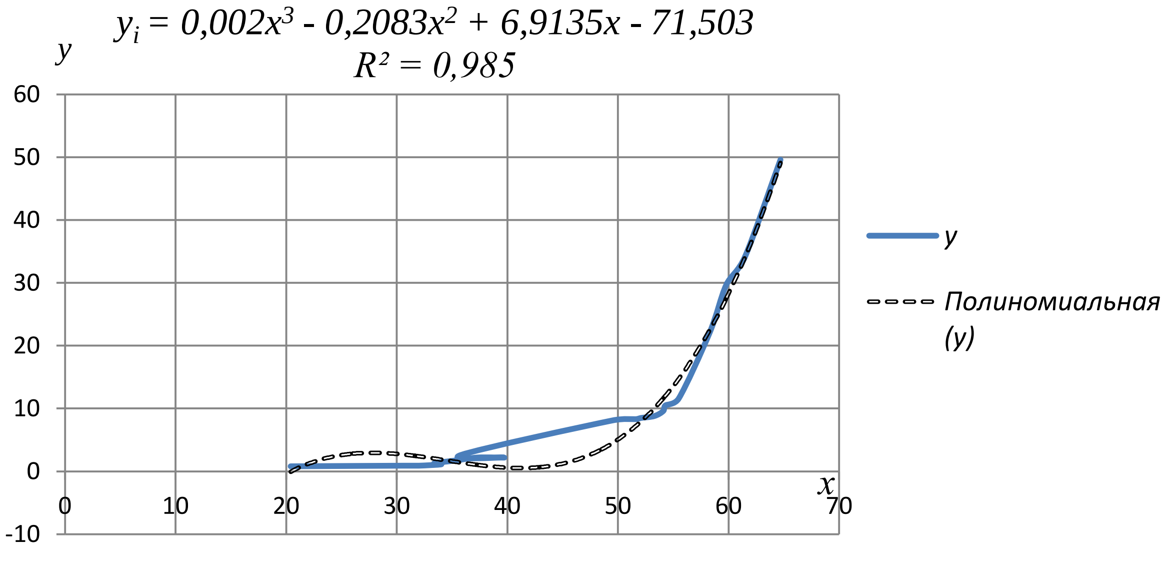



Рис. 3.7

2) *Знаходження статистичних оцінок для параметрів визначеної регресійної залежності*

Статистичні оцінки $b_0$, $b_1$, $b_2$, $b_3$ параметрів $\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$ для шуканого рівняння регресії можна знайти, якщо використати формулу (3.5) (розрахунки провести самостійно). Отримаємо:

$b_0 = -71{,}503$; $b_1 = 6{,}914$; $b_2 = -0{,}208$; $b_3 = 0{,}002$.

Регресійне рівняння залежності між прибутком банків від величини залучених коштів буде мати такий вигляд:

$$\hat{y}_i = -71{,}503 + 6{,}914\, x_i - 0{,}208\, x_i^2 + 0{,}002\, x_i^3.$$

Використовуючи засоби *Excel* («Диаграмма» → «Добавить линию тренда» → Вид «Полиномиальная» степени 3), можна також отримати рівняння регресії та її графік (теоретичну лінію регресії). Графік залежності між результативним та пояснювальним факторами, побудований за розрахованою моделлю, поданий на рис. 3.7.

3) *Перевірка адекватності моделі за коефіцієнтом детермінації та індексом кореляції.*

Коефіцієнт детермінації і індекс кореляції відповідно дорівнюють: $R^2 = 0{,}985$, $i_r = 0{,}992$.

Коефіцієнт детермінації показує, що 98,5% загальної зміни прибутків банків пояснюється зміною величини залучених коштів, у той час, як на інші фактори доводиться лише 1,5%. Коефіцієнт детермінації є мірою тісноти зв'язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже тісним. Отримана економетрична модель є адекватною.

Перевіримо значущість індекса кореляції $i_r$. Сформулюємо нульову гіпотезу $H_0 : i_r = 0$ при альтернативній гіпотезі $H_\alpha : i_r \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережуване значення статистичного $t$- критерію за формулою:

$$t = \sqrt{\frac{i_r^2}{1 - i_r^2} \cdot (n - m)} = \sqrt{\frac{0{,}985}{1 - 0{,}985} \cdot (20 - 4)} = 32{,}435$$

($m$ – число параметрів регресійного рівняння).

За таблицею 2 Додатку знаходимо

$$t''_{kp}(\alpha/2\,;k) = t''_{kp}(0{,}25;16) = 2{,}12$$

при рівні значущості $\alpha$ = 0,05 та степенях вільності $k = n - m = 16$ .

$$t'_{kp}(0{,}25;16) = -t''_{kp}(0{,}25;16) = 2{,}12.$$

*Висновок*. Оскільки $t \notin \left[t'_{kp}(\alpha/2\,;k);\ t''_{kp}(\alpha/2\,;k)\right]$, тобто $t \notin [-2{,}12;\ 2{,}12]$, то нульова гіпотеза $H_0$ відхиляється, що доводить *суттєвість зв'язку між залежною та пояснювальною змінними*.

4) *Статистична значущість параметрів моделі.*

Середньоквадратичні відхилення оцінок параметрів дорівнюють

$$S_{b_0} = 13{,}449;\ S_{b_1} = 1{,}0103;\ S_{b_2} = 0{,}0245;\ S_{b_3} = 0{,}00019.$$

Перевіримо статистичну значущість параметрів рівняння регресії $\beta_j$ $(j = \overline{0{,}3})$ аналогічно перевірці статистичної значущості $i_r$. Сформулюємо нульову гіпотезу: $H_0$: $\beta_j = 0$ при альтернативній гіпотезі $H_\alpha$: $\beta_j \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережувані значення даного критерію за формулою:

$$t_{b_j} = (b_j - 0)/S_{b_j}\ ,\qquad j = \overline{0{,}3}.$$

Отримаємо:

$$t_{b_0} = \frac{-71{,}503 - 0}{13{,}449} = -5{,}317;\quad t_{b_1} = \frac{6{,}914 - 0}{1{,}0103} = 6{,}843;$$

$$t_{b_2} = \frac{-0{,}208 - 0}{0{,}0245} = -8{,}510;\quad t_{b_3} = \frac{0{,}002 - 0}{0{,}00019} = 10{,}536.$$

*Висновок.* Оскільки $t''_{kp}(\alpha/2\,;k) = t''_{kp}(0{,}025;16) = 2{,}120$, то внаслідок того, що $t_{b_j} \notin [-2{,}120;2{,}120]$ $(j = \overline{0{,}3})$, нульова гіпотеза $H_0$ про рівність нулю параметрів $\beta_j$ $(j = \overline{0{,}3})$ *відхиляється*.

5) *Обчислення коефіцієнта еластичності при $x_i = 55{,}5$.*

$$\mathrm{E}_{55{,}5} = \left.\frac{x_i \cdot \hat{y}\,'}{\hat{y}}\right|_{x_i=55{,}5} = \left.\frac{x_i \cdot (6{,}914 - 2 \cdot 0{,}208 \cdot x_i + 3 \cdot 0{,}002 \cdot x_i^2)}{-71{,}503 + 6{,}914x_i - 0{,}208x_i^2 + 0{,}002x_i^3}\right|_{x_i=55{,}5} \approx 9{,}53.$$

*Висновок*. Якщо величина залучених коштів збільшиться на 1%, прибуток банку у середньому збільшиться наближено на 9,53%.

**Приклад 3.4**. Досліджується залежність між зміною заробітної плати $Y$ (%) та рівнем безробіття $X$ (%) на основі вибіркових даних, що наведені в таблиці 3.2

Таблиця 3.2

| $i$ | $y_i$ | $x_i$ | $1/x_i$ | $i$ | $y_i$ | $x_i$ | $1/x_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 0,6 | 8,4 | 0,119 | 16 | 4,1 | 3,9 | 0,256 |
| 2 | 0,8 | 8,0 | 0,125 | 17 | 4,4 | 3,7 | 0,270 |
| 3 | 0,9 | 7,5 | 0,133 | 18 | 4,8 | 3,5 | 0,286 |
| 4 | 1,0 | 6,9 | 0,145 | 19 | 5,1 | 3,3 | 0,303 |
| 5 | 1,1 | 6,6 | 0,152 | 20 | 5,4 | 3,0 | 0,333 |
| 6 | 1,2 | 6,5 | 0,154 | 21 | 5,5 | 2,9 | 0,345 |
| 7 | 1,7 | 6,0 | 0,167 | 22 | 5,9 | 2,9 | 0,345 |
| 8 | 1,8 | 5,9 | 0,169 | 23 | 6,0 | 2,8 | 0,357 |
| 9 | 2,0 | 5,0 | 0,200 | 24 | 6,2 | 2,5 | 0,400 |
| 10 | 2,1 | 5,2 | 0,192 | 25 | 2,3 | 2,6 | 0,385 |
| 11 | 2,4 | 4,0 | 0,250 | 26 | 6,4 | 2,6 | 0,385 |
| 12 | 3,0 | 4,2 | 0,238 | 27 | 6,5 | 2,5 | 0,400 |
| 13 | 3,2 | 4,3 | 0,233 | 28 | 6,8 | 2,4 | 0,417 |
| 14 | 3,6 | 4,4 | 0,227 | 29 | 7,0 | 2,3 | 0,435 |
| 15 | 4,0 | 3,8 | 0,263 | 30 | 7,1 | 2,4 | 0,417 |

1) Визначити форму регресійної залежності між зміною заробітної плати $Y$ та рівнем безробіття $X$.
2) Знайти статистичні оцінки для параметрів визначеної регресійної залежності.
3) Перевірити адекватність моделі за коефіцієнтом детермінації та індексом кореляції.
4) Перевірити статистичну значущість параметрів моделі.
5) Обчислити коефіцієнт еластичності при $x_i = 4$ (%).

***Розв'язання***

1) *Визначення форми регресійної залежності між змінною заробітної плати та рівнем безробіття*

Ідентифікуємо змінні: $Y$ – заробітна плата (залежна змінна); $X$ – рівень безробіття (незалежна змінна).

Побудуємо графік залежності (рис. 3.8) для вибіркових даних. На основі візуального аналізу розглядуваної залежності можна висунути припущення про гіперболічний характер зв'язку між факторами:

$$\hat{y} = b_0 + b_1 \cdot \frac{1}{x}.$$

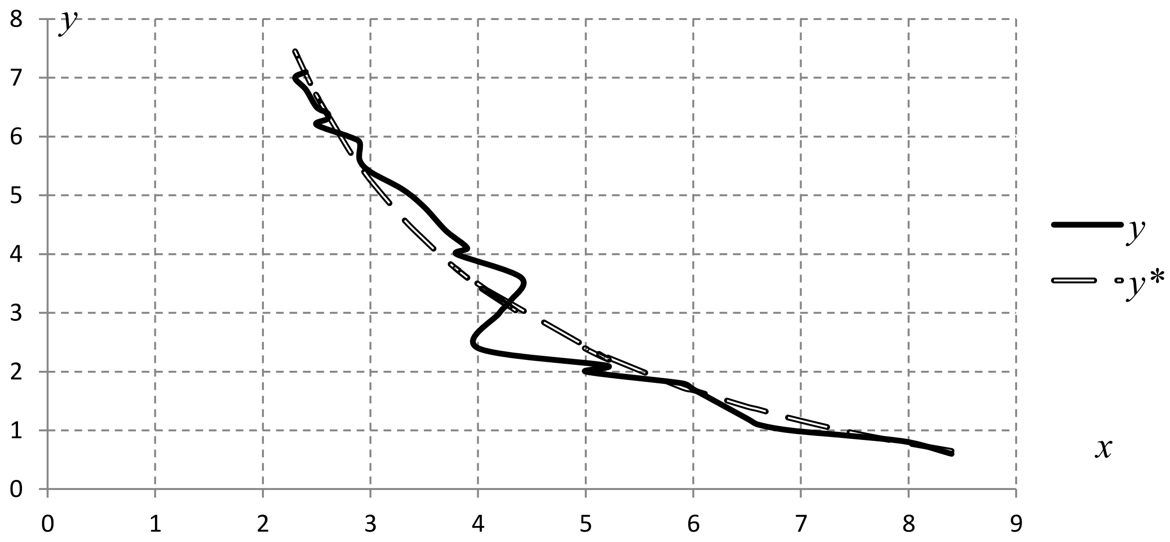



Рис. 3.8

2) *Знаходження статистичних оцінок для параметрів визначеної регресійної залежності*.

Статистичні оцінки $b_0$, $b_1$ параметрів $\beta_0$, $\beta_1$ для шуканого рівняння регресії можна знайти, якщо використати формулу (3.9) (розрахунки провести самостійно). Отримаємо:

$$b_0 = -1{,}911; \quad b_1 = 21{,}521.$$

Регресійне рівняння залежності між зміною заробітної плати та рівнем безробіттяв буде мати такий вигляд:

$$\hat{y}_i = -1{,}911 + 21{,}521 \cdot \frac{1}{x}.$$

Графік залежності між результативним та пояснювальним факторами, побудований за розрахованою моделлю, поданий на рис. 3.8.

3) *Перевірка адекватності моделі за коефіцієнтом детермінації та індексом кореляції*.

Коефіцієнт детермінації і індекс кореляції відповідно дорівнюють: $R^2 = 0{,}973$, $i_r = 0{,}986$.

Коефіцієнт детермінації показує, що 97,3% загальної зміни заробітної плати пояснюється зміною рівня безробіття, у той час, як на інші фактори доводиться лише 2,7%. Коефіцієнт детермінації є мірою тісноти зв'язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже тісним. Отримана економетрична модель є адекватною.

Перевіримо значущість індекса кореляції $i_r$. Сформулюємо нульову гіпотезу $H_0 : i_r = 0$ при альтернативній гіпотезі $H_\alpha : i_r \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережуване значення статистичного $t$- критерію за формулою:

$$t = \sqrt{\frac{i_r^2}{1 - i_r^2} \cdot (n - m)} = \sqrt{\frac{0{,}973}{1 - 0{,}973} \cdot (30 - 2)} = 31{,}802.$$

За таблицею 2 Додатку знаходимо

$$t''_{kp}(\alpha/2\,; k) = t''_{kp}(0{,}25; 28) = 2{,}048$$

при рівні значущості $\alpha = 0{,}05$ та степенях вільності $k = n - m = 28$ .

$$t'_{kp}(\alpha/2\,; k) = -t''_{kp}(\alpha/2\,; k) = -2{,}048.$$

*Висновок*. Оскільки $t \notin \left[t'_{kp}(\alpha/2\,; k);\ t''_{kp}(\alpha/2\,; k)\right]$, тобто $t \notin [-2{,}048;\ 2{,}048]$, то нульова гіпотеза $H_0$ відхиляється, що доводить *суттєвість зв'язку між залежною та пояснювальною змінними*.

4) *Статистична значущість параметрів моделі.*

Середньоквадратичні відхилення оцінок параметрів дорівнюють

$$S_{b_0} = 0{,}195;\ \ S_{b_1} = 0{,}677.$$

Перевіримо статистичну значущість параметрів рівняння регресії $\beta_j$ $(j = \overline{0{,}1})$ аналогічно перевірці статистичної значущості $i_r$. Сформулюємо нульову гіпотезу: $H_0$: $\beta_j = 0$ при альтернативній гіпотезі $H_\alpha$: $\beta_j \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережувані значення даного критерію за формулою:

$$t_{b_j} = (b_j - 0)/S_{b_j}\ , \qquad j = \overline{0,1}.$$

Отримаємо:

$$t_{b_0} = -1{,}911/0{,}195 = -9{,}813; \quad t_{b_1} = 21{,}521/0{,}677 = 31{,}802.$$

*Висновок.* Оскільки $t''_{kp}(\alpha/2\,; k) = t''_{kp}(0{,}025; 28) = 2{,}048,$ то внаслідок того, що $t_{b_j} \notin [-2{,}048; 2{,}048]$ $(j = \overline{0,1})$, нульова гіпотеза $H_0$ про рівність нулю параметрів $\beta_j$ $(j = \overline{0,1})$ *відхиляється*.

*5) Обчислення коефіцієнта еластичності при* $x_i = 4$**.**

$$\mathrm{E}_4 = \frac{x_i \cdot \hat{y}'}{\hat{y}}\bigg|_{x_i=4} = -\frac{21{,}521}{x_i\hat{y}}\bigg|_{x_i=4} = \frac{-21{,}521}{x_i(-1{,}911 + 21{,}521/x_i)}\bigg|_{x_i=4} = -1{,}55.$$

*Висновок*. При збільшенні рівня безробіття на 1% заробітна плата у середньому зменшиться наближено на 1,55%.

**Приклад 3.5**. Досліджується залежність між попитом на товари першої необхідності *Y* (грн.) і доходом *X* (грн.) на основі вибіркових даних, що наведені у таблиці 3.3.

Таблиця 3.3

| $i$ | $y_i$ | $x_i$ | $1/x_i$ | $i$ | $y_i$ | $x_i$ | $1/x_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 14,7 | 15,5 | 0,065 | 16 | 13,4 | 3,4 | 0,294 |
| 2 | 14,5 | 14,5 | 0,069 | 17 | 13,2 | 3,2 | 0,313 |
| 3 | 14,4 | 13,9 | 0,072 | 18 | 12,9 | 3,0 | 0,333 |
| 4 | 14,3 | 11,3 | 0,088 | 19 | 12,8 | 2,8 | 0,357 |
| 5 | 14,3 | 10,8 | 0,093 | 20 | 12,6 | 2,5 | 0,400 |
| 6 | 14,3 | 9,6 | 0,104 | 21 | 12,3 | 2,4 | 0,417 |
| 7 | 14,2 | 8,4 | 0,119 | 22 | 12,2 | 2,4 | 0,417 |
| 8 | 14,2 | 7,1 | 0,141 | 23 | 12,2 | 2,3 | 0,435 |
| 9 | 14,2 | 6,9 | 0,145 | 24 | 12,0 | 2,0 | 0,500 |
| 10 | 14,1 | 6,7 | 0,149 | 25 | 12,0 | 2,1 | 0,476 |
| 11 | 14,0 | 6,6 | 0,152 | 26 | 12,0 | 2,1 | 0,476 |
| 12 | 14,0 | 6,4 | 0,156 | 27 | 11,9 | 2,0 | 0,500 |
| 13 | 13,8 | 5,5 | 0,182 | 28 | 11,8 | 1,9 | 0,526 |
| 14 | 13,5 | 3,9 | 0,256 | 29 | 11,5 | 1,8 | 0,556 |
| 15 | 13,2 | 3,3 | 0,303 | 30 | 11,5 | 1,9 | 0,526 |

1) Визначити форму регресійної залежності між попитом на товари першої необхідності $Y$ і доходом $X$.
2) Знайти статистичні оцінки для параметрів визначеної регресійної залежності.
3) Перевірити адекватність моделі за коефіцієнтом детермінації та індексом кореляції.
4) Перевірити статистичну значущість параметрів моделі.
5) Обчислити коефіцієнт еластичності при $x_i = 3$ (грн.).

***Розв'язання***

1) *Визначення форми регресійної залежності між попитом на товари першої необхідності і доходом.*

Ідентифікуємо змінні: $Y$ – попит на товари першої необхідності (залежна змінна); $X$ – доход (незалежна змінна).

Побудуємо графік залежності (рис. 3.9) для вибіркових даних. На основі візуального аналізу розглядуваної залежності можна висунути припущення про гіперболічний характер зв'язку між факторами:

$$\hat{y} = b_0 + b_1 \cdot \frac{1}{x}.$$

2) *Знаходження статистичних оцінок для параметрів визначеної регресійної залежності.*

Статистичні оцінки $b_0$, $b_1$ параметрів $\beta_0$, $\beta_1$ для шуканого рівняння регресії можна знайти, якщо використати формулу (3.9) (розрахунки провести самостійно). Отримаємо:

$$b_0 = 14{,}984; \;\; b_1 = -\,6{,}208.$$

Регресійне рівняння залежності між попитом на товари першої необхідності і доходом буде мати такий вигляд:

$$\hat{y}_i = 14{,}984 - 6{,}208 \cdot \frac{1}{x}.$$

Графік залежності між результативним та пояснювальним факторами, побудований за розрахованою моделлю, поданий на рис. 3.9.

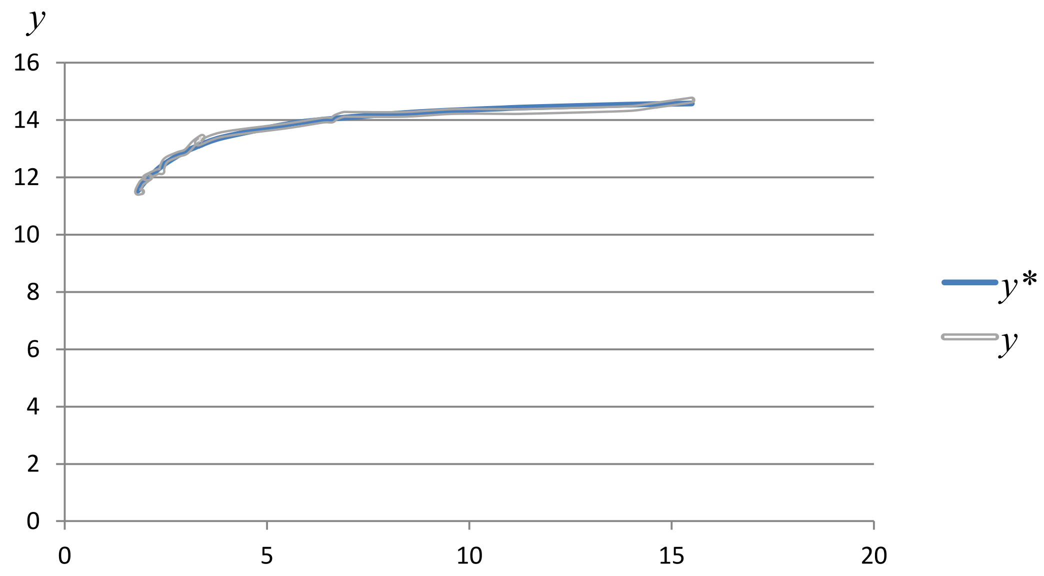



Рис. 3.9

3) *Перевірка адекватності моделі за коефіцієнтом детермінації та індексом кореляції.*

Коефіцієнт детермінації і індекс кореляції відповідно дорівнюють: $R^2 = 0{,}989$, $i_r = 0{,}995$.

Коефіцієнт детермінації показує, що 98,9% загальної зміни попиту на товари першої необхідності пояснюється зміною доходів, у той час, як на інші фактори доводиться лише 1,1%. Коефіцієнт детермінації є мірою тісноти зв’язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже тісним. Отримана економетрична модель є адекватною.

Перевіримо значущість індекса кореляції $i_r$. Сформулюємо нульову гіпотезу $H_0 : i_r = 0$ при альтернативній гіпотезі $H_\alpha : i_r \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережуване значення статистичного $t$- критерію за формулою:

$$t = \sqrt{\frac{i_r^2}{1 - i_r^2} \cdot (n - m)} = \sqrt{\frac{0{,}989}{1 - 0{,}989} \cdot (30 - 2)} = 50{,}893.$$

За таблицею 2 Додатку знаходимо

$$t''_{кр}(\alpha/2\,; k) = t''_{кр}(0{,}25;\ k) = 2{,}048$$

при рівні значущості $\alpha$ = 0,05 та степенях вільності $k = n - m = 28$.

$$t'_{kp}(\alpha/2\,;k) = -t''_{kp}(\alpha/2\,;k) = -2{,}048.$$

*Висновок*. Оскільки $t \notin [t'_{kp}(\alpha/2\,;k);\ t''_{kp}(\alpha/2\,;k)]$, тобто $t \notin [-2{,}048;\ 2{,}048]$, то нульова гіпотеза $H_0$ відхиляється, що доводить *суттєвість зв'язку між залежною та пояснювальною змінними*.

4) *Статистична значущість параметрів моделі.*

Середньоквадратичні відхилення оцінок параметрів дорівнюють

$$S_{b_0} = 0{,}040;\ S_{b_1} = 0{,}122.$$

Перевіримо статистичну значущість параметрів рівняння регресії $\beta_j$ $(j = \overline{0,1})$ аналогічно перевірці статистичної значущості $i_r$. Сформулюємо нульову гіпотезу: $H_0$: $\beta_j = 0$ при альтернативній гіпотезі $H_\alpha$: $\beta_j \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережувані значення даного критерію за формулою:

$$t_{b_j} = (b_j - 0)/S_{b_j}\ , \quad j = \overline{0,1}.$$

Отримаємо:

$$t_{b_0} = 14{,}984/0{,}040 = 370{,}829;\ \ t_{b_1} = -6{,}208/0{,}122 = -50{,}893.$$

*Висновок.* Оскільки $t''_{kp}(\alpha/2\,;k) = t''_{kp}(0{,}025; 28) = 2{,}048$, то внаслідок того, що $t_{b_j} \notin [-2{,}048; 2{,}048]$ $(j = \overline{0,1})$, нульова гіпотеза $H_0$ про рівність нулю параметрів $\beta_j$ $(j = \overline{0,1})$ *відхиляється*

5) *Обчислення коефіцієнта еластичності при* $x_i = 3$**.**

$$\mathrm{E}_4 = \frac{x_i \cdot \hat{y}'}{\hat{y}}\bigg|_{x_i=3} = -\frac{-6{,}208}{x_i\hat{y}}\bigg|_{x_i=3} = \frac{6{,}208}{x_i(14{,}984 - 6{,}208/x_i)}\bigg|_{x_i=3} = 0{,}16.$$

*Висновок*. При збільшенні доходу на 1% попит на товари першої необхідності у середньому збільшиться наближено на 0,16 %.

**Приклад 3.6.** Досліджується залежність між фіксованими витратами на одиницю продукції $Y$ (грн.) та кількісними обсягами виробництва продукції $X$ (т) на основі вибіркових даних, що наведені в таблиці 3.4.

Таблиця 3.4

| $i$ | $y_i$ | $x_i$ | $1/x_i$ | $i$ | $y_i$ | $x_i$ | $1/x_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 2,40 | 165 | 0,0061 | 16 | 2,21 | 550 | 0,0018 |
| 2 | 2,38 | 200 | 0,0050 | 17 | 2,21 | 575 | 0,0017 |
| 3 | 2,35 | 250 | 0,0040 | 18 | 2,21 | 600 | 0,0017 |
| 4 | 2,33 | 275 | 0,0036 | 19 | 2,21 | 650 | 0,0015 |
| 5 | 2,30 | 300 | 0,0033 | 20 | 2,21 | 700 | 0,0014 |
| 6 | 2,30 | 325 | 0,0031 | 21 | 2,20 | 750 | 0,0013 |
| 7 | 2,28 | 350 | 0,0029 | 22 | 2,20 | 800 | 0,0013 |
| 8 | 2,25 | 360 | 0,0028 | 23 | 2,20 | 850 | 0,0012 |
| 9 | 2,24 | 375 | 0,0027 | 24 | 2,20 | 900 | 0,0011 |
| 10 | 2,24 | 400 | 0,0025 | 25 | 2,18 | 950 | 0,0011 |
| 11 | 2,24 | 425 | 0,0024 | 26 | 2,18 | 1000 | 0,0010 |
| 12 | 2,22 | 450 | 0,0022 | 27 | 2,18 | 1050 | 0,0010 |
| 13 | 2,22 | 475 | 0,0021 | 28 | 2,18 | 1100 | 0,0009 |
| 14 | 2,22 | 500 | 0,0020 | 29 | 2,17 | 1150 | 0,0009 |
| 15 | 2,22 | 525 | 0,0019 | 30 | 2,17 | 1200 | 0,0008 |

1) Визначити форму регресійної залежності між попитом на товари першої необхідності $Y$ і доходом $X$.
2) Знайти статистичні оцінки для параметрів визначеної регресійної залежності.
3) Перевірити адекватність моделі за коефіцієнтом детермінації та індексом кореляції.
4) Перевірити статистичну значущість параметрів моделі.
5) Обчислити коефіцієнт еластичності при $x_i = 1000$ (грн.).

***Розв'язання***

1) *Визначення форми регресійної залежності між фіксованими витратами на одиницю продукції та кількісними обсягами виробництва продукції.*

Ідентифікуємо змінні: $Y$ – фіксовані витрати на одиницю продукції (залежна змінна); $X$ – обсяг виробництва продукції (незалежна змінна).

Побудуємо графік залежності (рис. 3.10) для вибіркових даних. На основі візуального аналізу розглядуваної залежності можна висунути припущення про гіперболічний характер зв'язку між факторами: $\hat{y} = b_0 + b_1 \cdot \dfrac{1}{x}$.

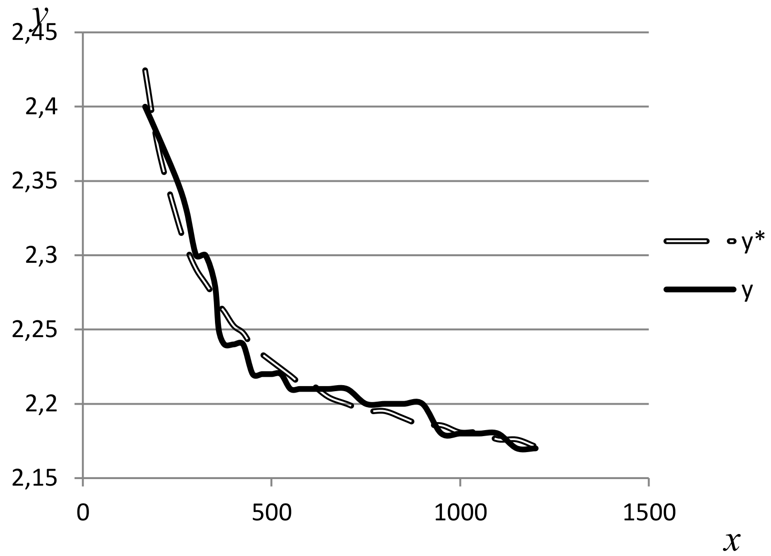



Рис. 3.10.

2) *Знаходження статистичних оцінок для параметрів визначеної регресійної залежності.*

Статистичні оцінки $b_0$, $b_1$ параметрів $\beta_0$, $\beta_1$ для шуканого рівняння регресії можна знайти, якщо використати формулу (3.9) (розрахунки провести самостійно). Отримаємо:

$$b_0 = 2{,}133; \quad b_1 = 47{,}786.$$

Регресійне рівняння залежності між фіксованими витратами на одиницю продукції та кількісними обсягами виробництва продукції:

$$\hat{y}_i = 2{,}133 + 47{,}786 \cdot \frac{1}{x}.$$

Графік залежності між результативним та пояснювальним факторами, побудований за розрахованою моделлю, поданий на рис. 3.9.

3) *Перевірка адекватності моделі за коефіцієнтом детермінації та індексом кореляції.*

Коефіцієнт детермінації і індекс кореляції відповідно дорівнюють: $R^2 = 0{,}959$, $i_r = 0{,}979$.

Коефіцієнт детермінації показує, що 95,9% загальної зміни між фіксованими витратами на одиницю продукції пояснюється зміною кількісних обсягів виробництва продукції, у той час, як на інші фактори доводиться лише 4,1%. Коефіцієнт детермінації є мірою тісноти зв’язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже тісним. Отримана економетрична модель є адекватною.

Перевіримо значущість індекса кореляції $i_r$. Сформулюємо нульову гіпотезу $H_0 : i_r = 0$ при альтернативній гіпотезі $H_\alpha : i_r \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережуване значення статистичного $t$- критерію за формулою:

$$t = \sqrt{\frac{i_r^2}{1 - i_r^2} \cdot (n - m)} = \sqrt{\frac{0{,}959}{1 - 0{,}959} \cdot (30 - 2)} = 25{,}692.$$

За таблицею 2 Додатку знаходимо

$$t''_{kp}(\alpha/2\,; k) = t''_{kp}(0{,}25;\ k) = 2{,}048$$

при рівні значущості $\alpha = 0{,}05$ та степенях вільності $k = n - m = 28.$

$$t'_{kp}(\alpha/2\,; k) = -t''_{kp}(\alpha/2\,; k) = -2{,}048.$$

*Висновок*. Оскільки $t \notin \left[t'_{kp}(\alpha/2\,; k);\ t''_{kp}(\alpha/2\,; k)\right]$, тобто $t \notin [-2{,}048;\ 2{,}048]$, то нульова гіпотеза $H_0$ відхиляється, що доводить *суттєвість зв'язку між залежною та пояснювальною змінними*.

4) *Статистична значущість параметрів моделі.*

Середньоквадратичні відхилення оцінок параметрів дорівнюють

$$S_{b_0} = 0{,}00466;\ S_{b_1} = 1{,}860.$$

Перевіримо статистичну значущість параметрів рівняння регресії $\beta_j$ $(j = \overline{0{,}1})$ аналогічно перевірці статистичної значущості $i_r$. Сформулюємо нульову гіпотезу: $H_0$: $\beta_j = 0$ при альтернативній гіпотезі $H_\alpha$: $\beta_j \neq 0$ та перевіримо її за $t$-критерієм Ст'юдента. Обчислимо спостережувані значення даного критерію за формулою:

$$t_{b_j} = (b_j - 0)/S_{b_j}\ , \quad j = \overline{0{,}1}.$$

Отримаємо:

$$t_{b_0} = 2{,}133/0{,}00466 = 457{,}982; \quad t_{b_1} = 47{,}786/1{,}860 = 25{,}692.$$

*Висновок.* Оскільки $t''_{kp}(\alpha/2\,; k) = t''_{kp}(0{,}025; 28) = 2{,}12$, то внаслідок того, що $t_{b_j} \notin [-2{,}12; 2{,}12]$ $(j = \overline{0{,}1})$, нульова гіпотеза $H_0$ про рівність нулю параметрів $\beta_j$ $(j = \overline{0{,}1})$ *відхиляється*

5) *Обчислення коефіцієнта еластичності при* $x_i = 1000$**.**

$$\text{E}_4 = \frac{x_i \cdot \hat{y}'}{\hat{y}}\bigg|_{x_i=1000} = -\frac{47{,}786}{x_i\hat{y}}\bigg|_{x_i=1000} = -\frac{47{,}786}{x_i(2{,}133 + 47{,}786/x_i)}\bigg|_{x_i=1000} \approx -0{,}02.$$

*Висновок*. При збільшенні кількісних обсягів виробництва продукції на 1% фіксовані витрати на одиницю продукції у середньому зменшаться наближено на 0,02 %.

**Приклад 3.7**. Досліджується залежність між часткою витрат на товари тривалого використання *Y* (% загальної суми витрат) та прибутком родини *X* (тис. грош. од.) на основі вибіркових даних, що наведені в таблиці 3.5.

Таблиця 3.5

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $y_i$ | 10 | 13,4 | 15,4 | 16,5 | 18,6 | 19,1 |
| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |

1) Визначити форму регресійної залежності між часткою витрат на товари тривалого використання *Y* та прибутком родини *X*.
2) Знайти статистичні оцінки для параметрів визначеної регресійної залежності.
3) Перевірити адекватність моделі за коефіцієнтом детермінації.

***Розв'язання***

1) *Визначення форми залежності між часткою витрат на товари тривалого використання Y та прибутком родини X.*

Ідентифікуємо змінні: *Y* – частка витрат на товари тривалого використання (залежна змінна); *X* – прибуток родини (незалежна змінна).

Побудуємо графік залежності (рис. 3.11) для вибіркових даних. На основі візуального аналізу розглядуваної залежності можна висунути припущення про логарифмічний характер зв'язку між факторами:

$$\hat{y} = b_0 + b_1 \cdot \ln x.$$

2) *Знаходження статистичних оцінок для параметрів визначеної регресійної залежності.*

Знайдемо статистичні оцінки $b_0$, $b_1$ параметрів $\beta_0$, $\beta_1$ для шуканого рівняння регресії. Зробимо заміну $z = \ln x$. Для знаходження параметрів рівняння регресії необхідно розв'язати таку систему нормальних рівнянь:

$$\begin{cases} b_0 + b_1 \cdot \bar{z} = \bar{y} \\ b_0\, \bar{z} + b_1 \cdot \overline{z^2} = \overline{zy}, \end{cases}$$

де

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} \ln x_i\,;\ \overline{z^2} = \frac{1}{n}\sum_{i=1}^{n} {z_i}^2 = \frac{1}{n}\sum_{i=1}^{n} (\ln x_i)^2,$$

$$\overline{zy} = \frac{1}{n}\sum_{i=1}^{n} z_i y_i = \frac{1}{n}\sum_{i=1}^{n} (\ln x_i) y_i.$$

Виконуючи необхідні обчислення, отримаємо таку систему рівнянь:

$$\begin{cases} b_0 + \dfrac{6{,}5792}{6} b_1 = 15{,}5 \\ \dfrac{6{,}5792}{6} \cdot b_0 + \dfrac{9{,}4099}{6} \cdot b_1 = 18{,}876. \end{cases}$$

Розв'язок системи: $b_0 = 9{,}876$, $b_1 = 5{,}129$.

Регресійне рівняння залежності між часткою витрат на товари тривалого використання та прибутком родини має такий вигляд:

$$\hat{y} = 9{,}876 + 5{,}129 \ln x.$$

Графік залежності між результативним та пояснювальним факторами, побудований за розрахованою моделлю, поданий на рис. 3.11 (значення $\hat{y}_i$ зображені трикутними маркерами, вихідні дані - квадратними).

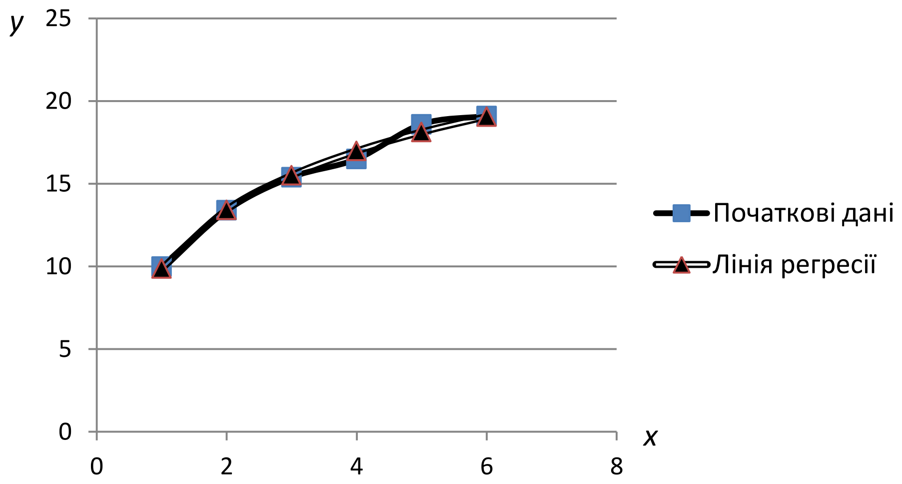



Рис. 3.11

3) *Перевірка адекватності моделі за коефіцієнтом детермінації.*

Коефіцієнт детермінації і індекс кореляції відповідно дорівнюють: $R^2 = 0{,}959$, $i_r = 0{,}979$.

Коефіцієнт детермінації показує, що 95,9% загальної зміни між часткою витрат на товари тривалого використання пояснюється зміною прибутку родини, у той час, як на інші фактори доводиться лише 4,1%. Коефіцієнт детермінації є мірою тісноти зв’язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже тісним. Отримана економетрична модель є адекватною.

### 3.3.1. Зразок виконання контрольного завдання

***Завдання*** [26]. Досліджується залежність між витратами на купівлю продовольчих товарів $Y$ (% від загальних витрат) та середньоденною заробітною платою одного робітника $X$ (грош.од.) на основі вибіркових даних семи районів деякої області, що наведені у таблиці 3.6.

Таблиця 3.6

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $y_i$ | 68,8 | 61,2 | 59,9 | 56,7 | 55 | 54,3 | 49,3 |
| $x_i$ | 45,1 | 59,2 | 57,2 | 61,8 | 58,8 | 47,2 | 55,2 |

Для характеристики залежності між витратами на купівлю продовольчих товарів $Y$ та середньоденною заробітною платою одного робітника $X$ знайти рівняння регресії в припущенні таких залежностей:

1) лінійної;
2) степеневої;
3) показникової;
4) рівнобічної гіперболи;
5) поліноміальної третьої степені.

Оцінити адекватність побудованих моделей.

***Розв'язання***

Ідентифікуємо змінні: $Y$ – витрати на купівлю продовольчих товарів (залежна змінна); $X$ – середньоденна заробітна плата одного робітника (незалежна змінна).

1) Нехай економетрична модель має *лінійну* форму:

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

де $\beta_0$ і $\beta_1$ – параметри моделі, $\varepsilon$ - залишок.

Оцінимо параметри моделі $\hat{y} = b_0 + b_1 x$ за допомогою МНК.

Для знаходження параметрів рівняння регресії $b_0$ та $b_1$ необхідно розв'язати таку систему нормальних рівнянь:

$$\begin{cases} nb_0 + b_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i \\ b_0 \sum\limits_{i=1}^{n} x_i + b_1 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i \end{cases}$$

Тут $n = 7$ – кількість досліджуваних районів. Вихідні дані та їх перетворення для побудови моделі наведені у таблиці 3.7.

Результати обчислень:

$$b_1 = \frac{cov(x,y)}{var(x)} = \frac{-11,88}{34,33} \approx -0,35;$$

$$b_0 = \overline{y} - b_1 \overline{x} \approx 57,89 + 0,35 \cdot 54,90 \approx 76,88.$$

Таблиця 3.7

| $i$ | $x$ | $y^e$ | $(x-\bar{x})^2$ | $(y-\bar{y}^{\,e})^2$ | $(x-\bar{x})\cdot(y-\bar{y}^{\,e})$ | $\hat{y}$ | *MAPE* |
|---|---|---|---|---|---|---|---|
| 1 | 45,10 | 68,80 | 96,04 | 119,12 | -106,96 | 61,28 | 10,94 |
| 2 | 59,00 | 61,20 | 16,81 | 10,98 | 13,59 | 56,47 | 7,73 |
| 3 | 57,20 | 59,90 | 5,29 | 4,06 | 4,63 | 57,09 | 4,69 |
| 4 | 61,80 | 56,70 | 47,61 | 1,41 | -8,18 | 55,50 | 2,12 |
| 5 | 58,80 | 55,00 | 15,21 | 8,33 | -11,25 | 56,54 | 2,79 |
| 6 | 47,20 | 54,30 | 59,29 | 12,86 | 27,61 | 60,55 | 11,51 |
| 7 | 55,20 | 49,30 | 0,09 | 73,71 | -2,58 | 57,78 | 17,20 |
| **Сума** | 384,30 | 405,20 | 240,34 | 230,47 | -83,14 | 405,20 | 56,99 |
| **Сер.зн** | 54,90 | 57,89 | **34,33** | **32,92** | **-11,88** | 57,89 | **8,14** |

Рівняння шуканої парної лінійної регресії має такий вигляд:

$$\hat{y} = 76{,}88 - 0{,}35\,x.$$

При збільшенні середньоденної заробітної плати одного робітника на 1 грош. од. частка витрат на купівлю продовольчих товарів в загальних витратах знижується у середньому на 0,35% .

Обчислимо лінійний коефіцієнт парної кореляції:

$$r = \frac{cov(x,y)}{\sqrt{var(x)\cdot var(y)}} = \frac{-11{,}88}{\sqrt{34{,}33\cdot 32{,}92}} \approx -0{,}353.$$

Регресійний зв'язок дуже слабкий, обернений.

У випадку парної лінійної регресійної моделі коефіцієнт детермінації дорівнює квадрату лінійного коефіцієнта кореляції:

$$R^2 = r^2 \approx 0{,}1248.$$

Коефіцієнт детермінації показує, що 12,48% загальної зміни частки витрат на купівлю продовольчих товарів в загальних витратах пояснюється зміною середньоденної заробітної плати одного робітника, у той час, як на інші фактори доводиться 87,52%.

Коефіцієнт детермінації є мірою тісноти зв’язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже слабкий. Отримана економетрична модель не є адекватною.

Середня похибка апроксимації

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}}{y_i}\right| \cdot 100\% \approx \frac{1}{7} \cdot 56{,}7\% \approx 8{,}14\%$$

показує, що у середньому розрахункові значення відхиляються від фактичних на 8,14%.

Обчислимо фактичне значення статистики $F$-критерію Фішера для парної лінійної регресії за формулою:

$$F_{факт} = \frac{r^2}{1-r^2}(n-2) = \frac{0{,}1248}{1-0{,}1248}\quad(7-2) \approx 0{,}71.$$

При рівні значущості α = 0,05 знайдемо за таблицею 1 Додатку критичне значення

$$F_{кр}\,(\alpha;\, k_1;\, k_2) = F_{кр}\,(\alpha;\, m-1;\, n-m) = F_{кр}\,(0{,}05;\, 1;\, 5) = 6{,}61.$$

Оскільки $F_{факт} < F_{кр}$, то приймається гіпотеза про випадкову природу наявної залежності і статистичної незначущості параметрів рівняння і показника тісноти зв'язку. Побудована модель не є адекватною.

Результати регресійного аналізу наведені на рисунку 3.12.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ВЫВОД ИТОГОВ | | | | | | | | |
| | | | | | | | | |
| рессионная статистика | | | | | | | | |
| Множеств | 0,35325729 | | | | | | | |
| R-квадрат | 0,12479072 | | | | | | | |
| Нормиро | -0,0502511 | | | | | | | |
| Стандартн | 6,35150744 | | | | | | | |
| Наблюдени | 7 | | | | | | | |
| | | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | Значимость F | | | |
| Регрессия | 1 | 28,7603379 | 28,76033785 | 0,71291928 | 0,436999564 | | | |
| Остаток | 5 | 201,708234 | 40,34164671 | | | | | |
| Итого | 6 | 230,468571 | | | | | | |
| | | | | | | | | |
| | Коэффициен | дартная ош | -статистик | P-Значение | Нижние 95% | Верхние 95% | Нижние 95,0% | Верхние 95,0% |
| Y-пересеч | 76,8770848 | 22,6201661 | 3,398608321 | 0,01928015 | 18,73009667 | 135,024073 | 18,73009667 | 135,024073 |
| Перемен | -0,3459266 | 0,40969794 | -0,84434548 | 0,43699956 | -1,399088692 | 0,70723548 | -1,399088692 | 0,707235484 |

Рис.3.12

2) Нехай економетрична модель має *степеневу* форму:

$$y = \beta_0 \cdot x^{\beta 1} \cdot \varepsilon,$$

де $\beta_0$ і $\beta_1$ – параметри моделі, $\varepsilon$ - залишок.

Оцінимо параметри моделі $\hat{y} = b_0 \cdot x^{b1}$ за допомогою МНК. Для цього спочатку виконаємо процедуру *лінеаризації* змінних. Логарифмуємо обидві частини рівняння:

$$ln\, \hat{y} = ln\, b_0 + b_1 \cdot ln\, x.$$

Отримаємо лінійне рівняння парної регресії

$$Y = B_0 + b_1 X,$$

де $$\hat{Y} = ln\, \hat{y}, \;\; X = ln\, x; \;\; B_0 = ln\, b_0.$$

Вихідні дані та їх перетворення для побудови моделі наведені у таблиці 3.8.

Таблиця 3.8

| $i$ | $x$ | $y^e$ | $X=lnx$ | $Y=lny$ | $(X-\bar{X})^2$ | $(X-\bar{X})\cdot(Y-\bar{Y})$ | $\hat{y}$ |
|---|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 3,809 | 4,231 | 0,0363 | -0,0338 | 60,98 |
| 2 | 59,0 | 61,2 | 4,078 | 4,114 | 0,0061 | 0,0047 | 56,28 |
| 3 | 57,2 | 59,9 | 4,047 | 4,093 | 0,0022 | 0,0018 | 56,80 |
| 4 | 61,8 | 56,7 | 4,124 | 4,038 | 0,0155 | -0,0020 | 55,51 |
| 5 | 58,8 | 55,0 | 4,074 | 4,007 | 0,0056 | -0,0035 | 56,34 |
| 6 | 47,2 | 54,3 | 3,854 | 3,995 | 0,0211 | 0,0086 | 60,16 |
| 7 | 55,2 | 49,3 | 4,011 | 3,898 | 0,0001 | -0,0018 | 57,41 |
| **Σ** | 384,3 | 405,2 | 27,996 | 28,376 | 0,0869 | -0,0259 | 403,48 |
| **Сер. зн.** | 54,9 | 57,89 | 3,999 | 4,054 | **0,0124** | **-0,0037** | 57,64 |

Результати обчислень:

$$b_1 = \frac{cov(X,Y)}{var(X)} = \frac{-0,0037}{0,0124} \approx -0,298;$$

$$B_0 = \overline{Y} - b_1 \overline{X} \approx 4,054 + 0,298 \cdot 3,999 \approx 5,247; \quad b_0 = e^{B_0} \approx e^{5,247} \approx 190,033.$$

Лінеарізована модель має такий вигляд:

$$\hat{Y} = 5,247 \; - 0,298 \cdot X.$$

Рівняння шуканої степеневої регресії має такий вигляд:

$$\hat{y} = 190{,}033 \cdot x^{-0{,}298}.$$

Знайдемо показники тісноти зв'язку: індекс кореляції $i_r$ і середню похибку апроксимації *MAPE*. Необхідні обчислення наведені у таблиці 3.9.

Таблиця 3.9

| $i$ | $x$ | $y^e$ | $\hat{y}$ | $(\hat{y} - \bar{y}^e)^2$ | $(y^e - \bar{y}^e)^2$ | *MAPE* |
|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 60,979 | 9,570 | 119,122 | 11,367 |
| 2 | 59,0 | 61,2 | 56,281 | 2,574 | 10,984 | 8,037 |
| 3 | 57,2 | 59,9 | 56,804 | 1,170 | 4,057 | 5,169 |
| 4 | 61,8 | 56,7 | 55,508 | 5,654 | 1,406 | 2,103 |
| 5 | 58,8 | 55,0 | 56,338 | 2,395 | 8,327 | 2,433 |
| 6 | 47,2 | 54,3 | 60,157 | 5,157 | 12,857 | 10,786 |
| 7 | 55,2 | 49,3 | 57,411 | 0,226 | 73,714 | 16,451 |
| **Сума** | 384,3 | 405,2 | 403,478 | **26,746** | **230,469** | 56,346 |
| **Сер. зн.** | 54,90 | 57,89 | 57,640 | 3,821 | 32,924 | **8,049** |

$$i_r = \sqrt{\frac{\sum\left(\hat{y}_i - \bar{y}\right)^2}{\sum\left(y_i^e - \bar{y}\right)^2}} = \sqrt{\frac{26{,}746}{230{,}469}} \approx 0{,}341.$$

Індекс кореляції є мірою тісноти зв’язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже слабкий. Отримана економетрична модель не є адекватною.

Середня похибка апроксимації *MAPE* ≈ 8,05% показує, що у середньому розрахункові значення відхиляються від фактичних на 8,05%.

3) Нехай економетрична модель має *показникову* форму:

$$y = \beta_0 \cdot \beta_1^x \cdot \varepsilon,$$

де $\beta_0$ і $\beta_1$ – параметри моделі, $\varepsilon$ - залишок.

Оцінимо параметри моделі $\hat{y} = b_0 \cdot b_1^x$ за допомогою МНК. Для цього спочатку виконаємо процедуру *лінеаризації* змінних. Логарифмуємо обидві частини рівняння:

$$\ln \hat{y} = \ln b_0 + x \cdot \ln b_1.$$

Отримаємо лінійне рівняння парної регресії

$$Y = B_0 + B_1 x,$$

де $$\hat{Y} = \ln \hat{y},\ B_0 = \ln b_0;\ B_1 = \ln b_1 .$$

Вихідні дані та їх перетворення для побудови моделі наведені у таблиці 3.10.

Таблиця 3.10

| $i$ | $x$ | $y^e$ | $Y=\ln y$ | $(x-\bar{x})^2$ | $(x-\bar{x})(Y-\bar{Y})$ | $\hat{y}$ |
|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 4,2312 | 96,0400 | -1,7400 | 60,704 |
| 2 | 59 | 61,2 | 4,1141 | 16,8100 | 0,2480 | 56,360 |
| 3 | 57,2 | 59,9 | 4,0927 | 5,2900 | 0,0898 | 56,904 |
| 4 | 61,8 | 56,7 | 4,0378 | 47,6100 | -0,1096 | 55,523 |
| 5 | 58,8 | 55 | 4,0073 | 15,2100 | -0,1807 | 56,420 |
| 6 | 47,2 | 54,3 | 3,9945 | 59,2900 | 0,4553 | 60,027 |
| 7 | 55,2 | 49,3 | 3,8979 | 0,0900 | -0,0467 | 57,515 |
| Σ | 384,3 | 405,2 | 28,3756 | 240,3400 | -1,2839 | 403,452 |
| **Сер.зн** | 54,9 | 57,88571 | 4,0537 | **34,3343** | **-0,1834** | 57,636 |

Результати обчислень:

$$B_1 = \frac{cov(x,Y)}{var(x)} = \frac{-0{,}1834}{34{,}3343} \approx -0{,}00534; \quad b_1 = e^{B_1} \approx e^{-0{,}00534} \approx 0{,}9947;$$

$$B_0 = \overline{Y} - b_1 \overline{x} \approx 4{,}0537 + 0{,}00534 \cdot 54{,}9 \approx 4{,}347; \quad b_0 = e^{B_0} \approx e^{4{,}347} \approx 77{,}24.$$

Лінеарізована модель має такий вигляд:

$$\hat{Y} = 4{,}347 - 0{,}00534 \cdot x.$$

Рівняння шуканої показникової регресії має такий вигляд:

$$\hat{y} = 77{,}24 \cdot 0{,}9947^{x}.$$

Знайдемо показники тісноти зв'язку: індекс кореляції $i_r$ і середню похибку апроксимації *MAPE*. Необхідні обчислення наведені у таблиці 3.11.

Індекс кореляції

$$i_r = \sqrt{\frac{\sum\left(\hat{y}_i - \overline{y}\right)^2}{\sum\left(y_i^e - \overline{y}\right)^2}} = \sqrt{\frac{3{,}384}{32{,}924}} \approx 0{,}321.$$

Таблиця 3.11

| $i$ | $x$ | $y^e$ | $\hat{y}$ | $(y^e - \bar{y})^2$ | $(\hat{y} - \bar{y})^2$ | *MAPE* |
|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 60,704 | 119,122 | 7,941 | 11,768 |
| 2 | 59 | 61,2 | 56,360 | 10,984 | 2,329 | 7,909 |
| 3 | 57,2 | 59,9 | 56,904 | 4,057 | 0,963 | 5,001 |
| 4 | 61,8 | 56,7 | 55,523 | 1,406 | 5,583 | 2,076 |
| 5 | 58,8 | 55 | 56,420 | 8,327 | 2,149 | 2,582 |
| 6 | 47,2 | 54,3 | 60,027 | 12,857 | 4,583 | 10,546 |
| 7 | 55,2 | 49,3 | 57,515 | 73,714 | 0,137 | 16,664 |
| **Σ** | 384,3 | 405,2 | 403,452 | 230,469 | 23,685 | 56,546 |
| **Сер.зн** | 54,9 | 57,88571 | 57,636 | **32,924** | **3,384** | **8,078** |

Індекс кореляції є мірою тісноти зв’язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже слабкий. Отримана економетрична модель не є адекватною.

Середня похибка апроксимації $MAPE \approx 8{,}08\,\%$ показує, що у середньому розрахункові значення відхиляються від фактичних на 8,08 %.

4) Нехай економетрична модель має форму *рівнобічної гіперболи:*

$$y = \beta_0 + \beta_1 \cdot \frac{1}{x} + \varepsilon,$$

де $\beta_0$ і $\beta_1$ – параметри моделі, $\varepsilon$ - залишок.

Оцінимо параметри моделі $\hat{y} = b_0 + b_1 \cdot \frac{1}{x}$ за допомогою МНК. Виконаємо лінеаризацію змінних заміною $X = 1 / x$ та отримаємо таке лінійне рівняння регресії:

$$y = b_0 + b_1 X.$$

Вихідні дані та їх перетворення для побудови моделі наведені у таблиці 3.12.

Результати обчислень:

$$b_1 = \frac{cov(X, y)}{var(X)} = \frac{0{,}0048}{0{,}000005} \approx 1054{,}67;$$

$$b_0 = \bar{y} - b_1 \bar{X} \approx 57{,}886 - 1054{,}67 \cdot 0{,}0184 \approx 38{,}435.$$

Таблиця 3.12

| $i$ | $x$ | $y^e$ | $X = 1 / x$ | $(X - \overline{X})^2$ | $(X - \bar{X})(y - \bar{y})$ | $\hat{y}$ |
|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 0,0222 | 0,000014 | 0,0407 | 61,820 |
| 2 | 59,0 | 61,2 | 0,0169 | 0,000002 | -0,0049 | 56,311 |
| 3 | 57,2 | 59,9 | 0,0175 | 0,000001 | -0,0019 | 56,874 |
| 4 | 61,8 | 56,7 | 0,0162 | 0,000005 | 0,0027 | 55,501 |
| 5 | 58,8 | 55,0 | 0,0170 | 0,000002 | 0,0041 | 56,372 |
| 6 | 47,2 | 54,3 | 0,0212 | 0,000008 | -0,0098 | 60,780 |
| 7 | 55,2 | 49,3 | 0,0181 | 0,000000 | 0,0028 | 57,542 |
| Σ | 384,3 | 405,2 | 0,1291 | 0,000032 | 0,0336 | 405,200 |
| Сер.зн | 54,9 | 57,886 | 0,0184 | **0,000005** | **0,0048** | 57,886 |

Лінеарізована модель має такий вигляд:

$$\hat{Y} = 38,435 + 1054,67 \cdot X.$$

Рівняння шуканої степеневої регресії має такий вигляд:

$$\hat{y} = 38,435 + 1054,67 \cdot (1/x)$$

Знайдемо показники тісноти зв'язку: індекс кореляції $i_r$ і середню похибку апроксимації *MAPE*. Необхідні обчислення наведені у таблиці 3.13.

Таблиця 3.13

| $i$ | $x$ | $y^e$ | $\hat{y}$ | $(y^e - \bar{y})^2$ | $(\hat{y} - \bar{y})^2$ | *MAPE* |
|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 61,82 | 119,1216 | 15,4824 | 10,145 |
| 2 | 59,0 | 61,2 | 56,311 | 10,9845 | 2,4794 | 7,988 |
| 3 | 57,2 | 59,9 | 56,874 | 4,0573 | 1,0243 | 5,052 |
| 4 | 61,8 | 56,7 | 55,501 | 1,4059 | 5,6859 | 2,114 |
| 5 | 58,8 | 55,0 | 56,372 | 8,3273 | 2,2916 | 2,494 |
| 6 | 47,2 | 54,3 | 60,78 | 12,8574 | 8,3771 | 11,934 |
| 7 | 55,2 | 49,3 | 57,542 | 73,7145 | 0,1184 | 16,717 |
| Σ | 384,3 | 405,2 | 405,2 | 230,4686 | 35,4592 | 56,445 |
| Сер.зн | 54,9 | 57,886 | 57,886 | **32,9241** | **5,0656** | **8,064** |

Індекс кореляції

$$i_r = \sqrt{\frac{\sum\left(\hat{y}_i - \bar{y}\right)^2}{\sum\left(y_i^e - \bar{y}\right)^2}} = \sqrt{\frac{5{,}0656}{32{,}924}} \approx 0{,}392\,.$$

Індекс кореляції є мірою тісноти зв'язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже слабкий. Отримана економетрична модель не є адекватною.

Середня похибка апроксимації $MAPE \approx 8{,}06\,\%$ показує, що у середньому розрахункові значення відхиляються від фактичних на 8,06%.

Обчислимо фактичне значення статистики $F$-критерію:

$$F_{факт} = \frac{i_r^{\,2}(n-m)}{(1-i_r^2)(m-1)} = \frac{0{,}392^2 \cdot (7-2)}{(1-0{,}392^2)(2-1)} \approx 0{,}91..$$

При рівні значущості α = 0,05 знайдемо за таблицею 1 Додатку критичне значення

$$F_{кр}\,(\alpha;\, k_1;\, k_2) = F_{кр}\,(\alpha;\, m - 1;\, n - m) = F_{кр}\,(0{,}05;\, 1;\, 5) = 6{,}61.$$

Оскільки $F_{факт} < F_{кр}$, то при рівні значущості α = 0,05 приймається гіпотеза $H_0$ про статистично *незначущі* параметри цього рівняння. Модель не є адекватною.

5) Нехай економетрична модель має форму *полінома третьої степені:* $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon,$

де $\beta_0,\ \beta_1, \beta_2, \beta_3$ – параметри моделі, $\varepsilon$ - залишок.

Статистичні оцінки $b_0,\ b_1,\ b_2,\ b_3$ параметрів $\beta_0,\ \beta_1,\ \beta_2,\ \beta_3$ для шуканого рівняння регресії можна знайти, якщо використати формулу (3.5):

$$X = \begin{pmatrix} 1 & 45{,}1 & 2034{,}01 & 91733{,}851 \\ 1 & 47{,}2 & 2227{,}84 & 105154{,}048 \\ 1 & 55{,}2 & 3047{,}04 & 168196{,}608 \\ 1 & 57{,}2 & 3271{,}84 & 187149{,}248 \\ 1 & 58{,}8 & 3457{,}44 & 203297{,}472 \\ 1 & 59 & 3481 & 205379 \\ 1 & 61{,}8 & 3819{,}24 & 236029{,}032 \end{pmatrix}$$

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 45,1 & 47,2 & 55,2 & 57,2 & 58,8 & 59 & 61,8 \\ 2034,01 & 2227,84 & 3047,04 & 3271,84 & 3457,44 & 3481 & 3819,24 \\ 91733,85 & 105154 & 168196,608 & 187149,248 & 203297,5 & 205379 & 236029 \end{pmatrix}$$

$$X^TX = \begin{pmatrix} 7 & 384,3 & 21338,41 & 1196939,259 \\ 384,3 & 21338,41 & 1196939,26 & 67747704,02 \\ 21338,41 & 1196939 & 67747704 & 3864942783 \\ 1196939 & 67747704 & 3864942783 & 2,22008E+11 \end{pmatrix}$$

$$(X^TX)^{-1} = \begin{pmatrix} 373707,5 & -20937,1 & 388,0902 & -2,38194 \\ -20937,1 & 1173,684 & -21,7684 & 0,133688 \\ 388,0902 & -21,7684 & 0,403993 & -0,00248 \\ -2,38194 & 0,133688 & -0,00248 & 1,53E-05 \end{pmatrix}$$

$$(X^TX)^{-1}X^T = \begin{pmatrix} 317,654 & -392,945 & -130,318 & 93,975 & 159,945 & 157,841 & -205,151 \\ -17,418 & 22,065 & 6,879 & -5,555 & -9,057 & -8,909 & 11,995 \\ 0,316 & -0,409 & -0,119 & 0,109 & 0,170 & 0,166 & -0,233 \\ -0,002 & 0,003 & 0,001 & -0,001 & -0,001 & -0,001 & 0,001 \end{pmatrix}$$

$$B(X^TX)^{-1}X^TY = \begin{pmatrix} 6546,839 \\ -357,084 \\ 6,499722 \\ -0,03917 \end{pmatrix}$$

Отримаємо:

$b_0 = 6546{,}839$; $b_1 = -357{,}084$; $b_2 = 6{,}500$; $b_3 = -\,0{,}039$.

Регресійне рівняння залежності буде мати такий вигляд:

$$\hat{y}_i = 6546{,}839 - 357{,}084\, x_i + 6{,}500\, x_i^2 - 0{,}039\, x_i^3.$$

Знайдемо показники тісноти зв'язку: індекс кореляції $i_r$ і середню похибку апроксимації *MAPE*. Необхідні обчислення наведені у таблиці 3.14.

Індекс кореляції

$$i_r = \sqrt{\frac{\sum\left(\hat{y}_i - \overline{y}\right)^2}{\sum\left(y_i^e - \overline{y}\right)^2}} = \sqrt{\frac{191{,}329}{230{,}469}} \approx 0{,}911.$$

Таблиця 3.14

| $i$ | $x$ | $y$ | $\hat{y}$ | $(y\text{-}\bar{y})^2$ | $(\hat{y}\text{-}\bar{y})^2$ | *MAPE* |
|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 69,252 | 119,122 | 129,192 | 0,657 |
| 2 | 47,2 | 54,3 | 53,493 | 12,857 | 19,292 | 1,485 |
| 3 | 55,2 | 49,3 | 51,763 | 73,714 | 37,488 | 4,996 |
| 4 | 57,2 | 59,9 | 56,281 | 4,057 | 2,574 | 6,041 |
| 5 | 58,8 | 55 | 58,705 | 8,327 | 0,671 | 6,736 |
| 6 | 59 | 61,2 | 58,880 | 10,984 | 0,988 | 3,791 |
| 7 | 61,8 | 56,7 | 56,826 | 1,406 | 1,123 | 0,222 |
| **Σ** | 384,3 | 405,2 | 405,20 | **230,469** | **191,329** | 23,928 |
| **Сер.зн.** | 54,9 | 57,886 | 57,886 | 32,924 | 27,333 | **3,418** |

Індекс кореляції є мірою тісноти зв’язку всіх пояснювальних змінних із залежною, тобто у даному випадку зв'язок між факторами є дуже тісним. Отримана економетрична модель є адекватною.

Середня похибка апроксимації $MAPE \approx 3{,}42\,\%$ показує, що у середньому розрахункові значення відхиляються від фактичних на 3,42 %.

Обчислимо фактичне значення статистики *F*-критерію Фішера:

$$F_{факт} = \frac{i_r^{\,2}(n-m)}{(1-i_r^2)(m-1)} = \frac{0{,}911^{\,2}\cdot(7-4)}{(1-0{,}911^{\,2})(4-1)} \approx 4{,}89..$$

При рівні значущості α = 0,05 знайдемо за таблицею 1 Додатку критичне значення

$$F_{кр}\,(\alpha;\, k_1;\, k_2) = F_{кр}\,(\alpha;\, m - 1;\, n - m) = F_{кр}\,(0{,}05;\, 3;\, 3) = 9{,}28.$$

Оскільки $F_{табл} > F_{кр}$, то приймається гіпотеза $H_0$ про статистично *значущі* параметри цього рівняння. Модель є адекватною.

Використовуючи засоби *Excel* («Диаграмма» → «Добавить линию тренда» → Вид «Полиномиальная» степени 3), можна також отримати рівняння регресії та її графік (теоретичну лінію регресії). Графік залежності між результативним та пояснювальним факторами, побудований за розрахованою моделлю, поданий на рис. 3.13.

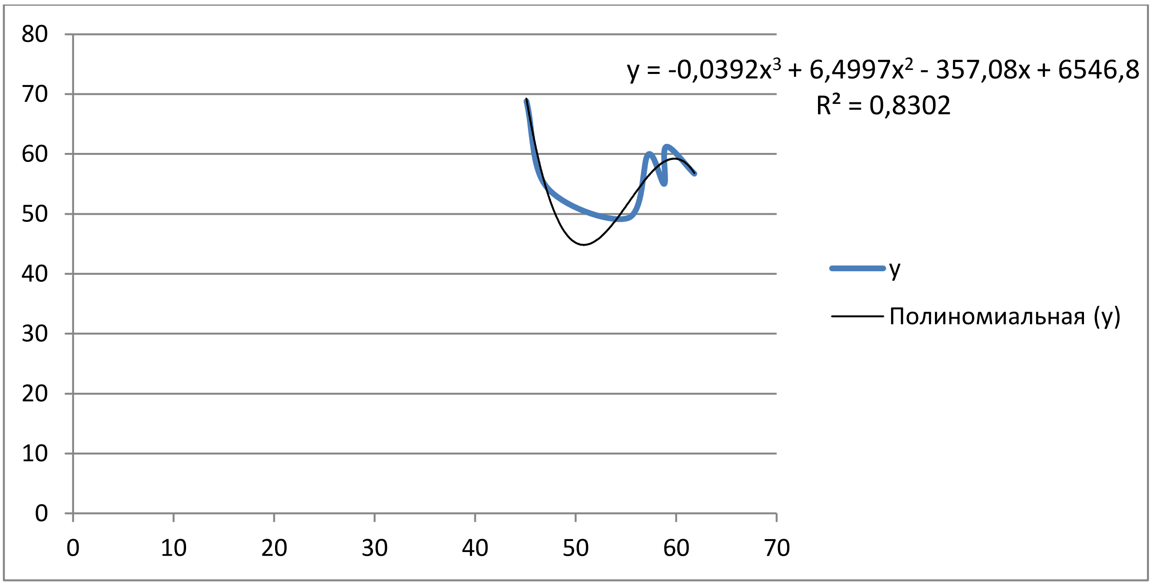



Рис. 3.13

Отримані результати запишемо у таблиці 3.15.

Таблиця 3.15

| Вигляд | Рівняння | Коефіцієнт детермінації | *MAPE* |
|---|---|---|---|
| Лінійна | $\hat{y} = 76{,}88 - 0{,}35\ x.$ | 0,125 | 8,14 |
| Степенева | $\hat{y} = 190{,}033 \cdot x^{-0{,}298}$ | 0,116 | 8,05 |
| Показникова | $\hat{y} = 77{,}24 \cdot 0{,}9947^{x}.$ | 0,103 | 8,08 |
| Гіперболічна | $\hat{y} = 38{,}435 + 1054{,}67 \cdot (1/x)$ | 0,154 | 8,06 |
| Поліноміальна степені 3 | $\hat{y}_i = 6546{,}839 - 357{,}084\ x_i + 6{,}500\ x_i^2 - 0{,}039\ x_i^3$ | 0,830 | 3,42 |

*Висновок*. Із побудованих економетричних моделей перші чотири (лінійна, степенева, показникова та гіперболічна) не є адекватними і не придатні для прогнозування. Побудова графіків залежностей для вибіркових даних та візуальний аналіз дали би інформацію про недоцільність використання цих моделей. Побудована поліноміальна третьої степені економетрична модель залежності між витратами на купівлю продовольчих товарів та середньоденною заробітною платою одного робітника є адекватною.

### 3.3.2. Виробнича функція Кобба-Дугласа

У реальних ситуаціях обсяг випуску продукції визначається, як правило, не одним, а багатьма факторами, тому частіше застосовують багатофакторні виробничі функції. Найпоширенішою серед них є виробнича функція Кобба-Дугласа.

**Приклад 3.8**. Побудувати виробничу функцію Кобба-Дугласа. Знайти основні економічні характеристики і зробити економічний аналіз на підставі даних таблиці 3.16. За допомогою критерія Фішера для рівня значущості $\alpha = 0{,}05$ перевірити адекватність знайденої моделі. Зробити прогноз $V\,(566;75)$.

Таблиця 3.16

Вибіркова сукупність даних про обсяг виробництва, виробничих фондів та вкладеної праці

| Підприємство | Обсяг виробництва, грош.од. | Виробничі фонди, од. | Вкладена праця, люд.-год. |
|---|---|---|---|
| № | $V$ | $K$ | $L$ |
| 1 | 232 | 605 | 30 |
| 2 | 289 | 195 | 181 |
| 3 | 351 | 515 | 85 |
| 4 | 1030 | 1070 | 470 |
| 5 | 568 | 570 | 254 |
| 6 | 244 | 904 | 21 |
| 7 | 132 | 69 | 94 |
| 8 | 283 | 158 | 216 |
| 9 | 316 | 251 | 166 |
| 10 | 127 | 73 | 84 |
| 11 | 209 | 87 | 192 |
| 12 | 348 | 458 | 104 |
| 13 | 458 | 231 | 412 |
| 14 | 160 | 61 | 170 |
| 15 | 580 | 420 | 340 |

***Розв'язування***

1) *Побудова економетричної моделі обсягу виробництва*.

Ідентифікація змінних:

*V* – обсяг виробництва (залежна змінна);

*K* - виробничі фонди,  *L* – вкладена праця (незалежні змінні).

Специфікація моделі.

Побудуємо графік залежності $V = f(K, L)$ (рис.3.14). Для цього використаємо пакет прикладних програм *Statistica*. Із умови прикладу переносимо дані до програми *Statistica*. Перейменувати стовпці згідно з їх значеннями (замість *var* записати *K, L, V*). На панелі інструментів знаходимо команду *Графика* → *Графики поверхностей.* У інформаційному вікні обираємо вкладнику *Быстрый*. Клацнувши на кнопку *Переменные*, вводимо значення змінних *K, L, V*. Далі виконуємо команду *Подгонка* → *Наименьшие квадраты*. Ставимо прапорець *Показать точки данных на поверхности*. Після натискання *ОК* з'являється графік поверхні.

Візуальний аналіз поверхні дозволяє припустити, що маємо нелінійну залежність виробничої функції Кобба – Дугласа: $\hat{V} = A\,K^{\alpha}\,L^{\beta}$.

Для лінеаризації моделі виконаємо логарифмування:

$$\ln \hat{V} = \ln(A\,K^{\alpha}\,L^{\beta}),$$

$$\ln \hat{V} = \ln A + \alpha \ln K + \beta \ln L.$$

Зробимо заміну:

$$\ln \hat{V} = \hat{y},\quad \ln A = b_0,\quad \alpha = b_1,\quad \beta = b_2,\quad \ln K = x_1,\quad \ln L = x_2.$$

Отримаємо таку лінійну регресійну модель:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2.$$

Побудуємо графік залежності $y(x_1, x_2)$ (рис.3.15). Перейменувати стовпці згідно з їх значеннями (замість *var* записати $x_1$, $x_2$, $y$). На панелі інструментів знаходимо команду *Графика* → *Графики поверхностей.* У інформаційному вікні обираємо вкладнику *Быстрый*. Клацнувши на кнопку *Переменные*, вводимо значення змінних $x_1$, $x_2$, $y$. Далі виконуємо команду *Подгонка* → *Линейная*.

Ставимо прапорець *Показать точки данных на поверхности*. Після натискання *OK* з'являється графік поверхні з виглядом її лінеарізованого рівняння. Графіком є площина, тому можна стверджувати, що маємо лінійну модель $\hat{y} = b_0+b_1x_1+b_2\,x_2$.

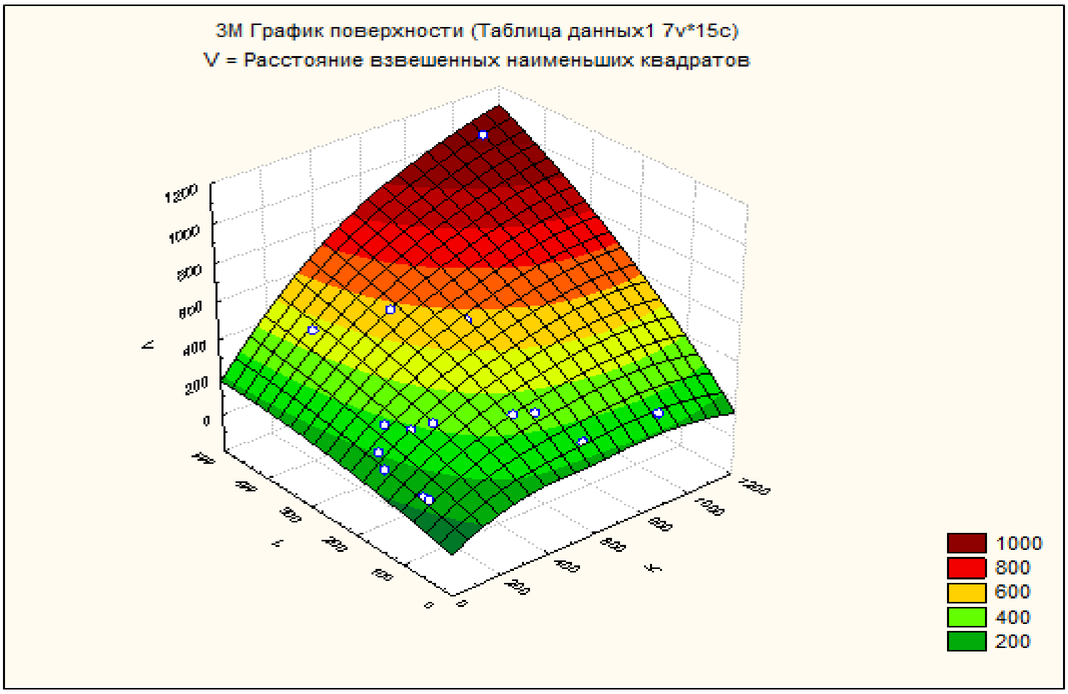



Рис. 3.14

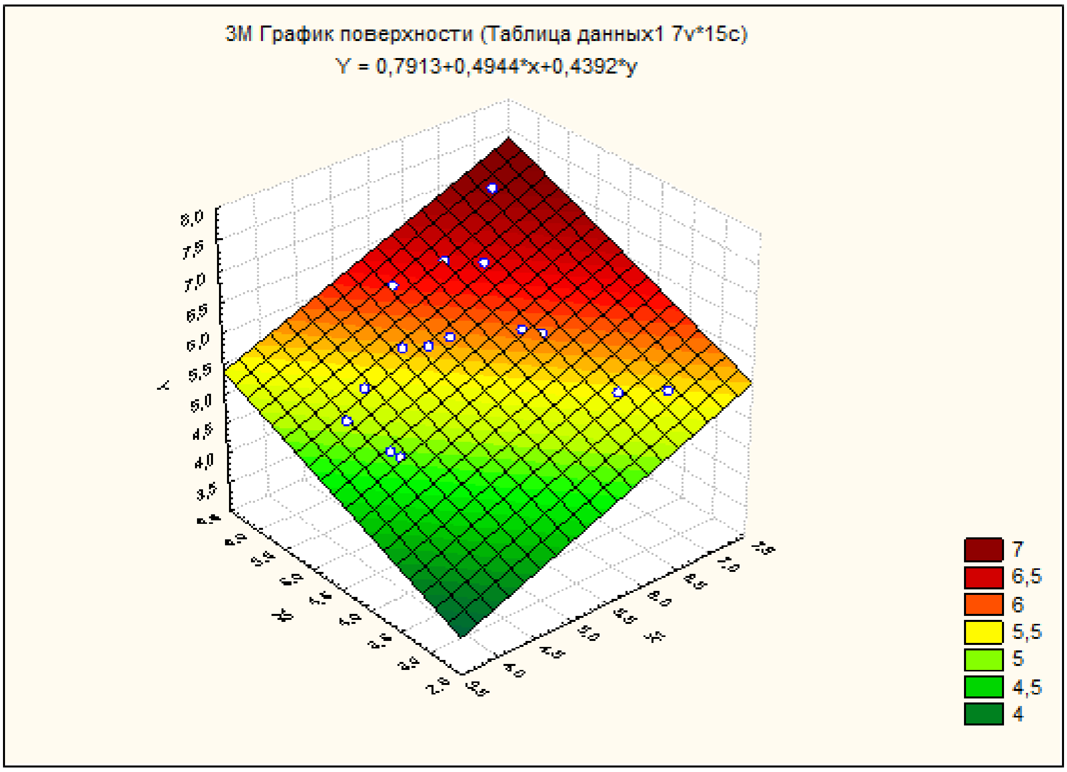



Рис. 3.15

Для знаходження коефіцієнтів $b_0$ , $b_1$ , $b_2$ розв'яжемо системy нормальних рівнянь:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{1i}x_{2i} = \sum_{i=1}^{n} x_{1i}\, y_i \\ b_0 \sum_{i=1}^{n} x_{2i} + b_1 \sum_{i=1}^{n} x_{1i}x_{2i} + b_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} x_{2i}\, y_i \end{cases}$$

Необхідні обчислення запишемо в таблиці 3.17.

Система нормальних рівнянь має вигляд:

$$\begin{cases} 15b_0 + 83{,}3492\, b_1 + 74{,}1260\, b_2 = 85{,}6383 \\ 83{,}3492\, b_0 + 476{,}0605\, b_1 + 410{,}4480\, b_2 = 481{,}6146 \\ 74{,}1260\, b_0 + 410{,}4480\, b_1 + 377{,}2341\, b_2 = 427{,}2868 \end{cases}$$

Розв'яжемо систему рівнянб методом оберненої матриці:

$$B = \begin{pmatrix} 15 & 83{,}349 & 74{,}126 \\ 83{,}349 & 476{,}06 & 410{,}45 \\ 74{,}126 & 410{,}45 & 377{,}23 \end{pmatrix}^{-1} \times \begin{pmatrix} 85{,}638 \\ 481{,}61 \\ 427{,}29 \end{pmatrix} =$$

$$= \begin{pmatrix} 5{,}33 & -0{,}488 & -0{,}517 \\ -0{,}488 & 0{,}079 & 0{,}0104 \\ -0{,}517 & 0{,}01 & 0{,}0929 \end{pmatrix} \times \begin{pmatrix} 85{,}64 \\ 481{,}6 \\ 427{,}3 \end{pmatrix} = \begin{pmatrix} 0{,}791 \\ 0{,}494 \\ 0{,}439 \end{pmatrix}$$

(при знаходженні матриць використовувались функції Excel МОБР( ) та МУМНОЖ( )).

Розв'язок системи: $b_0 = 0{,}7913$, $b_1 = 0{,}4944$, $b_2 = 0{,}4392$.

Лінеарізоване рівняння регресії має такий вигляд:

$$\hat{y} = 0{,}7913 + 0{,}4944\, x_1 + 0{,}4392\, x_2.$$

Виконаємо зворотне перетворення:

$A = exp(b_0) = e^{0{,}7919} = 2{,}2064$; $\alpha = b_1 = 0{,}4944$; $\beta = b_2 = 0{,}4392$.

Вибіркова виробнича функція Кобба-Дугласа набуває вигляду:

$$\hat{V} = 2{,}2064 \cdot K^{0{,}4944} \cdot L^{0{,}4392}.$$

Таблиця 3.17

Допоміжні змінні для побудови функції Кобба-Дугласа

| № | $V$ | $K$ | $L$ | $lnV$ ($y$) | $lnK$ ($x_1$) | $lnL$ ($x_2$) |
|---|---|---|---|---|---|---|
| 1 | *232* | *605* | *30* | *5,4467* | *6,4052* | *3,4012* |
| 2 | *289* | *195* | *181* | *5,6664* | *5,2730* | *5,1985* |
| 3 | *351* | *515* | *85* | *5,8608* | *6,2442* | *4,4427* |
| 4 | *1030* | *1070* | *470* | *6,9373* | *6,9754* | *6,1527* |
| 5 | *568* | *570* | *254* | *6,3421* | *6,3456* | *5,5373* |
| 6 | *244* | *904* | *21* | *5,4972* | *6,8068* | *3,0445* |
| 7 | *132* | *69* | *94* | *4,8828* | *4,2341* | *4,5433* |
| 8 | *283* | *158* | *216* | *5,6454* | *5,0626* | *5,3753* |
| 9 | *316* | *251* | *166* | *5,7557* | *5,5255* | *5,1120* |
| 10 | *127* | *73* | *84* | *4,8442* | *4,2905* | *4,4308* |
| 11 | *209* | *87* | *192* | *5,3423* | *4,4659* | *5,2575* |
| 12 | *348* | *458* | *104* | *5,8522* | *6,1269* | *4,6444* |
| 13 | *458* | *231* | *412* | *6,1269* | *5,4424* | *6,0210* |
| 14 | *160* | *61* | *170* | *5,0752* | *4,1109* | *5,1358* |
| 15 | *580* | *420* | *340* | *6,3630* | *6,0403* | *5,8289* |
| $\Sigma$ | ***5327*** | ***5667*** | ***2819*** | ***85,6383*** | ***83,3492*** | ***74,1260*** |

Таблиця 3.17 (продовження)

| № | $x_1^2$ | $x_2^2$ | $x_1 x_2$ | $yx_1$ | $yx_2$ |
|---|---|---|---|---|---|
| 1 | *41,0270* | *11,5681* | *21,78545* | *34,8876* | *18,52543* |
| 2 | *27,8045* | *27,0244* | *27,41167* | *29,87907* | *29,4569* |
| 3 | *38,9896* | *19,7372* | *27,74066* | *36,59573* | *26,03743* |
| 4 | *48,6564* | *37,8561* | *42,91786* | *48,39064* | *42,68344* |
| 5 | *40,2671* | *30,6621* | *35,13791* | *40,2448* | *35,11845* |
| 6 | *46,3329* | *9,2691* | *20,72354* | *37,41829* | *16,73625* |
| 7 | *17,9277* | *20,6415* | *19,23679* | *20,6743* | *22,18401* |
| 8 | *25,6299* | *28,8936* | *27,21286* | *28,58061* | *30,34585* |
| 9 | *30,5306* | *26,1324* | *28,24605* | *31,80308* | *29,42328* |
| 10 | *18,4080* | *19,6321* | *19,01024* | *20,78379* | *21,46371* |
| 11 | *19,9443* | *27,6413* | *23,47949* | *23,85837* | *28,0873* |
| 12 | *37,5385* | *21,5704* | *28,45558* | *35,85568* | *27,17992* |
| 13 | *29,6199* | *36,2527* | *32,76892* | *33,34498* | *36,89002* |
| 14 | *16,8993* | *26,3764* | *21,11262* | *20,8634* | *26,06507* |
| 15 | *36,4847* | *33,9766* | *35,20832* | *38,43431* | *37,08974* |
| $\Sigma$ | ***476,0605*** | ***377,2341*** | ***410,448*** | ***481,6146*** | ***427,2868*** |

*Висновки.*

Оскільки $\alpha + \beta = 0,9336 < 1$, то зміна кожного чинника на 1% сприятиме зростанню величини обсягу виробництва менш, ніж на 1%. Таким чином, темп зростання обсягу виробництва продукції нижче, ніж темп зростання виробничих ресурсів.

Збільшення на 1% витрат виробничих фондів при незмінному обсязі вкладеної праці обумовлює зростання обсягу виробництва наближено на 0,49%.

Збільшення на 1% витрат вкладеної праці при незмінному обсязі виробничих фондів обумовлює зростання обсягу виробництва наближено на 0,44%.

Збільшення ресурсівна 1% збільшує обсяг виробництва в 1,13 рази ($\alpha / \beta = 1,13$) більше, ніж 1% приросту вкладеної праці. Оскільки $\alpha > \beta$, то виробництво є більше фондомістким, ніж праце містким (інтенсивне зростання обсягу виробництва).

Коефіцієнт пропорційності $A = 2,2064$ можна трактувати як величину, яка враховує всі якісні (що не має виразу в обсягах виробничих фондів і витрат вкладеної праці) чинникі виробництва.

2) *Перевірка адекватності лінеарізованої моделі.*

Перевірка адекватності виконується виключно *для лінеарізованих моделей*. Справа у тому, що математичний апарат, що застосовується, розроблен для лінійних відносно параметрів моделей.

Запишемо допоміжні обчислення в таблиці 3.18.

Тут $y_i = \ln V_i$ і $\hat{y}_i$ - відповідно фактичне і розрахункове значення регресанду в лінеарізованій моделі, $\bar{y}$ - середнє значення регресанда, що обчислюється за формулою:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{85,6383}{15} \approx 5,7092.$$

Розрахункове значення величини $F_{\text{факт}}$ за формулою:

$$F_{\text{факт}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \cdot \frac{k_2}{k_1},$$

де $k_1 = m - 1 = 2$, $k_2 = n - m = 12$ ( $n = 15$ – кількість досліджених підприємств, $m = 3$ - число параметрів моделі: $A, \alpha, \beta$). Таким чином,

$$F_{\text{факт}} = \frac{4{,}6399}{0{,}0039} \cdot \frac{12}{2} = 7064{,}297.$$

При рівні значущості α = 0,05 знайдемо за таблицею 1 Додатку критичне значення

$$F_{\text{кр}}(\alpha; k_1; k_2) = F_{\text{кр}}(\alpha; m - 1; n - m) = F_{\text{кр}}(0{,}05; 2; 12) = 3{,}89.$$

Оскільки $F_{\text{факт}} > F_{\text{кр}}$, робимо висновок, що лінеарізована модель з надійністю 0,95 адекватна емпіричним даним.

Таблиця 3.18

| $i$ | $y$ | $\hat{y} = 0{,}7919 + 0{,}4944\, x_1 + 0{,}4392\, x_2$ | $(\hat{y} - \bar{y})^2$ | $(\hat{y} - y)^2$ |
|---|---|---|---|---|
| 1 | *5,4467* | *5,4522* | *0,0661* | *0,0000* |
| 2 | *5,6664* | *5,6818* | *0,0008* | *0,0002* |
| 3 | *5,8608* | *5,8300* | *0,0146* | *0,0010* |
| 4 | *6,9373* | *6,9426* | *1,5213* | *0,0000* |
| 5 | *6,3421* | *6,3609* | *0,4247* | *0,0004* |
| 6 | *5,4972* | *5,4941* | *0,0463* | *0,0000* |
| 7 | *4,8828* | *4,8803* | *0,6871* | *0,0000* |
| 8 | *5,6454* | *5,6554* | *0,0029* | *0,0001* |
| 9 | *5,7557* | *5,7686* | *0,0035* | *0,0002* |
| 10 | *4,8442* | *4,8588* | *0,7232* | *0,0002* |
| 11 | *5,3423* | *5,3086* | *0,1605* | *0,0011* |
| 12 | *5,8522* | *5,8606* | *0,0229* | *0,0001* |
| 13 | *6,1269* | *6,1268* | *0,1744* | *0,0000* |
| 14 | *5,0752* | *5,0796* | *0,3964* | *0,0000* |
| 15 | *6,3630* | *6,3380* | *0,3954* | *0,0006* |
| **Сума** | ***85,6383*** | *85,6383* | **4,6399** | ***0,0039*** |
| **Сер.зн** | *5,7092* | *5,7092* | 0,3093 | *0,0003* |

Результати регресійного аналізу для лінеарізованого рівняння регресії засобами Excel зображені на рисунку 3.16.

| ВЫВОД ИТОГОВ | |
|---|---|
| *Регрессионная статистика* | |
| Множественный R | 0,9995756 |
| R-квадрат | 0,999151379 |
| Нормированный R-квадрат | 0,999009943 |
| Стандартная ошибка | 0,018122033 |
| Наблюдения | 15 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость F* |
|---|---|---|---|---|---|
| Регрессия | 2 | 4,639944742 | 2,319972 | 7064,2974 | 3,73492E-19 |
| Остаток | 12 | 0,003940897 | 0,000328 | | |
| Итого | 14 | 4,643885639 | | | |

| | *Коэффициенты* | *'тандартная ошибк* | *татисти* | *P-Значение* | *Нижние 95%* | *ерхние 95'* | *іжние 95,0* | *Верхние 95,0%* |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | 0,79134613 | 0,04183884 | 18,91415 | 2,674E-10 | 0,700187128 | 0,882505 | 0,700187 | 0,882505132 |
| Переменная X 1 | 0,494429358 | 0,005079003 | 97,34771 | 9,235E-19 | 0,48336316 | 0,505496 | 0,483363 | 0,505495556 |
| Переменная X 2 | 0,439223286 | 0,005523918 | 79,513 | 1,044E-17 | 0,427187703 | 0,451259 | 0,427188 | 0,451258869 |

Рис. 3.16.

3) *Знаходження основних економічних характеристик для виробничої функції* $\hat{V} = 2{,}2064 \cdot K^{0{,}4944} \cdot L^{0{,}4392}$.

*Фондовіддача:*

$$\hat{V}/K = A \cdot K^{\alpha-1} L^{\beta} = 2{,}2064 \cdot K^{-0{,}5656} \cdot L^{0{,}4392}.$$

*Продуктивність праці:*

$$\hat{V}/L = A \cdot K^{\alpha} L^{\beta-1} = 2{,}2064 \cdot K^{0{,}4944} \cdot L^{-0{,}5608}.$$

*Фондомісткість*

$$K/\hat{V} = \frac{1}{A} \cdot K^{1-\alpha} L^{-\beta} = 0{,}4532 \cdot K^{0{,}5056} \cdot L^{-0{,}4392}.$$

*Працемісткість:*

$$\frac{L}{\hat{V}} = \frac{1}{A} \cdot K^{-\alpha} \cdot L^{1-\beta} = 0{,}4532 \cdot K^{-0{,}4944} \cdot L^{0{,}5608}.$$

*Гранична фондовіддача:*

$$\partial\hat{V}/\partial K = A \cdot \alpha \cdot K^{\alpha-1} L^{\beta} = \alpha \cdot \frac{\hat{V}}{K} = 0{,}4944 \cdot \frac{\hat{V}}{K}.$$

*Гранична продуктивність праці:*

$$\partial\hat{V}/\partial L = A \cdot \beta \cdot K^{\alpha} L^{\beta-1} = \beta \cdot \frac{\hat{V}}{L} = 0{,}4392 \cdot \frac{\hat{V}}{L}.$$

*Еластичність випуску за основними виробничими фондами:*

$$E_K = \alpha = 0{,}4944,$$

тобто збільшення на 1% витрат виробничих фондів при незмінному обсязі вкладеної праці обумовлює зростання обсягу виробництва наближено на 0,49 %.

*Еластичність випуску за витратами праці:*

$$E_L = \beta = 0{,}4392,$$

тобто збільшення витрат на 1% вкладеної праці при незмінному обсязі виробничих фондів обумовлює зростання обсягу виробництва наближено на 0,44%.

*Граничні технологічні норми заміщення ресурсів:*
вкладеної праці виробничими фондами

$$h_{KL} = \frac{\beta}{\alpha} \cdot \frac{K}{L} = \frac{0{,}4392}{0{,}4944} \cdot \frac{K}{L} = 0{,}88835 \cdot \frac{K}{L},$$

виробничих фондів вкладеною працею

$$h_{LK} = \frac{\alpha}{\beta} \cdot \frac{L}{K} = \frac{0{,}4944}{0{,}4392} \cdot \frac{L}{K} = 1{,}12568 \cdot \frac{L}{K}.$$

Норми показують, на скільки одиниць необхідно збільшити використання ресурсу $K$ ($L$), якщо ресурс $L$ ($K$) зменшити на одну одиницю, щоб обсяг продукції $V$ залишився незмінним.

*Еластичність заміщення ресурсів*

$$\delta_{KL} = \partial \ln(K/L) / \partial \ln(h_{KL})$$

показує, на скільки відсотків зросте фондоозброєність, якщо гранична норма заміщення збільшиться на 1 %.

Аналогічно визначається еластичність заміщення ресурсів $\delta_{LK}$. Можна довести, що для функції Кобба - Дугласа $\delta_{KL} = \delta_{LK} = 1$.

Запишемо значення розглянутих економічних характеристик для виробничої функції в таблиці 3.19.

*Висновки* (*на прикладі 14 підприємства*)

Фондовіддача $\hat{V}/K$ = *2,6347*, тобто на одну одиницю виробничих фондів доводиться у середньому *2,6347* гр. од. виробленої продукції.

Продуктивність праці $\hat{V}/L$ = *0,9454*, тобто на одну людино-годину вкладеної праці доводиться у середньому *0,9454* грош. од. виготовленої продукції.

Фондомісткість $\frac{K}{\hat{V}}$ = *0,3795*, тобто на одну грош.од. обсягу виробництва доводиться у середньому *0,3795* одиниць виробничих фондів.

Працемісткість $\frac{L}{\hat{V}}$ = *1,0578*, тобто на одну грош.од. обсягу продукції доводиться у середньому *1, 0578* люд.-год. вкладеної праці. Це означає, що для 14 підприємства залучення додаткової живої праці набагато збільшить прибуток, ніж введення в виробництво додаткових виробничих фондів.

Фондоозброєність $\frac{K}{L}$ = *0,3588*, тобто на одну люд.-год. вкладеної праці доводиться у середньому *0,3588* одиниць виробничих фондів.

Працеозброєність $\frac{L}{K}$ = *2,7869*, тобто на одиницю витрат виробничих фондів доволиться у середньому *2,7869* люд.-год. вкладеної праці.

Гранична фондовіддача $\frac{\partial \hat{V}}{\partial K}$= *1,3027*, тобто кожна додаткова одиниця виробничих фондів збільшить обсяг виробництва на *1,3027* одиниць.

Гранична продуктивність праці $\frac{\partial \hat{V}}{\partial L}$ = *0,4152*, тобто кожний додатковий робітник збільшить обсяг виробництва на *0,4152* одиниць

Гранична норма заміщення $h_{KL}$ = *0,3188*, тобто при збільшенні на *0,3188* одиниць виробничих фондів і зменшенні на одну одиницю вкладеної праці обсяг виробництва залишиться незмінним.

Прогнозне значення виробничої функції: якщо основні виробничі фонди складають $K$ = 566 одиниць, а витрати праці $L$ = 75 людино-годин, то обсяг виробництва

$\hat{V}(566; 75) = 2{,}2064 \cdot 566^{0{,}4944} \cdot 75^{0{,}4392} \approx 337{,}5368$ грош.од.

Таблиця 3.19

Значення економічних показників

| № | $\hat{V}$ | $K$ | $L$ | $\hat{V}/K$ | $\hat{V}/L$ | $K/\hat{V}$ | $L/\hat{V}$ | $K/L$ | $L/K$ | $\partial\hat{V}/\partial K$ | $\partial\hat{V}/\partial L$ | $h=(\beta/\alpha)(K/L)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 233,2625 | 605 | 30 | 0,3856 | 7,7754 | 2,5936 | 0,1286 | 20,1667 | 0,0496 | 0,1906 | 3,4151 | 17,9149 |
| 2 | 293,4693 | 195 | 181 | 1,5050 | 1,6214 | 0,6645 | 0,6168 | 1,0773 | 0,9282 | 0,7441 | 0,7121 | 0,9571 |
| 3 | 340,3456 | 515 | 85 | 0,6609 | 4,0041 | 1,5132 | 0,2497 | 6,0588 | 0,1650 | 0,3268 | 1,7587 | 5,3823 |
| 4 | 1035,479 | 1070 | 470 | 0,9677 | 2,2031 | 1,0333 | 0,4539 | 2,2766 | 0,4393 | 0,4785 | 0,9677 | 2,0224 |
| 5 | 578,7909 | 570 | 254 | 1,0154 | 2,2787 | 0,9848 | 0,4388 | 2,2441 | 0,4456 | 0,5021 | 1,0009 | 1,9935 |
| 6 | 243,2446 | 904 | 21 | 0,2691 | 11,5831 | 3,7164 | 0,0863 | 43,0476 | 0,0232 | 0,1330 | 5,0876 | 38,2411 |
| 7 | 131,6746 | 69 | 94 | 1,9083 | 1,4008 | 0,5240 | 0,7139 | 0,7340 | 1,3623 | 0,9435 | 0,6153 | 0,6521 |
| 8 | 285,8277 | 158 | 216 | 1,8090 | 1,3233 | 0,5528 | 0,7557 | 0,7315 | 1,3671 | 0,8944 | 0,5812 | 0,6498 |
| 9 | 320,0881 | 251 | 166 | 1,2753 | 1,9282 | 0,7842 | 0,5186 | 1,5120 | 0,6614 | 0,6305 | 0,8469 | 1,3432 |
| 10 | 128,8686 | 73 | 84 | 1,7653 | 1,5341 | 0,5665 | 0,6518 | 0,8690 | 1,1507 | 0,8728 | 0,6738 | 0,7720 |
| 11 | 202,0745 | 87 | 192 | 2,3227 | 1,0525 | 0,4305 | 0,9501 | 0,4531 | 2,2069 | 1,1484 | 0,4623 | 0,4025 |
| 12 | 350,9258 | 458 | 104 | 0,7662 | 3,3743 | 1,3051 | 0,2964 | 4,4038 | 0,2271 | 0,3788 | 1,4821 | 3,9121 |
| 13 | 457,9733 | 231 | 412 | 1,9826 | 1,1116 | 0,5044 | 0,8996 | 0,5607 | 1,7835 | 0,9802 | 0,4882 | 0,4981 |
| 14 | 160,717 | 61 | 170 | **2,6347** | **0,9454** | **0,3795** | **1,0578** | **0,3588** | **2,7869** | **1,3027** | **0,4152** | **0,3188** |
| 15 | 565,6831 | 420 | 340 | 1,3469 | 1,6638 | 0,7425 | 0,6010 | 1,2353 | 0,8095 | 0,6659 | 0,7308 | 1,0974 |

**Приклад 3.9**. Побудувати виробничу функцію Кобба-Дугласа та проаналізувати використання факторів на підставі даних таблиці 3.20.

Таблиця 3.20

Вибіркова сукупність даних про валовий регіональний продукт, зайнятість населення та основних засобів по регіонах України [11]

| **№ з/п** | **Область** | **Валовий регіональний продукт у факт. цінах, млн.грн.** | **Зайнятість населення, тис. осіб** | **Основні засоби у факт.цінах, млн..грн.** |
|---|---|---|---|---|
| | | ***Y*** | ***L*** | ***C*** |
| 1 | АР Крим | *9901* | *899,7* | *43758* |
| 2 | Вінницька | *8123* | *720,8* | *25993* |
| 3 | Волинська | *4994* | *423,9* | *14962* |
| 4 | Дніпропетровська | *30040* | *1554,7* | *128686* |
| 5 | Донецька | *45617* | *2086,0* | *129043* |
| 6 | Житомирська | *5947* | *561,6* | *20207* |
| 7 | Закарпатська | *5297* | *537,8* | *15182* |
| 8 | Запорізька | *15255* | *824,8* | *59225* |
| 9 | Івано-Франківська | *7311* | *513,5* | *26811* |
| 10 | Київська | *11883* | *764,4* | *42289* |
| 11 | Кіровоградська | *5594* | *444,5* | *19729* |
| 12 | Луганська | *14672* | *1019,8* | *51946* |
| 13 | Львівська | *13992* | *1057,0* | *51890* |
| 14 | Миколаївська | *7934* | *534,7* | *26451* |
| 15 | Одеська | *17029* | *1039,4* | *57173* |
| 16 | Полтавська | *13983* | *670,3* | *52493* |
| 17 | Рівненська | *5599* | *437,4* | *18964* |
| 18 | Сумська | *6275* | *534,4* | *26542* |
| 19 | Тернопільська | *3948* | *387,7* | *13913* |
| 20 | Харківська | *20524* | *1285,7* | *75594* |
| 21 | Херсонська | *5200* | *477,6* | *18219* |
| 22 | Хмельницька | *6344* | *573,6* | *22399* |
| 23 | Черкаська | *6623* | *561,1* | *23238* |
| 24 | Чернівецька | *3277* | *353,4* | *12893* |
| 25 | Чернігівська | *6181* | *506,1* | *27430* |
| 26 | м. Київ | *61357* | *1348,9* | *129325* |
| 27 | м. Севастополь | *2213* | *176,9* | *6714* |

***Розв'язування***

1) *Побудова економетричної моделі продуктивності праці.*

Ідентифікація змінних:

$Y$ – величина валового регіонального продукту (залежна змінна);

$L$ – зайнятість населення, $C$ – основні засоби у фактичних цінах (незалежні змінні).

Специфікація моделі.

Самостійно побудувати графік залежності $Y = f(L, C)$ за допомогою пакету прикладних програм *Statistica*. Візуальний аналіз поверхні дозволить припустити наявність нелінійної залежності виробничої функції Кобба – Дугласа: $\hat{Y} = A \cdot L^{\alpha} C^{\beta}$.

Для лінеаризації моделі виконаємо логарифмування:

$$\ln \hat{Y} = \ln(A\, L^{\alpha} C^{\beta}),$$

$$\ln \hat{Y} = \ln A + \alpha \ln L + \beta \ln C.$$

Зробимо заміну:

$$\ln \hat{Y} = \hat{y},\quad \ln A = b_0,\quad \alpha = b_1,\quad \beta = b_2,\quad \ln L = x_1,\quad \ln C = x_2.$$

Отримаємо таку лінійну регресійну модель:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2.$$

Побудуємо графік залежності $y(x_1, x_2)$ (рис.3.17). Графіком є площина, тому можна стверджувати, що маємо лінійну модель

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2.$$

Результати регресійного аналізу для лінеарізованого рівняння регресії засобами Excel зображені на рисунку 3.18.

Лінеарізоване рівняння регресії має такий вигляд:

$$\hat{y} = -1{,}3038 + 0{,}1113\, x_1 + 0{,}9356\, x_2.$$

Виконаємо зворотне перетворення:

$$A = \exp(b_0) = e^{-1{,}3038} = 0{,}2715;\quad \alpha = b_1 = 0{,}1113;\quad \beta = b_2 = 0{,}9356.$$

Вибіркова виробнича функція Кобба-Дугласа набуває вигляду:

$$\hat{Y} = 0{,}2715 \cdot L^{0{,}1113} \cdot C^{0{,}9356}.$$

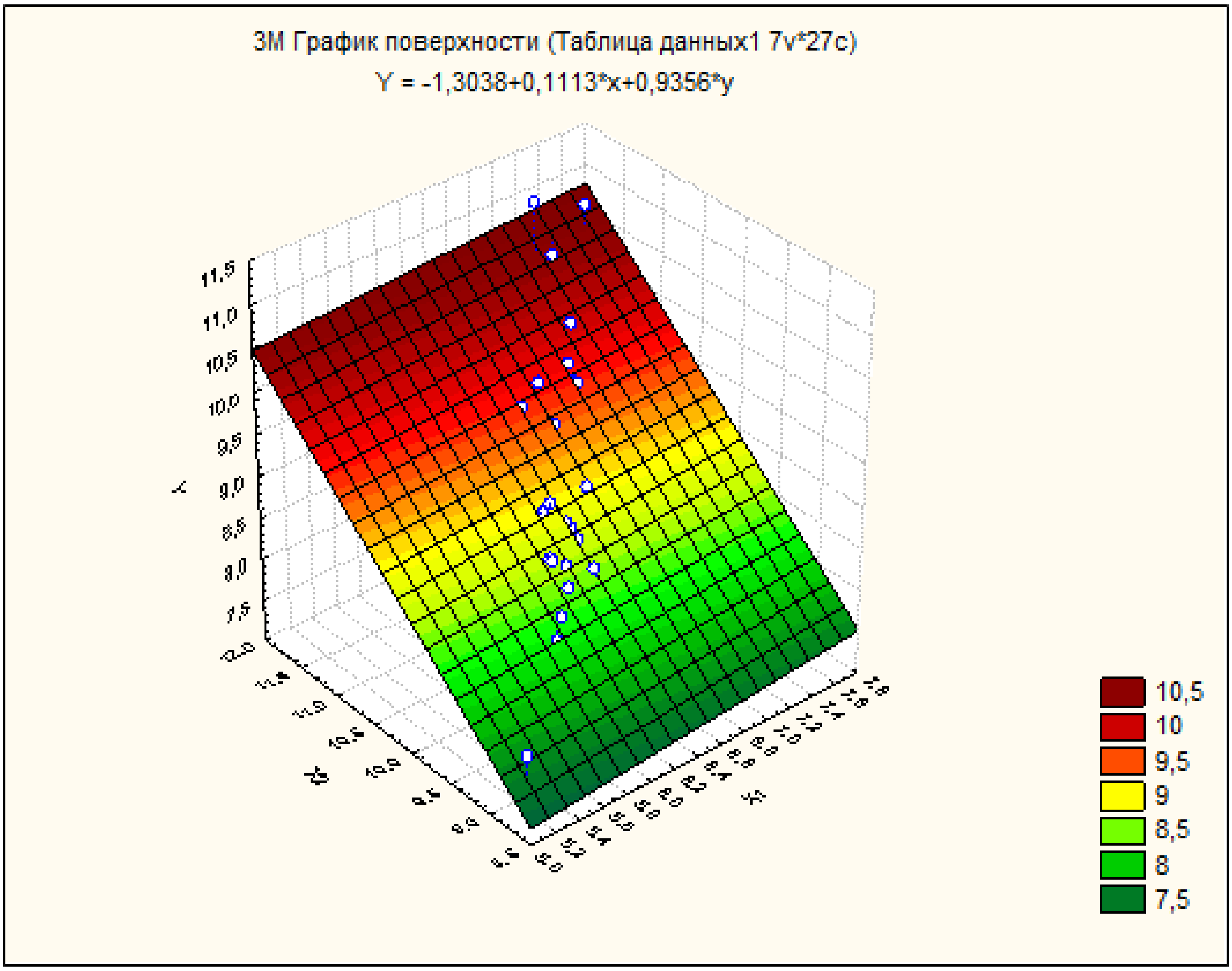



Рис. 3.17

| ВЫВОД ИТОГОВ | |
|---|---|
| *Регрессионная статистика* | |
| Множественный R | 0,98004489 |
| R-квадрат | 0,96048798 |
| Нормированный R-кв | 0,95719531 |
| Стандартная ошибка | 0,16036971 |
| Наблюдения | 27 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость F* |
|---|---|---|---|---|---|
| Регрессия | 2 | 15,0044 | 7,502200281 | 291,70506 | 1,44794E-17 |
| Остаток | 24 | 0,617243 | 0,025718444 | | |
| Итого | 26 | 15,62164 | | | |

| | *Коэффициент* | *дартная о* | *t-статистика* | *P-Значение* | *Нижние 95%* | *ерхние 95%* | *ижние 95,0* | *рхние 95,0%* |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | -1,3037726 | 0,443847 | -2,93743788 | 0,0071968 | -2,219827461 | -0,38772 | -2,21983 | -0,38772 |
| Переменная X 1 | 0,11126152 | 0,204327 | 0,544525776 | 0,5911038 | -0,310449451 | 0,532972 | -0,31045 | 0,532972 |
| Переменная X 2 | 0,93564718 | 0,142743 | 6,554751862 | 8,854E-07 | 0,641039403 | 1,230255 | 0,641039 | 1,230255 |

Рис. 3.18

*Висновки.*

Оскільки $\alpha + \beta =$ 1,0469 > 1, то зміна кожного чинника на 1% сприятиме зростанню величини валового регіонального продукту більш, ніж на 1%. Отже, темп зростання валового регіонального продукту вище, ніж темп зростання виробничих ресурсів.

Збільшення на 1% зайнятості населення при незмінному обсязі основних засобів обумовлює зростання валового регіонального продукту наближено на 0,1113 %.

Збільшення на 1% витрат основних засобів при незмінній зайнятості населення обумовлює зростання валового регіонального продукту наближено на 0,9356 %.

Збільшення основних засобів на 1% збільшує валовий регіональний продукт в 8,41 раз ($\beta / \alpha = 8{,}41$) більше, ніж 1% приросту зайнятості населення. Виробництво є значно більше фондомістким, ніж працемістким. Звернемо увагу на те, що дані таблиці – економічні показники України за 2005 рік. Самостійно проаналізувати, для якої економіки можлива така ситуація.

Коефіцієнт пропорційності $A$ = 0,2715 можна трактувати як величину, яка враховує всі якісні (що не має виразу в кількості зайнятого населення і основних засобів) чинникі виробництва.

2) *Перевірка адекватності лінеаризованої моделі.*

Розрахункове значення величини $F_{\text{факт}}$ = 291,71 (рис. 3.18).

При рівні значущості α = 0,05 знайдемо за таблицею 1 Додатку критичне значення

$$F_{\text{кр}}(\alpha; k_1; k_2) = F_{\text{кр}}(\alpha; m - 1; n - m) = F_{\text{кр}}(0{,}05; 2; 24) = 3{,}40.$$

Оскільки $F_{\text{факт}} > F_{\text{кр}}$, робимо висновок, що лінеаризована модель з надійністю 0,95 адекватна емпіричним даним.

3) *Знаходження основних економічних характеристик для виробничої функції* $\hat{Y} = 0{,}2715 \cdot L^{0,1113} \cdot C^{0,9356}$.

*Середня ефективність кількості зайнятого населення*:

$$\frac{\hat{Y}}{L} = 0{,}2715\, L^{-0,8887}\, C^{0,93564}.$$

*Середня ефективність використання основних засобів*:

$$\frac{\hat{Y}}{C} = 0{,}2715\, L^{\,0{,}1113}\, C^{\,-0{,}0644}.$$

*Гранична ефективність кількості зайнятого населення*:

$$\frac{\partial \hat{Y}}{\partial L} = A \cdot \alpha \cdot L^{\alpha-1} C^{\beta} = \alpha \cdot \frac{\hat{Y}}{L} = 0{,}1113 \cdot \frac{\hat{Y}}{L}.$$

*Гранична ефективність оснвних засобів*:

$$\frac{\partial \hat{Y}}{\partial C} = A \cdot \beta \cdot L^{\alpha} C^{\beta-1} = \beta \cdot \frac{\hat{Y}}{C} = 0{,}9356 \cdot \frac{\hat{Y}}{C}.$$

*Еластичність за кількістю зайнятого населення*:

$$E_L = \alpha = 0{,}1113,$$

тобто збільшення кількості зайнятого населення на 1% при незмінних основних засобах призведе до збільшення валового регіонального продукту наближено на 0,1113 %.

*Еластичність за величиною основних засобів*:

$$E_C = \beta = 0{,}9356,$$

тобто збільшення основних засобів на 1% при незмінній зайнятості населення призведе до збільшення валового регіонального продукту наближено на 0,9356%.

*Гранична норма еквівалентної заміни величини основних засобів кількістю зайнятого населення*:

$$h_{CL} = \frac{\beta}{\alpha} \cdot \frac{L}{C} = \frac{0{,}9356}{0{,}1113} \cdot \frac{L}{C} = 8{,}4061 \cdot \frac{L}{C}.$$

*Гранична норма еквівалентної заміни кількості зайнятих величиною основних засобів*:

$$h_{LC} = \frac{\alpha}{\beta} \cdot \frac{C}{L} = \frac{0{,}1113C}{0{,}9356L} = 0{,}1190\frac{C}{L}.$$

*Еластичність заміщення ресурсів*

$$\delta_{LC} = \frac{\partial \ln(L/C)}{\partial lnh}.$$

Можна довести, що для функції Кобба - Дугласа $\delta_{LC} = \delta_{CL} = 1$.

Розраховані значення показників ефективності використання кожного з ресурсів у регіонах України наведено в табл. 3.21.

Таблиця 3.21

Показники ефективності і використання основних засобів та зайнятих працівників

| № з/п | Область | Середня ефективність використання | | Гранична ефективність використання | | Гранична норма еквівалентної заміни | |
|---|---|---|---|---|---|---|---|
| | | **Кількості зайнятих** | **Основних засобів** | **Кількості зайнятих** | **Основних засобів** | **Основних засобів кількістю зайнятих** | **Кількості зайнятих основними засобами** |
| 1 | АР Крим | 14,1498 | 0,2909 | 1,5743 | 0,2722 | 5,7835 | 0,1729 |
| 2 | Вінницька | 10,5847 | 0,2935 | 1,1777 | 0,2746 | 4,2882 | 0,2332 |
| 3 | Волинська | 10,1192 | 0,2867 | 1,1259 | 0,2682 | 4,1972 | 0,2383 |
| 4 | Дніпропетровська | 23,8758 | 0,2885 | 2,6565 | 0,2699 | 9,8428 | 0,1016 |
| 5 | Донецька | 18,4340 | 0,2980 | 2,0510 | 0,2788 | 7,3562 | 0,1359 |
| 6 | Житомирська | 10,4397 | 0,2901 | 1,1615 | 0,2715 | 4,2787 | 0,2337 |
| 7 | Закарпатська | 8,3027 | 0,2941 | 0,9238 | 0,2752 | 3,3569 | 0,2979 |
| 8 | Запорізька | 20,2903 | 0,2826 | 2,2575 | 0,2644 | 8,5387 | 0,1171 |
| 9 | Івано-Франківська | 14,7284 | 0,2821 | 1,6387 | 0,2639 | 6,2088 | 0,1611 |
| 10 | Київська | 15,8408 | 0,2863 | 1,7625 | 0,2679 | 6,5787 | 0,1520 |
| 11 | Кіровоградська | 12,5666 | 0,2831 | 1,3982 | 0,2649 | 5,2780 | 0,1895 |
| 12 | Луганська | 14,8624 | 0,2918 | 1,6536 | 0,2730 | 6,0572 | 0,1651 |
| 13 | Львівська | 14,3821 | 0,2930 | 1,6002 | 0,2741 | 5,8377 | 0,1713 |
| 14 | Миколаївська | 14,0297 | 0,2836 | 1,5610 | 0,2654 | 5,8825 | 0,1700 |
| 15 | Одеська | 15,9845 | 0,2906 | 1,7785 | 0,2719 | 6,5410 | 0,1529 |
| 16 | Полтавська | 21,7929 | 0,2783 | 2,4247 | 0,2604 | 9,3125 | 0,1074 |
| 17 | Рівненська | 12,2846 | 0,2833 | 1,3668 | 0,2651 | 5,1557 | 0,1940 |
| 18 | Сумська | 14,0819 | 0,2835 | 1,5668 | 0,2653 | 5,9061 | 0,1693 |
| 19 | Тернопільська | 10,2344 | 0,2852 | 1,1387 | 0,2668 | 4,2673 | 0,2343 |
| 20 | Харківська | 17,1834 | 0,2923 | 1,9119 | 0,2734 | 6,9917 | 0,1430 |
| 21 | Херсонська | 10,9431 | 0,2869 | 1,2175 | 0,2684 | 4,5362 | 0,2204 |
| 22 | Хмельницька | 11,2817 | 0,2889 | 1,2552 | 0,2703 | 4,6436 | 0,2154 |
| 23 | Черкаська | 11,9076 | 0,2875 | 1,3249 | 0,2690 | 4,9248 | 0,2031 |
| 24 | Чернівецька | 10,3485 | 0,2837 | 1,1514 | 0,2654 | 4,3383 | 0,2305 |
| 25 | Чернігівська | 15,2417 | 0,2812 | 1,6958 | 0,2631 | 6,4450 | 0,1552 |
| 26 | м.Київ | 27,2130 | 0,2838 | 3,0278 | 0,2656 | 11,4008 | 0,0877 |
| 27 | м.Севастополь | 10,3954 | 0,2739 | 1,1566 | 0,2563 | 4,5132 | 0,2216 |

*Висновки.* Аналіз показників використання основних засобів та кількості зайнятих показав, що найефективніше трудові ресурси використовують у місті Києві (величина валового регіонального продукту, яка припадає на одного зайнятого, становить 27,2130 тис.грн/осіб (середня ефективність використання ресурсу)), Дніпропетровській (23,8758), Полтавській (21,7929) та Запорозькій (20,2903) областях. Найнижчими показники є у Закарпатській (8,3027), Волинській (10,1192), Тернопільській (10,2344), Чернівецькій (10,3485), Житомирській (10,4397), Вінницькій (10,5847) та Херсонській (10,9431) областях, а також у місті Севастополі (10,3954).

Найбільший приріст валового регіонального продукту при додатковому залученні кількості зайнятих можливий у місті Києві (при збільшенні кількості зайнятих на 3,0278 млн. грн. (гранична ефективність використання ресурсу)), Дніпропетровській (2,6565), Полтавській (2,4247) та Запорозькій (2,2575) областях.

Найменший приріст валового регіонального продукту при додатковому залученні кількості зайнятих спостерігається у Закарпатській (при збільшенні кількості зайнятих на 1 тис. осіб величина валового регіонального продукту зросте на 0,9238 млн. грн.), Волинській (1,1259), Тернопільській (1,1387) областях.

Серед регіонів, у яких потрібно найменшу величину основних засобів для еквівалентної заміни кількості зайнятих, можна виділити місто Київ (для еквівалентної заміни 1 тис. зайнятих необхідно 0,0877 млн. грн.), Дніпропетровську (0,1016), Полтавську (0,1074) та Запорізьку області. Такі самі висновки можна зробити і для величини основних засобів.

Коефіцієнт еластичності кожного з ресурсів – кількості зайнятих та величини основних засобів – для виробничої функції Кобба-Дугласа постійний і дорівнює відповідному коефіцієнту $\alpha = 0{,}1113$ та $\beta = 0{,}9356$.

### 3.3.3. Степенева модель виробничої функції чотирьох змінних

**Приклад 3.10**. Розглядається вибіркова залежність між рівнем продуктивності праці та фондомісткістю продукції, коефіцієнтом плинності робочої сили, рівнем втрат робочого часу та стажу праці:

Таблиця 3.22

| **Місяць** | **Продуктивність праці, гр. од/люд.год.** | **Фондомісткість продукції, гр.од.** | **Коефіцієнт плинності робочої сили, %** | **Рівень витрат робочого часу, %** | **Стаж, роки** |
|---|---|---|---|---|---|
| | $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
| *1* | *52* | *75* | *13,2* | *2,8* | *4,4* |
| *2* | *53* | *76* | *12,8* | *2,8* | *5,1* |
| *3* | *50* | *64* | *11,5* | *3* | *5,5* |
| *4* | *51* | *65* | *11,3* | *3,1* | *6* |
| *5* | *54* | *70* | *11,2* | *3,2* | *6,5* |
| *6* | *55* | *72* | *10* | *3,3* | *7* |
| *7* | *57* | *74* | *9,8* | *3,4* | *8* |
| *8* | *52* | *63* | *7,5* | *3,5* | *9* |
| *9* | *60* | *91,2* | *5* | *3,7* | *10,1* |
| *10* | *60* | *87,9* | *4,8* | *3,8* | *11* |
| *11* | *62* | *89* | *5* | *3,9* | *11,9* |
| *12* | *64* | *102* | *3,5* | *4* | *13* |
| *13* | *65* | *102* | *3,6* | *4,1* | *13,5* |
| *14* | *67* | *104* | *3,7* | *4,3* | *14* |
| *15* | *67* | *105* | *3,6* | *4,2* | *14,5* |
| *16* | *62* | *76* | *5,2* | *4,5* | *15* |
| *17* | *63* | *77* | *5* | *4,8* | *15,5* |
| ***18*** | ***66*** | ***87*** | ***4,0*** | ***4,9*** | ***17,0*** |
| ***19*** | ***68*** | ***90*** | ***4,0*** | ***5,0*** | ***16,5*** |
| ***20*** | ***70*** | ***92*** | ***3,0*** | ***4,7*** | ***17,5*** |
| ***21*** | | ***92*** | ***4,0*** | ***5,2*** | ***17,6*** |
| ***22*** | | ***93*** | ***5,0*** | ***5,3*** | ***17,7*** |
| ***23*** | | ***93*** | ***5,0*** | ***5,4*** | ***17,8*** |
| ***24*** | | ***94*** | ***6,0*** | ***5,4*** | ***17,9*** |

1) Побудувати економетричну модель продуктивності праці, що характеризує залежність між продуктивністю праці і основними чинниками, що впливають на неї (використати дані перших 17 спостережень таблиці 3.22).
2) За допомогою критерія Фішера для рівня значущості $\alpha = 0{,}05$ перевірити адекватність знайденої моделі.
3) Знайти основні економічні характеристики для знайденої виробничої функції і зробити економічний аналіз
4) Виконати прогноз продуктивності праці на наступні чотири місяці (21 - 24) для заданих в таблиці очікуваних значень чинників (використати 18-те, 19-те та 20-те спостереження).

***Розв'язування***

1) *Побудова економетричної моделі продуктивності праці.*

Ідентифікація змінних:

$Y$ – продуктивність праці (залежна змінна);

$X_1$ – фондомісткість продукції,

$X_2$ – коефіцієнт плинності робочого часу,

$X_3$ – процент втрат робочого часу,

$X_4$ – стаж роботи (незалежні змінні).

Специфікація моделі.

Загальний вигляд мультиплікативної моделі продуктивності:

$$\hat{Y} = a_0 X_1^{a_1} X_2^{a_2} X_3^{a_3} X_4^{a_4}.$$

Для лінеаризації моделі виконаємо логарифмування:

$$\ln \hat{Y} = \ln a_0 + a_1 \ln X_1 + a_2 \ln X_2 + a_3 \ln X_3 + a_4 \ln X_4 ,$$

Зробимо заміну:

$$\hat{y} = \ln \hat{Y},\ \ b_0 = \ln a_0 ,\quad x_j = \ln X_j \quad ( j = \overline{1{,}4} ).$$

Отримаємо таку лінійну регресійну модель:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4.$$

Запишемо логарифми вихідних даних:

$$y=lnY=\begin{pmatrix} 3{,}95124\\ 3{,}97029\\ 3{,}91202\\ 3{,}93183\\ 3{,}98898\\ 4{,}00733\\ 4{,}04305\\ 3{,}95124\\ 4{,}09435\\ 4{,}09435\\ 4{,}12713\\ 4{,}15888\\ 4{,}17439\\ 4{,}20469\\ 4{,}20469\\ 4{,}12713\\ 4{,}14314 \end{pmatrix} \qquad lnX=\begin{pmatrix} 4{,}31749 & 2{,}58022 & 1{,}02962 & 1{,}48161\\ 4{,}33073 & 2{,}54945 & 1{,}02962 & 1{,}62924\\ 4{,}15888 & 2{,}44235 & 1{,}09861 & 1{,}70475\\ 4{,}17439 & 2{,}42480 & 1{,}13140 & 1{,}79176\\ 4{,}24850 & 2{,}41591 & 1{,}16315 & 1{,}87180\\ 4{,}27667 & 2{,}30259 & 1{,}19392 & 1{,}94591\\ 4{,}30407 & 2{,}28238 & 1{,}22378 & 2{,}07944\\ 4{,}14313 & 2{,}01490 & 1{,}25276 & 2{,}19723\\ 4{,}51305 & 1{,}60944 & 1{,}30833 & 2{,}31254\\ 4{,}47620 & 1{,}56862 & 1{,}33500 & 2{,}39790\\ 4{,}48864 & 1{,}60944 & 1{,}36098 & 2{,}47654\\ 4{,}62497 & 1{,}25276 & 1{,}38629 & 2{,}56495\\ 4{,}62497 & 1{,}28093 & 1{,}41099 & 2{,}60269\\ 4{,}64439 & 1{,}30833 & 1{,}45862 & 2{,}63906\\ 4{,}65396 & 1{,}28093 & 1{,}43508 & 2{,}67415\\ 4{,}33073 & 1{,}64866 & 1{,}50408 & 2{,}70805\\ 4{,}34381 & 1{,}60944 & 1{,}56862 & 2{,}74084 \end{pmatrix}$$

*Оцінимо параметри лінеаризованої моделі* з використанням МНК та засобів *Excel*.

Оператор оцінювання параметрів лінеаризованої моделі:

$$B=(x^Tx)^{-1}x^Ty,$$

$$x=\begin{pmatrix} 1 & 4{,}317488 & 2{,}580217 & 1{,}029619 & 1{,}481605\\ 1 & 4{,}330733 & 2{,}549445 & 1{,}029619 & 1{,}629241\\ 1 & 4{,}158883 & 2{,}442347 & 1{,}098612 & 1{,}704748\\ 1 & 4{,}174387 & 2{,}424803 & 1{,}131402 & 1{,}791759\\ 1 & 4{,}248495 & 2{,}415914 & 1{,}163151 & 1{,}871802\\ 1 & 4{,}276666 & 2{,}302585 & 1{,}193922 & 1{,}94591\\ 1 & 4{,}304065 & 2{,}282382 & 1{,}223775 & 2{,}079442\\ 1 & 4{,}143135 & 2{,}014903 & 1{,}252763 & 2{,}197225\\ 1 & 4{,}513055 & 1{,}609438 & 1{,}308333 & 2{,}312535\\ 1 & 4{,}4762 & 1{,}568616 & 1{,}335001 & 2{,}397895\\ 1 & 4{,}488636 & 1{,}609438 & 1{,}360977 & 2{,}476538\\ 1 & 4{,}624973 & 1{,}252763 & 1{,}386294 & 2{,}564949\\ 1 & 4{,}624973 & 1{,}280934 & 1{,}410987 & 2{,}60269\\ 1 & 4{,}644391 & 1{,}308333 & 1{,}458615 & 2{,}639057\\ 1 & 4{,}65396 & 1{,}280934 & 1{,}435085 & 2{,}674149\\ 1 & 4{,}330733 & 1{,}648659 & 1{,}504077 & 2{,}70805\\ 1 & 4{,}343805 & 1{,}609438 & 1{,}568616 & 2{,}74084 \end{pmatrix}$$

$$x^T = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 4,317488 & 4,330733 & 4,158883 & \dots & 4,65396 & 4,330733 \\ 2,580217 & 2,549445 & 2,442347 & \dots & 1,280934 & 1,648659 \\ 1,029619 & 1,029619 & 1,098612 & \dots & 1,435085 & 1,504077 \\ 1,481605 & 1,629241 & 1,704748 & \dots & 2,674149 & 2,70805 \end{pmatrix}$$

$$x^T x = \begin{pmatrix} 17 & 74,65458 & 32,18115 & 21,89085 & 37,81844 \\ 74,65458 & 328,3341 & 140,1396 & 96,40882 & 166,872 \\ 32,18115 & 140,1396 & 64,92179 & 40,26989 & 68,43106 \\ 21,89085 & 96,40882 & 40,26989 & 28,62141 & 49,78572 \\ 37,81844 & 166,872 & 68,43106 & 49,78572 & 86,95359 \end{pmatrix}$$

$$(x^T x)^{-1} = \begin{pmatrix} 545,1318 & -76,95038 & -58,116727 & 2,733856908 & -45,2458 \\ -76,9504 & 11,7956119 & 7,43906832 & -0,169148534 & 5,073321 \\ -58,1167 & 7,43906832 & 7,97399288 & -11,2524478 & 11,1675 \\ 2,733857 & -0,1691485 & -11,252448 & 109,6228482 & -54,774 \\ -45,2458 & 5,07332115 & 11,1674961 & -54,77403871 & 32,52646 \end{pmatrix}$$

$$x^T y = \begin{pmatrix} 69,08474 \\ 303,6323 \\ 130,03 \\ 89,19316 \\ 154,3018 \end{pmatrix}$$

$$B = (x^T x)^{-1} x^T y = \begin{pmatrix} 1,0823 \\ 0,4692 \\ 0,1443 \\ 0,2262 \\ 0,1602 \end{pmatrix}$$

Лінеарізоване рівняння регресії має такий вигляд:

$$\hat{y} = 1,0823 + 0,4692\, x_1 + 0,1443\, x_2 + 0,2262\, x_3 + 0,1602\, x_4.$$

Виконаємо зворотне перетворення:

$$a_0 = exp(b_0) = e^{1,0823} = 2,9514;$$

$$a_1 = b_1 = 0,46923; \quad a_2 = b_2 = 0,1443; \quad a_3 = b_3 = 0,2262; \quad a_4 = b_4 = 0,1602.$$

Вибіркова виробнича функція Кобба-Дугласа набуває вигляду:

$$\hat{Y} = 2{,}9514X_1^{0{,}4692}X_2^{0{,}1443}X_3^{0{,}2262}X_4^{0{,}1602}. \qquad (3.22)$$

*Висновки.*

Оскільки $a_1+a_2+a_3+a_4 =$ 0,9999 < 1, то зміна кожного чинника на 1% сприятиме зростанню величини продуктивності праці менш, ніж на 1%. Отже, темп зростання продуктивності праці нижче, ніж темп зростання виробничих ресурсів (темп зростання продуктивності праці майже дорівнює темпу зростання виробничих ресурсів).

Збільшення на 1% витрат фондомісткості продукції при незмінних інших факторах (коефіцієнта плинності робочої сили, рівня витрат робочого часу і стажу роботи) обумовлює зростання продуктивності праці наближено на 0,4692 %.

Збільшення на 1% коефіцієнта плинності робочої сили при незмінних інших факторах (фондомісткості продукції, рівня витрат робочого часу і стажу роботи) обумовлює зростання продуктивності праці наближено на 0,1443 %.

Збільшення на 1% рівня витрат робочого часу при незмінних інших факторах (фондомісткості продукції, коефіцієнта плинності робочої сили і стажу роботи) обумовлює зростання продуктивності праці наближено на 0,2262 %.

Збільшення на 1% стажу роботи при незмінних інших факторах (фондомісткості продукції, коефіцієнта плинності рабочої сили і рівня витрат робочого часу) обумовлює зростання продуктивності праці наближено на 0,1602 %.

Коефіцієнт пропорційності $a_0$ = 2,9514 можна трактувати як величину, яка враховує всі якісні (що не має виразу в фондомісткості продукції, коефіцієнта плинності рабочої сили, рівня витрат робочого часу і стажу роботи) чинникі виробництва.

Результати регресійного аналізу для лінеарізованого рівняння регресії засобами Excel зображені на рисунку 3.19.

ВЫВОД ИТОГОВ

| Регрессионная статистика | |
|---|---|
| Множественный R | 0,999617095 |
| R-квадрат | 0,999234336 |
| Нормированный R- | 0,998979115 |
| Стандартная ошибк | 0,00320416 |
| Наблюдения | 17 |

Дисперсионный анализ

| | df | SS | MS | F | Значимость F |
|---|---|---|---|---|---|
| Регрессия | 4 | 0,160783 | 0,040196 | 3915,168 | 1,40942E-18 |
| Остаток | 12 | 0,000123 | 1,03E-05 | | |
| Итого | 16 | 0,160906 | | | |

| | Коэффициент | артная о | татисти | Р-Значение | Нижние 95% | ерхние 95 | жние 95,0 | Верхние 95,0% |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | 1,082285376 | 0,074811 | 14,46694 | 5,87E-09 | 0,919286382 | 1,245284 | 0,919286 | 1,245284371 |
| Переменная X 1 | 0,469231747 | 0,011005 | 42,63958 | 1,8E-14 | 0,445254772 | 0,493209 | 0,445255 | 0,493208722 |
| Переменная X 2 | 0,14429428 | 0,009048 | 15,94766 | 1,93E-09 | 0,124580402 | 0,164008 | 0,12458 | 0,164008159 |
| Переменная X 3 | 0,226225097 | 0,033548 | 6,743355 | 2,06E-05 | 0,153130598 | 0,29932 | 0,153131 | 0,299319595 |
| Переменная X 4 | 0,160233476 | 0,018274 | 8,768406 | 1,45E-06 | 0,120417941 | 0,200049 | 0,120418 | 0,20004901 |

Рис. 3.19

2) *Перевірка адекватності лінеарізованої моделі.*

Для оцінювання достовірності лінеарізованої моделі, обчислимо розрахункове значення $F$-критерія Фішера за формулою:

$$F_{\text{факт}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \cdot \frac{k_2}{k_1},$$

де $y_i$ і $\hat{y}_i$ - відповідно фактичне і розрахункове значення регресанду в лінеарізованій моделі; $\bar{y}$ - середнє значення регресанда; число ступенів вільності $k_1 = m - 1 = 4$, $k_2 = n - m = 12$ ( $n = 17$ – кількість досліджених місяців, $m = 5$ - число параметрів моделі).

Середнє значення регресанда $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} lnY_i}{n} = \frac{69{,}0847}{17} \approx 4{,}0638.$

Запишемо допоміжні обчислення в таблиці 3.23. Маємо:

$$F_{\text{факт}} = \frac{0{,}160783}{0{,}000123} \cdot \frac{12}{4} \approx 3915{,}168.$$

Таблиця 3.23

| Місяць | $Y_i$ | $y_i = lnY_i$ | $\hat{y}_i$ | $(\hat{y}_i - \bar{y})^2$ | $(\hat{y}_i - y_i)^2$ |
|---|---|---|---|---|---|
| *1* | *52* | *3,9512* | *3,9508* | *0,0128* | *1,73827E-07* |
| *2* | *53* | *3,9703* | *3,9763* | *0,0077* | *3,55933E-05* |
| *3* | *50* | *3,9120* | *3,9079* | *0,0243* | *1,72189E-05* |
| *4* | *51* | *3,9318* | *3,9340* | *0,0169* | *4,62825E-06* |
| *5* | *54* | *3,9890* | *3,9875* | *0,0058* | *2,27411E-06* |
| *6* | *55* | *4,0073* | *4,0032* | *0,0037* | *1,72662E-05* |
| *7* | *57* | *4,0431* | *4,0413* | *0,0005* | *3,17668E-06* |
| *8* | *52* | *3,9512* | *3,9526* | *0,0124* | *1,81263E-06* |
| *9* | *60* | *4,0943* | *4,0987* | *0,0012* | *1,90571E-05* |
| *10* | *60* | *4,0943* | *4,0952* | *0,0010* | *7,95684E-07* |
| *11* | *62* | *4,1271* | *4,1254* | *0,0038* | *2,87042E-06* |
| *12* | *64* | *4,1589* | *4,1578* | *0,0088* | *1,0853E-06* |
| *13* | *65* | *4,1744* | *4,1735* | *0,0120* | *7,18622E-07* |
| *14* | *67* | *4,2047* | *4,2032* | *0,0194* | *2,20818E-06* |
| *15* | *67* | *4,2047* | *4,2040* | *0,0197* | *4,21985E-07* |
| *16* | *62* | *4,1271* | *4,1265* | *0,0039* | *4,34385E-07* |
| *17* | *63* | *4,1431* | *4,1468* | *0,0069* | *1,34641E-05* |
| **сума** | ***994*** | ***69,0847*** | ***69,0847*** | ***0,1608*** | ***0,0001232*** |

При рівні значущості α = 0,05 знайдемо за таблицею 1 Додатку критичне значення

$$F_{кр}\,(\alpha;\, k_1;\, k_2) = F_{кр}\,(\alpha;\, m - 1;\, n - m) = F_{кр}\,(0{,}05;\, 4;\, 12) = 3{,}26.$$

Оскільки $F_{факт} > F_{кр}$, робимо висновок, що лінеарізована модель з надійністю 0,95 адекватна емпіричним даним.

Отримані результати збігаються з результатами, що наведені на рис. 3.19.

Такі ж результати розрахунку економетричної моделі (рис.3.20) можна отримати на основі стандартної програми «Линейн». Для цього спочатку треба виділити (5×5) масив для відповіді і скористатися *Мастер*'ом *функции* $\boldsymbol{f_x}$. Далі вибрати у категорії *Статистическая* функцію *ЛИНЕЙН*. У поле *Известные_значения_ у* ввести стовпець лінеарізованих значень *y*. Аналогічно у поле *Известные_значения_x* ввести матрицю (17×4) лінеарізованих значень $x = (x_1,\ldots,x_4)$. У полях *Конст* і *Статистика* – ввести 1 і одночасно натиснути *Ctrl + Shiift + Enter*. Перший рядок результатів розрахунку містить оцінки параметрів моделі ($lnb_0$ та $b_1, b_2, b_3, b_4$) Другий рядок таблиці містить стандартні похибки оцінок параметрів лінеарізованої моделі. Третій рядок містить два показника: коефіцієнт детермінації $R^2$ і дисперсію залишків. Четвертий рядок містить *F*-критерій та ступені вільності ($n - m$).

| | | | | | |
|---|---|---|---|---|---|
| **0,160233** | **0,226225** | **0,144294** | **0,469232** | 1,082285 | 2,951417 |
| 0,018274 | 0,033548 | 0,009048 | 0,011005 | 0,074811 | |
| **0,999234** | 0,003204 | #Н/Д | #Н/Д | #Н/Д | |
| **3915,168** | **12** | #Н/Д | #Н/Д | #Н/Д | |
| **0,160783** | **0,000123** | #Н/Д | #Н/Д | #Н/Д | |

Рис. 3.20

3) *Знаходження основних економічних характеристик для знайденої виробничої функції*

Обчислимо для знайденої моделі виробничої функції основні економічні характеристики взаємозв'язку.

*Середня ефективність впливу чинників на продуктивність праці* обчислюється за формулою:

$$\Pi_j = \frac{\overline{\hat{Y}}}{\overline{X}_j}, \quad j = \overline{1,4}.$$

Необхідні обчислення запишемо в таблиці 3.24. Отримаємо:

$$\Pi_1 = \frac{58,47026}{81,94706} = 0,71,$$

$$\Pi_2 = \frac{58,47026}{7,45529} = 7,85,$$

$$\Pi_3 = \frac{58,47026}{3,670588} = 15,93,$$

$$\Pi_4 = \frac{58,47026}{10} = 5,85.$$

Таблиця 3.24

| Місяць | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $\hat{Y}$ |
|---|---|---|---|---|---|
| 1 | *75* | *13,2* | *2,8* | *4,4* | *51,98* |
| 2 | *76* | *12,8* | *2,8* | *5,1* | *53,32* |
| 3 | *64* | *11,5* | *3* | *5,5* | *49,79* |
| 4 | *65* | *11,3* | *3,1* | *6* | *51,11* |
| 5 | *70* | *11,2* | *3,2* | *6,5* | *53,92* |
| 6 | *72* | *10* | *3,3* | *7* | *54,77* |
| 7 | *74* | *9,8* | *3,4* | *8* | *56,9* |
| 8 | *63* | *7,5* | *3,5* | *9* | *52,07* |
| 9 | *91,2* | *5* | *3,7* | *10,1* | *60,26* |
| 10 | *87,9* | *4,8* | *3,8* | *11* | *60,05* |
| 11 | *89* | *5* | *3,9* | *11,9* | *61,9* |
| 12 | *102* | *3,5* | *4* | *13* | *63,93* |
| 13 | *102* | *3,6* | *4,1* | *13,5* | *64,94* |
| 14 | *104* | *3,7* | *4,3* | *14* | *66,9* |
| 15 | *105* | *3,6* | *4,2* | *14,5* | *66,96* |
| 16 | *76* | *5,2* | *4,5* | *15* | *61,96* |
| 17 | *77* | *5* | *4,8* | *15,5* | *63,23* |
| Сума | *1393* | *126,7* | *62,4* | *170* | *994* |
| **Сер.Зн.** | ***81,95*** | ***7,453*** | ***3,671*** | ***10*** | ***58,47*** |

*Висновки*

Збільшення фондомісткості продукції на одну грошову одиницю при незмінних інших факторах (коефіцієнта плинності робочої сили, рівня витрат робочого часу і стажу роботи) у середньому обумовлює збільшення продуктивності праці наближено на 0,71 гр.од./люд.-год.

Збільшення на 1% коефіцієнта плинності робочої сили при незмінних інших факторах (фондомісткості продукції, рівня витрат робочого часу і стажу роботи) у середньому обумовлює збільшення продуктивності праці наближено на 7,85 грош.од./люд.-год.

Збільшення на 1% рівня витрат робочого часу при незмінних інших факторах (фондомісткості продукції, коефіцієнта плинності робочої сили і стажу роботи) у середньому обумовлює збільшення продуктивності праці наближено на 15,93 грош.од./люд.-год..

Збільшення на 1% стажу роботи при незмінних інших факторах (фондомісткості продукції, коефіцієнта плинності рабочої сили і рівня витрат робочого часу) в середньому обумовлює зростання продуктивності праці наближено на 5,85 грош.од./люд.-год.

*Гранична ефективність продуктивності праці для кожного виробничого ресурсу* обчислюється за формулами:

$$\frac{\partial \hat{Y}}{\partial X_1} = a_0 a_1 X_1^{a_1-1} X_2^{a_2} X_3^{a_3} X_4^{a_4} = a_1 \frac{\hat{Y}}{X_1};$$

$$\frac{\partial \hat{Y}}{\partial X_2} = a_0 X_1^{a_1} a_2 X_2^{a_2-1} X_3^{a_3} X_4^{a_4} = a_2 \frac{\hat{Y}}{X_2};$$

$$\frac{\partial \hat{Y}}{\partial X_3} = a_0 X_1^{a_1} X_2^{a_2} a_3 X_3^{a_3-1} X_4^{a_4} = a_3 \frac{\hat{Y}}{X_2};$$

$$\frac{\partial \hat{Y}}{\partial X_4} = a_0 X_1^{a_1} X_2^{a_2} X_3^{a_3} a_4 X_4^{a_4-1} = a_4 \frac{\hat{Y}}{X_4}.$$

Значення граничних ефективностей продуктивності праці для виробничих ресурсів запишемо в таблиці 3.25.

*Коефіцієнти еластичності продуктивності праці* обчислюються за формулою:

$$E_{Y/X_j} = \frac{\partial Y}{Y} : \frac{\partial X_j}{X_j} = \frac{\partial Y}{\partial X_j} : \frac{Y}{X_j}.$$

Для степеневої моделі виробничої функції коефіцієнти еластичності дорівнюють параметрам моделі $a_1$, $a_2$,$a_3$,$a_4$. Наприклад,

$$E_{\hat{Y}/X_1} = \frac{\partial \hat{Y}}{\partial X_1} : \frac{\hat{Y}}{X_1} = \frac{a_0 a_1 X_1^{a_1-1} X_2^{a_2} X_3^{a_3} X_4^{a_4} X_1}{a_0 X_1^{a_1} X_2^{a_2} X_3^{a_3} X_4^{a_4}} = a_1.$$

$$E_{\hat{Y}/X_1} = a_1 = 0{,}4692; \quad E_{\hat{Y}/X_2} = a_2 = 0{,}1443;$$

$$E_{\hat{Y}/X_3} = a_3 = 0{,}2262; \quad E_{\hat{Y}/X_4} = a_4 = 0{,}1602.$$

Таблиця 3.25

| Місяць | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $\hat{Y}$ | $\partial\hat{Y}/\partial X_1$ | $\partial\hat{Y}/\partial X_2$ | $\partial\hat{Y}/\partial X_3$ | $\partial\hat{Y}/\partial X_4$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 75 | 13,2 | 2,8 | 4,4 | 51,9783 | 0,3252 | 0,5682 | 4,1996 | 1,8929 |
| 2 | 76 | 12,8 | 2,8 | 5,1 | 53,3171 | 0,3292 | 0,6010 | 4,3077 | 1,6751 |
| 3 | 64 | 11,5 | 3 | 5,5 | 49,7930 | 0,3651 | 0,6248 | 3,7548 | 1,4506 |
| 4 | 65 | 11,3 | 3,1 | 6 | 51,1098 | 0,3690 | 0,6526 | 3,7298 | 1,3649 |
| 5 | 70 | 11,2 | 3,2 | 6,5 | 53,9186 | 0,3614 | 0,6947 | 3,8118 | 1,3292 |
| 6 | 72 | 10 | 3,3 | 7 | 54,7719 | 0,3570 | 0,7903 | 3,7548 | 1,2538 |
| 7 | 74 | 9,8 | 3,4 | 8 | 56,8985 | 0,3608 | 0,8378 | 3,7858 | 1,1396 |
| 8 | 63 | 7,5 | 3,5 | 9 | 52,0701 | 0,3878 | 1,0018 | 3,3656 | 0,9270 |
| 9 | 91,2 | 5 | 3,7 | 10,1 | 60,2625 | 0,3101 | 1,7391 | 3,6846 | 0,9560 |
| 10 | 87,9 | 4,8 | 3,8 | 11 | 60,0535 | 0,3206 | 1,8053 | 3,5752 | 0,8748 |
| 11 | 89 | 5 | 3,9 | 11,9 | 61,8950 | 0,3263 | 1,7862 | 3,5903 | 0,8334 |
| 12 | 102 | 3,5 | 4 | 13 | 63,9334 | 0,2941 | 2,6358 | 3,6158 | 0,7880 |
| 13 | 102 | 3,6 | 4,1 | 13,5 | 64,9449 | 0,2988 | 2,6031 | 3,5835 | 0,7708 |
| 14 | 104 | 3,7 | 4,3 | 14 | 66,9005 | 0,3018 | 2,6090 | 3,5197 | 0,7657 |
| 15 | 105 | 3,6 | 4,2 | 14,5 | 66,9565 | 0,2992 | 2,6837 | 3,6065 | 0,7399 |
| 16 | 76 | 5,2 | 4,5 | 15 | 61,9592 | 0,3825 | 1,7193 | 3,1148 | 0,6619 |
| 17 | 77 | 5 | 4,8 | 15,5 | 63,2316 | 0,3853 | 1,8248 | 2,9801 | 0,6537 |
| Сума | 1393,1 | 126,7 | 62,4 | 170 | 993,9945 | 5,7742 | 25,1775 | 61,9803 | 18,0774 |
| Сер.зн. | 81,9471 | 7,45294 | 3,67059 | 10 | 58,4703 | 0,3397 | 1,4810 | 3,6459 | 1,0634 |

*Граничні технологічні норми заміщення виробничих ресурсів* обчислюються за формулами:

$$h_{lj} = \frac{\partial\hat{Y}}{\partial X_l} : \frac{\partial\hat{Y}}{\partial X_j} \quad (l, j = \overline{1, k}).$$

Звернемо увагу на те, що індекс $l$ – номер ресурсу, що замінюється; $j$ – номер ресурсу, що заміщує.

Граничні технологічні норми заміщення виробничих ресурсів $h_{l\,j}$ показують, на скільки одиниць необхідно збільшити використання $j$-го ресурсу при зменшенні на одну одиницю $l$-го ресурсу, щоб продуктивність праці залишилася незмінною.

Граничні технологічні норми заміщення виробничих ресурсів знаходяться за формулою:

$$h_{lj} = \frac{\partial\hat{Y}}{\partial X_l} : \frac{\partial\hat{Y}}{\partial X_j} = a_l \frac{\hat{Y}}{X_l} : a_j \frac{\hat{Y}}{X_j} = \frac{a_l}{a_j} \cdot \frac{X_j}{X_l} \quad (l, j = \overline{1, k}),$$

тобто дорівнюють добутку відношень еластичностей продуктивності праці випуску по $l$-ому та $j$-ому ресурсам і обсягів $j$-ого та $l$-ого ресурсів. Значення граничних технологічних норм заміщення виробничих ресурсів наведені у таблиці 3.26.

*Висновки*

При зменшенні фондомісткості продукції на одну грош.од. для незмінності продуктивності праці у середньому збільшиться плинність робочої сили на 4,55%, або рівень витрат робочого часу – на 10,86%, або стаж праці - на 3,14 роки (за умови, що решта ресурсів не зміняться).

При зменшенні коефіцієнту плинності робочої сили на 1% для незмінності продуктивності праці у середньому збільшиться фондомісткість продукції на 0,3227 грош.од., або рівень витрат робочого часу – на 3,49%, або стаж праці – на 1,16 роки (за умови, що решта ресурсів не зміняться).

При зменшенні рівня витрат робочого часу на 1% для незмінності продуктивності праці у середньому збільшиться фондомісткість продукції на 0,0942 грош.од., або коефіцієнт плинності робочої сили – 0,03%, або стаж праці – на 0,29 роки (за умови, що решта ресурсів не зміняться).

При зменшенні стажу праці на 1 рік для незмінності продуктивності праці у середньому збільшиться фондомісткість продукції на 0,3524 грош.од., або коефіцієнт плинності робочої сили – на 1,75 %, або рівень витрат робочого часу – на 3,71 % (за умови, що решта ресурсів не зміняться).

4). *Оцінювання очікуваного рівня продуктивності праці*

Оскільки прийняття рішень завжди пов'язане з оцінюванням прогнозного (очікуваного) значення, визначимо прогнозні якості моделі. Для цього з 20-ти спостережень вихідної інформації для побудови моделі продуктивності праці використано 17.

Таблиця 3.26

| Місяць | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $h_{12}$ | $h_{13}$ | $h_{14}$ | $h_{21}$ | $h_{23}$ | $h_{24}$ | $h_{31}$ | $h_{32}$ | $h_{34}$ | $h_{41}$ | $h_{42}$ | $h_{43}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 75 | 13,2 | 2,8 | 4,4 | 1,7472 | 12,9139 | 5,8207 | 0,5723 | 7,3911 | 3,3314 | 0,0774 | 0,0238 | 0,4507 | 0,1718 | 0,3002 | 2,2186 |
| 2 | 76 | 12,8 | 2,8 | 5,1 | 1,8259 | 13,0861 | 5,0887 | 0,5477 | 7,1671 | 2,7870 | 0,0764 | 0,0235 | 0,3889 | 0,1965 | 0,3588 | 2,5716 |
| 3 | 64 | 11,5 | 3 | 5,5 | 1,7114 | 10,2852 | 3,9736 | 0,5843 | 6,0099 | 2,3219 | 0,0972 | 0,0299 | 0,3863 | 0,2517 | 0,4307 | 2,5884 |
| 4 | 65 | 11,3 | 3,1 | 6 | 1,7689 | 10,1089 | 3,6994 | 0,5653 | 5,7149 | 2,0914 | 0,0989 | 0,0304 | 0,3660 | 0,2703 | 0,4782 | 2,7326 |
| 5 | 70 | 11,2 | 3,2 | 6,5 | 1,9219 | 10,5463 | 3,6775 | 0,5203 | 5,4873 | 1,9134 | 0,0948 | 0,0292 | 0,3487 | 0,2719 | 0,5226 | 2,8678 |
| 6 | 72 | 10 | 3,3 | 7 | 2,2141 | 10,5189 | 3,5124 | 0,4517 | 4,7509 | 1,5864 | 0,0951 | 0,0292 | 0,3339 | 0,2847 | 0,6304 | 2,9948 |
| 7 | 74 | 9,8 | 3,4 | 8 | 2,3220 | 10,4932 | 3,1587 | 0,4307 | 4,5190 | 1,3603 | 0,0953 | 0,0293 | 0,3010 | 0,3166 | 0,7351 | 3,3220 |
| 8 | 63 | 7,5 | 3,5 | 9 | 2,5831 | 8,6781 | 2,3904 | 0,3871 | 3,3596 | 0,9254 | 0,1152 | 0,0354 | 0,2754 | 0,4183 | 1,0806 | 3,6305 |
| 9 | 91,2 | 5 | 3,7 | 10,1 | 5,6090 | 11,8836 | 3,0835 | 0,1783 | 2,1187 | 0,5497 | 0,0841 | 0,0259 | 0,2595 | 0,3243 | 1,8191 | 3,8540 |
| 10 | 87,9 | 4,8 | 3,8 | 11 | 5,6313 | 11,1522 | 2,7287 | 0,1776 | 1,9804 | 0,4846 | 0,0897 | 0,0276 | 0,2447 | 0,3665 | 2,0637 | 4,0869 |
| 11 | 89 | 5 | 3,9 | 11,9 | 5,4737 | 11,0022 | 2,5539 | 0,1827 | 2,0100 | 0,4666 | 0,0909 | 0,0280 | 0,2321 | 0,3916 | 2,1432 | 4,3079 |
| 12 | 102 | 3,5 | 4 | 13 | 8,9618 | 12,2940 | 2,6793 | 0,1116 | 1,3718 | 0,2990 | 0,0813 | 0,0250 | 0,2179 | 0,3732 | 3,3448 | 4,5885 |
| 13 | 102 | 3,6 | 4,1 | 13,5 | 8,7128 | 11,9942 | 2,5801 | 0,1148 | 1,3766 | 0,2961 | 0,0834 | 0,0256 | 0,2151 | 0,3876 | 3,3770 | 4,6488 |
| 14 | 104 | 3,7 | 4,3 | 14 | 8,6436 | 11,6605 | 2,5367 | 0,1157 | 1,3490 | 0,2935 | 0,0858 | 0,0264 | 0,2175 | 0,3942 | 3,4074 | 4,5967 |
| 15 | 105 | 3,6 | 4,2 | 14,5 | 8,9691 | 12,0530 | 2,4728 | 0,1115 | 1,3438 | 0,2757 | 0,0830 | 0,0255 | 0,2052 | 0,4044 | 3,6271 | 4,8742 |
| 16 | 76 | 5,2 | 4,5 | 15 | 4,4944 | 8,1424 | 1,7302 | 0,2225 | 1,8117 | 0,3850 | 0,1228 | 0,0378 | 0,2125 | 0,5780 | 2,5977 | 4,7062 |
| 17 | 77 | 5 | 4,8 | 15,5 | 4,7357 | 7,7340 | 1,6964 | 0,2112 | 1,6331 | 0,3582 | 0,1293 | 0,0398 | 0,2193 | 0,5895 | 2,7916 | 4,5591 |
| Сума | 1393,1 | 126,7 | 62,4 | 170 | 77,3259 | 184,5466 | 53,3829 | 5,4852 | 59,3949 | 19,7255 | 1,6007 | 0,4922 | 4,8748 | 5,9911 | 29,7082 | 63,1486 |
| **Сер.зн.** | **81,9471** | **7,4529** | **3,6706** | **10** | **4,5486** | **10,8557** | **3,1402** | **0,3227** | **3,4938** | **1,1603** | **0,0942** | **0,0290** | **0,2868** | **0,3524** | **1,7475** | **3,7146** |

Для оцінювання якості прогнозу на основі моделі, використаємо 18-те, 19-те та 20-те значення спостережень.

Для останніх трьох спостережень підставимо значення чинників в модель продуктивності праці (3.22) і знайдемо прогнозні значення продуктивності праці. Отримаємо:

$$\hat{Y}_{18} = 2{,}9514 \cdot 87^{0{,}4692} \cdot 4{,}0^{0{,}1443} \cdot 4{,}9^{0{,}2262} \cdot 17^{0{,}1602} \approx 66{,}1131;$$

$$\hat{Y}_{19} = 2{,}9514 \cdot 90^{0{,}4692} \cdot 4{,}0^{0{,}1443} \cdot 5{,}0^{0{,}2262} \cdot 16{,}5^{0{,}1602} \approx 67{,}1589;$$

$$\hat{Y}_{20} = 2{,}9514 \cdot 92^{0{,}4692} \cdot 3{,}0^{0{,}1443} \cdot 4{,}7^{0{,}2262} \cdot 17{,}5^{0{,}1602} \approx 64{,}7992.$$

Визначимо відхилення між фактичними і розрахунковими значеннями продуктивності праці:

$$u_{18} = 66 - 66{,}1131 = -0{,}1131;$$

$$u_{19} = 68 - 67{,}1589 = 0{,}8411;$$

$$u_{20} = 70 - 64{,}7992 = 5{,}2008.$$

Знайдемо похибки прогнозу. Обчислимо середньоарифметичну абсолютну похибку за формулою:

$$MAE = \frac{|u_{18}| + |u_{19}| + |u_{20}|}{3}.$$

$$MAE = \frac{|0{,}1131| + |0{,}8411| + |5{,}2008|}{3} \approx 2{,}0517.$$

Обчислимо середньоквадратичну похибку прогнозу за формулою:

$$MSE = \frac{(u_{18})^2 + (u_{19})^2 + (u_{20})^2}{3}$$

$$MSE = \frac{0{,}1131^2 + 0{,}8411^2 + 5{,}2008^2}{3} \approx 9{,}2562.$$

Відносна похибка прогнозу:

$$MAPE = \frac{1}{3} \cdot \frac{|u_{18}| + |u_{19}| + |u_{20}|}{Y_{18} + Y_{19} + Y_{20}} \cdot 100\%.$$

$$MAPE = \frac{0,1131 + 0,8411 + 5,2008}{3(66 + 68 + 70)} 100\% \approx 1,0057\ \%.$$

Коефіцієнт невідповідальності Тейла [23]:

$$K_T = \frac{\sqrt{\frac{(u_{18})^2 + (u_{19})^2 + (u_{20})^2}{3}}}{\sqrt{\frac{{Y_{18}}^2 + {Y_{19}}^2 + {Y_{20}}^2}{3}} + \sqrt{\frac{{\hat{Y}_{18}}^2 + {\hat{Y}_{19}}^2 + {\hat{Y}_{20}}^2}{3}}}.$$

$$K_T = \frac{\sqrt{\frac{0,1131^2 + 0,8411^2 + 5,2008^2}{3}}}{\sqrt{\frac{66^2 + 68^2 + 70^2}{3}} + \sqrt{\frac{66,11305^2 + 67,15885^2 + 64,79921^2}{3}}} \approx$$

$$\approx 0,022.$$

*Висновок*. Похибки прогнозу свідчать про те, що модель продуктивності праці (3.22) має великі прогнозні можливості - абсолютна та відносна похибки незначні за рівнем, а це вказує на добру апроксимацію моделі. Коефіцієнт невідповідальності Тейла наближається до нуля, що свідчить також про якісний прогноз.

Здійснимо прогноз продуктивності праці на наступні чотири місяці. Задамо очікувані значення чинників на цей період. Отримаємо:

$$\hat{Y}_{21} = 2,9514 \cdot 92^{0,4692} \cdot 4,0^{0,1443} \cdot 5,2^{0,2262} \cdot 17,6^{0,1602} \approx 69,1714\ ;$$

$$\hat{Y}_{22} = 2,9514 \cdot 93^{0,4692} \cdot 5,0^{0,1443} \cdot 5,3^{0,2262} \cdot 17,7^{0,1602} \approx 72,1737;$$

$$\hat{Y}_{23} = 2,9514 \cdot 93^{0,4692} \cdot 5,0^{0,1443} \cdot 5,4^{0,2262} \cdot 17,8^{0,1602} \approx 72,5450\ ;$$

$$\hat{Y}_{24} = 2,9514 \cdot 94^{0,4692} \cdot 6,0^{0,1443} \cdot 5,4^{0,2262} \cdot 17,9^{0,1602} \approx 74,9208\ .$$

Прогнозний рівень продуктивності праці може бути основою для прийняття рішень стосовно зміни цього показника на останній квартал року.

### 3.4. Побудова нелінійних регресій в Excel

Обчислити коефіцієнти нелінійної регресії в Excel можна одним із таких способів:

- використовуючи команду *Добавить линию тренда*;
- використовуючи команду *Поиск решения*.

Команда *Добавить линию тренда* використовується для виділення тренда (повільних змін) при аналізі часових рядів. Однак цю команду можна використовувати і для побудови рівняння регресії, розглядаючи як $t$ незалежну змінну $x$.

Ця команда дає змогу побудувати такі регресії:

- лінійну $\hat{y} = b_0 + b_1 x$;
- поліноміальну $\hat{y} = b_0 + b_1 x + \ldots + b_k x^k$ ( $k \leq 6$);
- логарифмічну $\hat{y} = b_0 + b_1 \ln x$;
- степеневу $\hat{y} = b_0 . x^{b_1}$;
- експоненціальну $\hat{y} = b_0 . e^{b_1 x}$.

Для побудови однієї з вищезазначених регресій треба виконати дії:

1. У вибраному аркуші Excel ввести стовпці вихідних даних $\{x_i, y_i\}$, $i = \overline{1, n}$ (таблиця 3.27).

Таблиця 3.27

| | | | | | | |
|---|---|---|---|---|---|---|
| *x* | *1* | *2* | *3* | *4* | *5* | *6* |
| *y* | *10* | *13,4* | *15,4* | *16,5* | *18,6* | *19,1* |

*2*. За цими даними побудувати графік (*Мастер диаграмм → Точечная,* Excel 97 - 2003, або *Вставка → Диаграмма →Точечная* ) в декартовій системі координат.

*3*. Установити курсор на побудованому графіку, клацнути правою кнопкою миші. У контекстному меню обрати команду *Добавить линию тренда*. На екрані з'явиться діалогове вікно *Линия тренда* (рис. 3.22 *а*) для Excel 97 - 2003), або *Формат лінії тренда* для наступних версій (рис. 3.22 *в*)).

4. На вкладці *Тип* діалогового вікна *Линия тренда* обрати тип лінії тренда. За замовчуванням активним є тип *Линейная*. У нашому випадку треба обрати рівняння регресії, що розглядається.

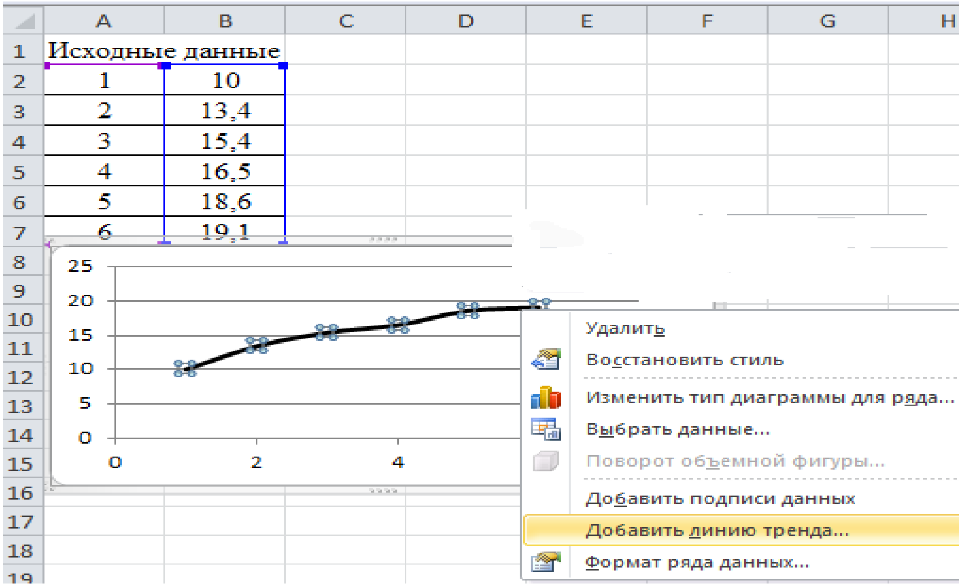
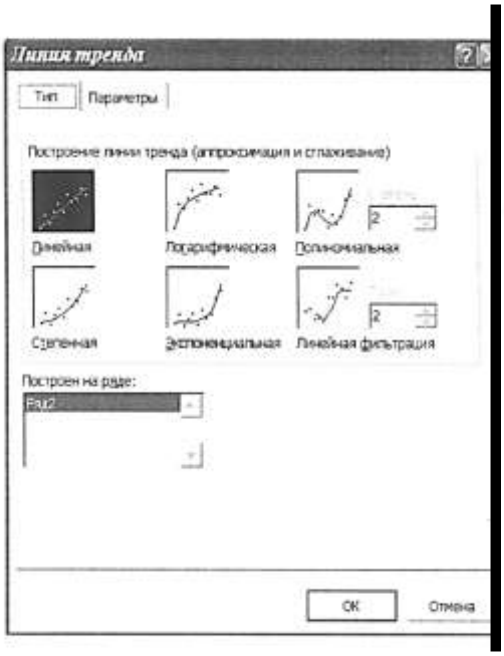



Рис. 3.21. Побудова графіку за початковими даними

5. На вкладці *Параметры* (Excel 97 - 2003) діалогового вікна *Лінія тренда* (рис. 3.22 б)) встановити параметри лінії тренда, тобто «увімкнути» (поставити прапорці) потрібні нам опції:

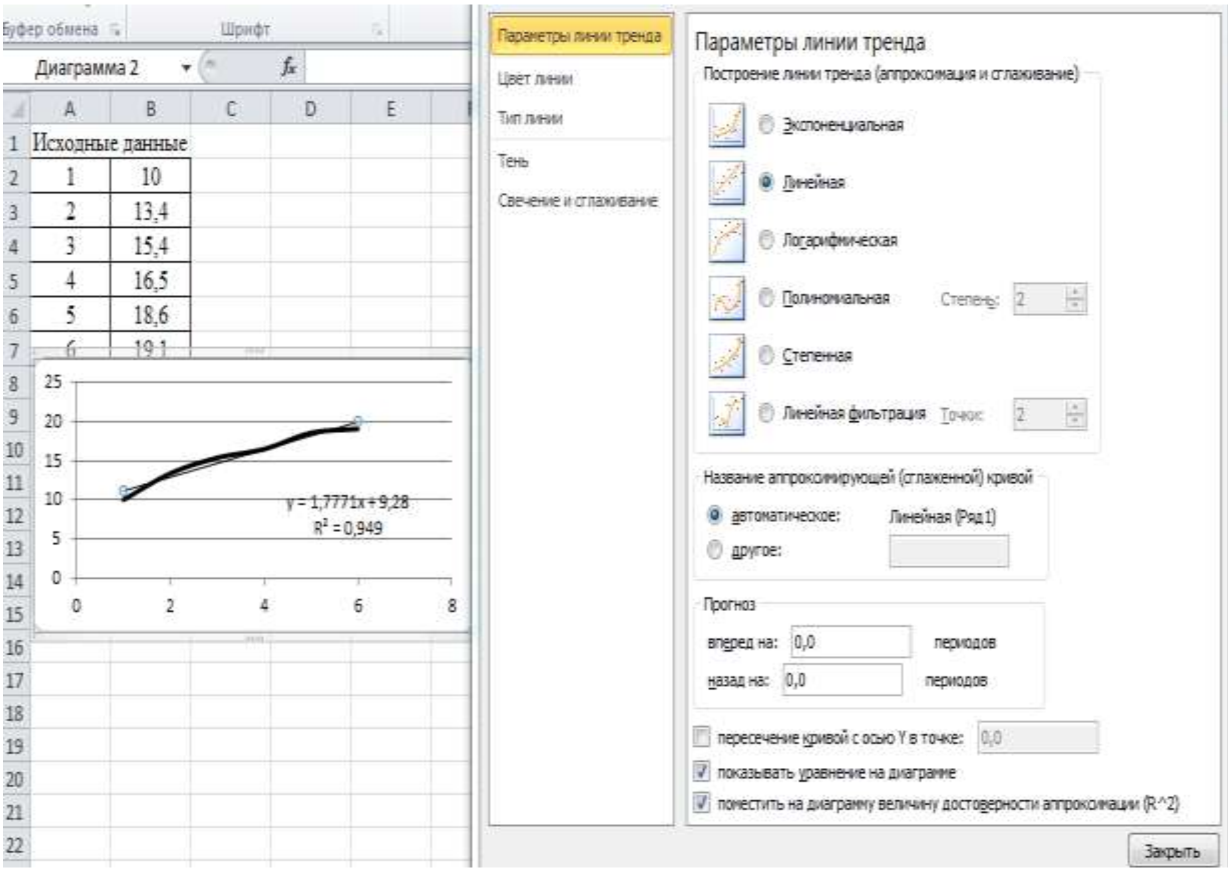
- «*Показывать уравнение на диаграмме*» - на діаграмі буде показано обране рівняння регресії з обчисленими коефіцієнтами;
- «Поместить на диаграмму величину достоверности аппроксимации (R^2)» - на діаграмі буде показано значення коефіцієнта детермінації $R^2$, який можна використовувати для перевірки значущості побудованої регресії за допомогою $F$-критерію. Якщо за побудованим рівнянням маємо зробити прогноз, то треба вказати число періодів прогнозу. Призначення інших опцій зрозумілі із їхніх назв.

6. Після виконання усіх вищезазначених дій, клацнути «ОК». На діаграмі з'явиться формула побудованого рівняння регресії і значення індексу детермінації.





а) Вибір рівняння регресії    б) Задання опцій виведення інформації




в) Вікно *Формат лінії тренда*

Рис. 3.22

**Приклад 3.11**. За даними таблиці 3.27 за допомогою команди *Добавить линию тренда* побудувати усі передбачені рівняння регресії і за значенням індексу детермінації обрати найкраще регресійне рівняння.

***Розв'язання***. Побудова кожного з п'яти рівнянь регресії виконується за вищезазначеними кроками 1 – 6.

Індекс детермінації характеризує близькість побудованої регресії до вихідних даних, які містять «небажану» випадкову складову ε. Очевидно, що взявши поліном п'ятої степені, отримаємо «ідеальне» значення $R^2 = 1$. Але таке регресійне рівняння містить у собі не тільки незалежну змінну $x$, а й складову ε, що знижує точність використання побудованого рівняння для прогнозу. Тому при обранні рівняння регресії треба мати на увазі не тільки величину коефіцієнта детермінації, а ще й «складність» регресійного рівняння, що визначається якістю коефіцієнтів рівняння. Таке урахування вдало реалізоване у скоригованому (виправленому) індексі детермінації:

$$\hat{R}^2 = 1 - \frac{n-1}{n-m} \cdot \left(1 - R^2\right),$$

де $m$ – кількість коефіцієнтів регресії, що обчислюються. Збільшення $m$ зменшує значення скоригованого індексу детермінації. Якщо кількість параметрів у регресійних рівняннях, що порівнюються, однакове, то відбір найкращого рівняння регресії можна виконати за величиною індексу детермінації.

Запишемо вигляди регресійних рівнянь, значення звичайних та скоригованих індексів детермінації у таблицю 3.28.

Порівняння величин індексів детермінації регресійних рівнянь виявляє найкраще з них - степеневу регресію

$$\hat{y} = 10{,}18 \cdot x^{0{,}3226}.$$

Згідно з величиною скоригованого індексу детермінації найкращою регресією також є степенева регресія.

Таблиця 3.28

| № | Рівняння | $R^2$ | $\hat{R}^2$ |
|---|---|---|---|
| 1 | $\hat{y} = 9{,}28+1{,}77\,x$ | *0,949* | *0,938* |
| 2 | $\hat{y} = 9{,}8759+5{,}1289 \cdot \ln x$ | *0,9916* | *0,9895* |
| 3 | $\hat{y} = 6{,}93+3{,}5396\,x - 0{,}2518\,x^2$ (поліноміальна, $p$ = 2) | *0,9896* | *0,9827* |
| 4 | $\hat{y} = 5{,}8333+4{,}9192\,x - 0{,}7087\,x^2 - 0{,}0435\,x^3$ (поліноміальна, $p$ = 3) | *0,9896* | *0,9827* |
| **5** | $\boldsymbol{\hat{y} = 10{,}18 \cdot x^{0{,}3226}}$ | ***0,9921*** | ***0,9901*** |
| 6 | $\hat{y} = 9{,}8675 \cdot e^{0{,}1225\,x}$ | *0,9029* | *0,8786* |

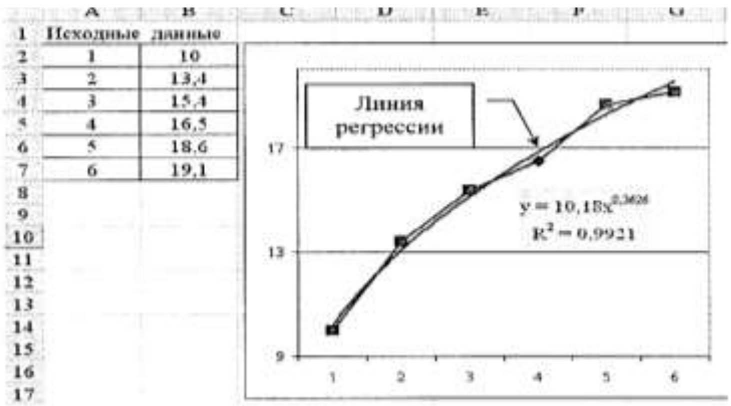



Рис. 3.23. Графік та рівняння побудованої степеневої регресії

*Команда* **Поиск решения** (*пункт меню Сервис для Excel 97 – 2003; для наступних версій – Данные → Поиск решения*). Використовується для обчислення параметрів, при яких деякий функціонал, що залежить від цих параметрів, досягає найменшого чи найбільшого значення.

Ця команда дозволяє також розв'язати *задачі умовної оптимізації*, тобто коли шукається мінімум або максимум функціонала із урахуванням додаткових обмежень (лінійних або нелінійних) на значення шуканих параметрів.

Наприклад, шуканий параметр $b$ має задовольняти обмеженню $0{,}2 \le b < 1$. Ця можливість обумовлює істотну перевагу розглянутого підходу у порівнянні з командою *Добавить линию тренда*.

Недоліком є необхідність програмувати «руками» обчислення індексу детермінації.

Використання команди *Поиск решения* для обчислення коефіцієнтів нелінійної регресії на основі МНК проілюстроване у наступному прикладі.

**Приклад 3.12**. За даними таблиці 3.27 побудувати рівняння степеневої регресії, використовуючи команду «*Поиск решения*».

***Розв'язання***.

1. Спочатку на аркуші Excel вводимо початкові дані: значення $x_i$ у комірках A2÷A7; значення $y_i$ в комірках B2÷B7.

2. У комірку B9 вводимо будь-яке значення коефіцієнта $b$. На рис. 3.24 зображений фрагмент аркуша Excel із введеними даними.

3. Далі обчислюємо за рівнянням регресії значення $\hat{y}_i = b_0 \cdot x_i^{b_1}$, $i = \overline{1,6}$. Для обчислення значення $\hat{y}_1$ у комірці C2 програмується вираз $B$9*A2^$B$10. Використання абсолютних адрес для комірок B9, B10 дозволяє «розмножити» (копіювати) цей вираз на комірки C3 – C7.

4. У комірках D2 – D7 обчислюється квадрат відхилу при відповідному значенні $x_i$. Так, у комірці D2 вводиться вираз

= (C2 - B2)^2 ,

який копіюється у комірки D3 – D7.

5. Значення функціоналу, що мінімізується МНК, обчислюється у комірці D9. На цьому необхідна для команди «Поиск решения» підготовка інформації завершується.

6. Для виконання команди *Поиск решения* треба звернутися до пункту основного меню ***Сервис*** та у меню клацнути мишею на команді *Поиск решения*. Далі у діалоговому вікні виконати нижчезазначені дії.

7. У полі введення *Установить целевую ячейку* ввести адресу комірки, в якій обчислюється значення функціоналу, що мінімізується (у нашому прикладі - D9).

|  | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Исходные данные |  | $\hat{y}_i$ | $(\hat{y}_i - y_i)^2$ |  |  |
| 2 | 1 | 10 | 1 | 81,000 |  |  |
| 3 | 2 | 13,4 | 1,231 | 148,081 |  |  |
| 4 | 3 | 15,4 | 1,390 | 196,269 |  |  |
| 5 | 4 | 16,5 | 1,516 | 224,529 |  |  |
| 6 | 5 | 18,6 | 1,621 | 288,298 |  |  |
| 7 | 6 | 19,1 | 1,712 | 302,351 |  |  |
| 8 |  |  |  |  |  |  |
| 9 | $b_0$ | 1 | $F(b_0, b_1)$ | 1240,528 |  |  |
| 10 | $b_1$ | 0,3 |  | =СУММ(D2:D7) |  |  |

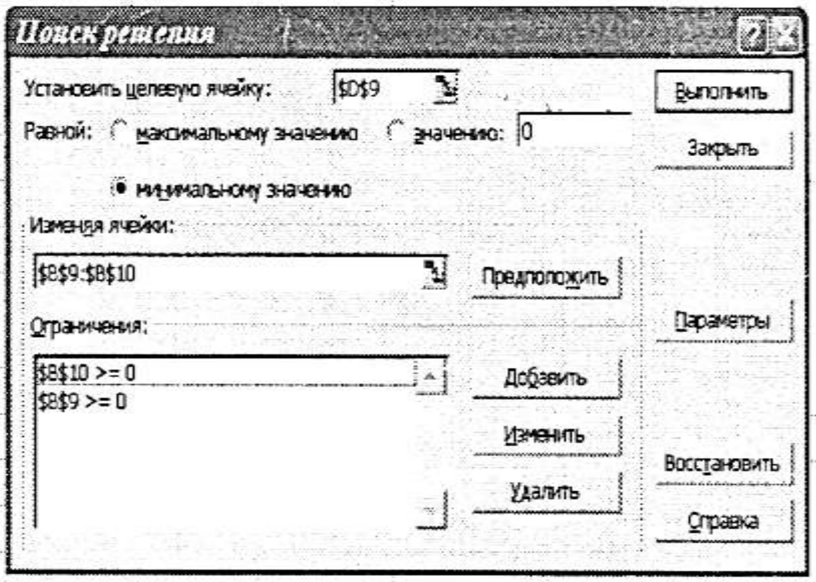



Рис. 3.24. Завдання параметрів команди *Поиск решения*

8. Вмикнути опцію *Минимальное значение* (йде пошук значень параметрів, при яких функціонал досягає свого мінімального значення).

9. У поле введення *Изменяя значения* ввести адреси комірок, де знаходяться значення шуканих параметрів (у прикладі це комірки B9, B10).

10. Клацнувши мишею на кнопку *Добавить*, формуємо обмеження на значення шуканих коефіцієнтів (у нашому прикладі це вимоги невід'ємності шуканих коефіцієнтів).

11. Після виконання наведених операцій клацнути на кнопку *Выполнить*. Починається пошук розв'язування введеної оптимізаційної задачі і через деякий час на екрані з'явиться нове діалогове вікно Результаты поиска решения.

12. Для збереження знайдених значень коефіцієнтів у відповідних комірках треба вибрати опцію *Сохранить найденное решение* і клацнути «ОК».

Обчислені значення коефіцієнтів знаходяться у комірках В9, В10 і дорівнюють: $b_0$ = 10,28299, $b_1$ = 0,354496. Комірка D9 містить значення функціоналу, що мінімізується. Знайдені у прикладі 3.11 значення параметрів степеневої регресії незначно відрізняються від значень, обчислених за допомогою команди *Добавить линию тренда*.

Використання табличного процесора Excel у двох розглянутих підходах дозволяє побудувати нелінійну парну регресію будь-якої «складності».

## 3.5. Завдання для самостійної роботи

### 3.5.1. Тести

*Вибрати правильну відповідь на запитання*

1. Регресійна модель вважається лінійною**,** коли вона:
   а) лінійна за змінними;
   б) лінійна за параметрами;
   в) лінійна за змінними та параметрами.
2. Для опису деякого економічного процесу придатні дві моделі. Обидві адекватні за *F*- критерієм Фішера. Треба надати перевагу моделі, у якій:
   а) більший коефіцієнт детермінації;
   б) менший коефіцієнт детермінації;
   в) більше значення *F*-критерію Фішера;
   г) менше значення *F*-критерію Фішера?
3. Для опису деякого економічного процесу прийнятні дві моделі. Обидві адекватні за *F*-критерієм Фішера. Треба надати перевагу моделі, у якій:

а) менше значення середньої похибки апроксимації $\bar{A}$;

б) більше значення середньої похибки апроксимації $\bar{A}$;

в) більше значення $F$-критерію Фішера;

г) менше значення $F$-критерію Фішера?

4. Досліджується відношення витрат споживача на товари ($y$) до його загального доходу ($x$). Модель для кривої витрат Енгеля:

а) $y = \beta_0 + \beta_1 x + \varepsilon$;

б) $y = \beta_0 + \beta_1 / x + \varepsilon$;

в) $y = \beta_0 + \beta_1 \ln x + \varepsilon$;

г) $\ln y = \ln \beta_0 + \beta_1 \ln x + \varepsilon$.

5. Крива Філліпса краще за все описується моделлю:

а) $\ln y = \ln \beta_0 + \beta_1 / x + \varepsilon$;

б) $\ln y = \ln \beta_0 + \beta_1 \ln x + \varepsilon$;

в) $y = \beta_0 + \beta_1 x + \varepsilon$;

г) $y = \beta_0 + \beta_1 / x + \varepsilon$.

6. До якого класу нелінійності відноситься модель

$$\hat{y}_i = b_0 + b_1 \cdot (1/x_i) + \varepsilon_i$$

а) регресія, яка нелінійна відносно пояснюючих змінних, але лінійна за параметрами, що оцінюються;

б) регресія, яка нелінійна за параметрами, що оцінюються;

в) регресія, яка нелінійна за залежною змінною.

7. Якщо при розрахунку параметрів параболи другого порядку отримаємо, що $b_1 < 0$ і $b_2 > 0$, то

а) крива симетрична відносно найвищої точки;

б) крива симетрична відносно найнижчої точки;

в) має місце повільно зростаюча функція з верхньою асимптотою.

8. Якщо при розрахунку параметрів параболи другого порядку отримаємо, що $b_1 > 0$ і $b_2 < 0$, то

а) крива симетрична відносно найвищої точки;

б) крива симетрична відносно найнижчої точки;

в) має місце повільно зростаюча функція з верхньою асимптотою.

9. Нелінійною формою залежності *y* від фактора(ів) не є рівняння

а) $y = \frac{1}{\beta_0+\beta_1 x} + \varepsilon$;

б) $y = \beta_0 + \beta_1 \cdot (1/x) + \varepsilon$;

в) $y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \varepsilon$;

г) $y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \varepsilon$.

10. Рівняння, що лінійне за параметрами, але нелінійне за змінними, є регресійною моделлю

а) $y = \beta_0 \cdot x^{\beta_1} \cdot \varepsilon$;

б) $y = \beta_0 \cdot \beta_1^x \cdot \varepsilon$;

в) $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$;

г) $y = \frac{1}{\beta_0+\beta_1 x} + \varepsilon$.

11. Рівняння, яке нелінійне за параметрами, є регресійною моделлю вигляду:

а) $y = \beta_0 + \beta_1 \cdot (1/x) + \varepsilon$;

б) $y = \beta_0 + \beta_1 \cdot \ln x + \varepsilon$;

в) $y = \frac{1}{\beta_0+\beta_1 x} + \varepsilon$;

г) $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$.

12. Визначте, яке з рівнянь найбільш точно описує залежність, що досліджується:

а) $\hat{y} = 6{,}79\, x^{-0,39}$; $R^2=0{,}94$;

б) $\hat{y} = 0{,}0424\, x^2 – 0{,}9029\, x +7{,}15$; $R^2=0{,}78$;

в) $\hat{y} = 2{,}3+5{,}7 \cdot (1/x)$; $R^2=0{,}94$;

г) $\hat{y} = 5{,}29\, e^{-0,06x}$; $R^2=0{,}66$.

13. Регресійна модель є нелінійною, якщо вона нелінійна

а) тільки за своїми змінними;

б) тільки за своїми параметрами;

в) і за своїми змінними, і за параметрами;

г) за своїми змінними, або за своїми параметрами.

### 3.5.2. Контрольні запитання

1. Записати в загальному вигляді нелінійну модель.
2. Класифікація нелінійних моделей.
3. Записати у загальному вигляді поліноміальну модель.
4. В яких випадках використовується квадратична функція?
5. Яку еволюцію відображає парабола, виходячи із значень її рівняння?
6. Записати загальний вигляд гіперболічної моделі.
7. В яких випадках використовується гіперболічна модель?
8. Записати рівняння кривої Філліпса та накреслити її графік.
9. Записати рівняння кривої Торнквіста та накреслити її графік.
10. Який вигляд має показникова (експоненціальна) модель в загальному вигляді?
11. До якого типу належать моделі $Y = 5{,}3\, x^{2{,}3}$ і $Y=5{,}3\cdot 2{,}3^{\,x}$ ?
12. Наведіть приклад гіперболічної моделі.
13. Що таке лінеаризація моделі?
14. Які методи лінеаризації нелінійних моделей вам відомі? В яких випадках вони використовуються?
15. Записати вигляд степеневої функції та накреслити її графіки залежно від значень параметрів.
16. Записати двофакторну модель виробничої функції Кобба – Дугласа та дати економічний сенс її параметрів.
17. Назвати середні значення ефективності ресурсів двофакторної виробничої функції Кобба–Дугласа, дати їх економічний сенс.
18. Назвати граничні значення ефективності ресурсів двофакторної виробничої функції Кобба - Дугласа та дати їх економічний сенс.
19. Записати формули граничних норм заміщення ресурсів та дати їх економічний сенс.
20. Записати формули еластичності заміщення ресурсів та дати їх економічний сенс.

### 3.5.3. Варіанти завдань контрольної роботи.

*Завдання 1.*

Досліджується залежність між витратами на купівлю продовольчих товарів $Y$ (% від загальних витрат) та середньоденною заробітною платою одного робітника $X$ (грош.од.) на основі вибіркових даних семи районів деякої області, що наведені у таблиці 3.29.

Таблиця 3.29

| Район | Доля витрат на купівлю продовольчих товарів у загальних витратах, %, ***y*** | Середньоденна заробітна плата одного працівника, грош.од., ***x*** |
|---|---|---|
| 1 | 68,8 +N | 45,1+ N |
| 2 | 61,2+ N | 59,0+ N |
| 3 | 59,9+ N | 57,2+ N |
| 4 | 56,7+ N | 61,8+ N |
| 5 | 55,0+ N | 58,8+ N |
| 6 | 54,3+ N | 47,2+ N |
| 7 | 49,3+ N | 55,2+ N |

($N$ – номер варіанту за списком журналу студентської групи).

Для характеристики залежності між витратами на купівлю продовольчих товарів $Y$ та середньоденною заробітною платою одного робітника $X$ знайти рівняння регресії в припущенні таких залежностей:

1) лінійної;
2) степеневої;
3) показникової;
4) рівнобічної гіперболи;
5) поліноміальної третьої степені.

Оцінити адекватність побудованих моделей.

*Завдання 2.*

Розглядається вибіркова залежність між рівнем продуктивності праці та фондомісткістю продукції, коефіцієнтом плинності робочої сили, рівнем втрат робочого часу та стажу праці (*N* - номер варіанту):

| Місяць | Продуктив-ність праці, гр.од./люд.-год. | Фондоміст-кість продукції, гр.од. | Коефіцієнт плинності робочої сили, % | Рівень витрат робочого часу, % | Стаж, роки |
|---|---|---|---|---|---|
| | $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
| 1 | 52+N | 75+N | 13,2+0,1·N | 2,8+0,1·N | 4,4+0,1·N |
| 2 | 53+N | 76+N | 12,8+0,1·N | 2,8+0,1·N | 5,1+0,1·N |
| 3 | 50+N | 64+N | 11,5+0,1·N | 3+0,1·N | 5,5+0,1·N |
| 4 | 51+N | 65+N | 11,3+0,1·N | 3,1+0,1·N | 6+0,1·N |
| 5 | 54+N | 70+N | 11,2+0,1·N | 3,2+0,1·N | 6,5+0,1·N |
| 6 | 55+N | 72+N | 10+0,1·N | 3,3+0,1·N | 7+0,1·N |
| 7 | 57+N | 74+N | 9,8+0,1·N | 3,4+0,1·N | 8+0,1·N |
| 8 | 52+N | 63+N | 7,5+0,1·N | 3,5+0,1·N | 9+0,1·N |
| 9 | 60+N | 91,2+N | 5+0,1·N | 3,7+0,1·N | 10+0,1·N |
| 10 | 60+N | 87,9+N | 4,8+0,1·N | 3,8+0,1·N | 11+0,1·N |
| 11 | 62+N | 89+N | 5+0,1·N | 3,9+0,1·N | 12+0,1·N |
| 12 | 64+N | 102+N | 3,5+0,1·N | 4+0,1·N | 13+0,1·N |
| 13 | 65+N | 102+N | 3,6+0,1·N | 4,1+0,1·N | 13,5+0,1·N |
| 14 | 67+N | 104+N | 3,7+0,1·N | 4,3+0,1·N | 14+0,1·N |
| 15 | 67+N | 105+N | 3,6+0,1·N | 4,2+0,1·N | 14,5+0,1·N |
| 16 | 62+N | 76+N | 5,2+0,1·N | 4,5+0,1·N | 15+0,1·N |
| 17 | 63+N | 77+N | 5+0,1·N | 4,8+0,1·N | 15,5+0,1·N |
| 18 | 66+N | 87+N | 4,0+0,1·N | 4,9+0,1·N | 17+0,1·N |
| 19 | 68+N | 90+N | 4,0+0,1·N | 5,0+0,1·N | 16,5+0,1·N |
| 20 | 70+N | 92+N | 3,0+0,1·N | 4,7+0,1·N | 17,5+0,1·N |
| 21 | | 92+N | 4,0+0,1·N | 5,2+0,1·N | 17,6+0,1·N |
| 22 | | 93+N | 5,0+0,1·N | 5,3+0,1·N | 17,7+0,1·N |
| 23 | | 93+N | 5,0+0,1·N | 5,4+0,1·N | 17,8+0,1·N |
| 24 | | 94+N | 6,0+0,1·N | 5,4+0,1·N | 17,9+0,1·N |

1) Побудувати економетричну модель продуктивності праці, що характеризує залежність між продуктивністю праці і основними чинниками, що впливають на неї.
2) За допомогою критерія Фішера для рівня значущості $\alpha = 0,05$ перевірити адекватність знайденої моделі.

3) Знайти основні економічні характеристики для знайденої виробничої функції і зробити економічний аналіз.
4) Виконати прогноз продуктивності праці на наступні чотири місяці, якщо задані такі очікувані значення чинників, що впливають на неї.

### 3.5.4. Завдання для індивідуальної роботи

Вибіркова сукупність даних про обсяг виробництва на підприємствах (*V*, грош.од.), виробничі фонди (*K*, од.) та вкладену працю (*L*, люд.-год.) наведені у таблицях.

1) Побудувати виробничу функцію Кобба - Дугласа.
2) Знайти основні економічні характеристики і зробити економічний аналіз на підставі даних таблиці.
3) За допомогою *F*-критерію Фішера на рівні значущості $\alpha = 0{,}05$ перевірити адекватність знайденої моделі.
4) Зробити прогноз обсягу виробництва для вказаних чинників.

| ***Варіанти №№*** | | |
|---|---|---|
| *V*(*565,76*) | | |
| ***V*** | ***K*** | ***L*** |
| *233* | *606* | *31* |
| *288* | *194* | *180* |
| *352* | *516* | *86* |
| *1029* | *1069* | *469* |
| *569* | *571* | *255* |
| *243* | *903* | *20* |
| *133* | *70* | *95* |
| *282* | *157* | *215* |
| *316* | *251* | *166* |
| *128* | *74* | *85* |
| *208* | *86* | *191* |
| *349* | *459* | *105* |

| ***Варіант № 2, 12, 22*** | | |
|---|---|---|
| ***V(62,52) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *596* | *41* | *242* |
| *2879* | *129* | *952* |
| *1006* | *56* | *382* |
| *1186* | *947* | *58* |
| *2319* | *88* | *981* |
| *1850* | *259* | *307* |
| *271* | *49* | *65* |
| *139* | *35* | *31* |
| *4433* | *616* | *565* |
| *437* | *28* | *185* |
| *4905* | *502* | *724* |
| *261* | *73* | *45* |

| ***Варіант № 3,13,23*** | | |
|---|---|---|
| ***V(48,66) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *105* | *153* | *27* |
| *364* | *764* | *68* |
| *282* | *73* | *462* |
| *108* | *30* | *139* |
| *721* | *798* | *296* |
| *388* | *134* | *455* |
| *457* | *876* | *104* |
| *138* | *256* | *25* |
| *259* | *772* | *34* |
| *201* | *309* | *52* |
| *124* | *30* | *194* |
| *475* | *320* | *309* |

| ***Варіант № 4,14,24*** | | |
|---|---|---|
| ***V(81,83) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *196* | *207* | *59* |
| *368* | *68* | *397* |
| *740* | *120* | *876* |
| *181* | *82* | *103* |
| *693* | *1322* | *132* |
| *532* | *47* | *998* |
| *158* | *65* | *95* |
| *195* | *198* | *59* |
| *134* | *33* | *105* |
| *151* | *129* | *50* |
| *388* | *88* | *393* |
| *787* | *267* | *582* |

| ***Варіант № 5,15,25*** | | |
|---|---|---|
| ***V(980,65) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *334* | *138* | *69* |
| *461* | *133* | *141* |
| *2539* | *891* | *481* |
| *1555* | *622* | *255* |
| *426* | *131* | *122* |
| *149* | *25* | *115* |
| *1200* | *1031* | *82* |
| *917* | *1039* | *45* |
| *499* | *216* | *88* |
| *762* | *966* | *34* |
| *336* | *93* | *111* |
| *410* | *56* | *312* |

| ***Варіант № 6,16,26*** | | |
|---|---|---|
| ***V(132,76) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *173* | *40* | *55* |
| *307* | *179* | *21* |
| *2894* | *703* | *348* |
| *592* | *46* | *624* |
| *163* | *46* | *40* |
| *699* | *324* | *54* |
| *690* | *482* | *27* |
| *466* | *334* | *21* |
| *1003* | *106* | *586* |
| *173* | *65* | *26* |
| *1579* | *157* | *783* |
| *863* | *183* | *178* |

| ***Варіант № 7,17,27*** | | |
|---|---|---|
| ***V(733,22) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *173* | *40* | *55* |
| *307* | *179* | *21* |
| *2894* | *703* | *348* |
| *592* | *46* | *624* |
| *163* | *46* | *40* |
| *699* | *324* | *54* |
| *690* | *482* | *27* |
| *466* | *334* | *21* |
| *1003* | *106* | *586* |
| *173* | *65* | *26* |
| *1579* | *157* | *783* |
| *863* | *183* | *178* |

| ***Варіант № 8,18,28*** | | |
|---|---|---|
| ***V(54,25) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *193* | *29* | *268* |
| *174* | *53* | *83* |
| *321* | *49* | *427* |
| *95* | *41* | *30* |
| *461* | *303* | *65* |
| *618* | *107* | *630* |
| *93* | *39* | *31* |
| *794* | *213* | *425* |
| *298* | *283* | *22* |
| *209* | *42* | *190* |
| *375* | *94* | *254* |
| *792* | *263* | *296* |

| ***Варіант № 9,19,29*** | | |
|---|---|---|
| ***V(285,82) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *275* | *72* | *444* |
| *234* | *538* | *27* |
| *411* | *259* | *231* |
| *297* | *197* | *163* |
| *673* | *897* | *171* |
| *322* | *336* | *100* |
| *232* | *170* | *103* |
| *273* | *670* | *30* |
| *910* | *724* | *436* |
| *898* | *1055* | *254* |
| *110* | *29* | *164* |
| *351* | *435* | *84* |

| ***Варіант № 10,20,30*** | | |
|---|---|---|
| ***V(131,819) -?*** | | |
| ***V*** | ***К*** | ***L*** |
| *4801* | *92* | *723* |
| *3844* | *432* | *78* |
| *6528* | *119* | *889* |
| *3136* | *73* | *525* |
| *7769* | *1163* | *71* |
| *2804* | *104* | *257* |
| *2856* | *120* | *280* |
| *297* | *29* | *38* |
| *1219* | *128* | *52* |
| *7754* | *956* | *87* |
| *4640* | *139* | *385* |
| *1913* | *49* | *357* |

# ЧАСТИНА 2
# РОЗДІЛ 4. МУЛЬТИКОЛІНЕАРНІСТЬ

## 4.1. Поняття мультиколінеарності та причини її виникнення

Як зазначалось у розділі 2, однією з умов використання МНК для знаходження параметрів економетричної моделі є те, що пояснювальні змінні у матриці $X$ мають бути незалежними між собою, тобто визначник $\det(X^T X) \neq 0$ (припущення 6°). Проте на практиці можуть мати місце випадки, коли пояснювальні змінні пов'язані між собою, що стає перешкодою до використання МНК.

Термін «мультиколінеарність» вперше було впроваджено у 1934 році Р. Фрішем. Явище мультиколінеарності є дуже поширеним для масивів статистичної інформації, за якими визначаються кількісні характеристики взаємозв'язку між економічними показниками.

*Мультиколінеарність* [23] - це явище, яке полягає в тому, що в множинній регресійній моделі існує тісна лінійна залежність, або сильна кореляція між двома або більше пояснювальними змінними, тобто парні коефіцієнти кореляції між незалежними змінними прямують до одиниці:

$$r_{x_i x_j} = \frac{cov(x_i, x_j)}{\sqrt{var(x_i) var(x_j)}} \to 1, \qquad i \neq j.$$

Так, при дослідженні залежності між ціною акції, дивідендами на акцію та отриманим прибутком на акцію виникає проблема мультиколінеарності, оскільки дивіденди та отриманий прибуток на одну акцію мають високий ступінь кореляції. Виникає ситуація, коли два колінеарних фактори змінюються в одному напрямку. Це робить неможливим оцінювання впливу кожного з факторів на досліджуваний показник. Мультиколінеарність у матричному вигляді – це залежність між стовпцями матриці факторних змінних $X$, отже, однією з основних причин наявності мультиколінеарності у моделі множинної регресії є «погана» матриця факторних змінних $X$.

Розрізняють функціональну та стохастичну (імовірнісну) форми мультиколінеарності. При *функціональній формі* мультиколінеарності в моделі повинен бути присутнім хоча б один фактор, що пов'язаний функціональною залежністю з будь-яким іншим фактором моделі або зі всіма іншими, тобто

$$x_i = \sum_{j \neq i} c_j x_j.$$

У випадку, коли у багатофакторній регресійній моделі існують декілька лінійно-залежних факторів, то має місце *досконала мультиколінеарність*. У випадку класичної лінійної множинної регресії одним з припущень є відсутність досконалої мультиколінеарності. Це пов'язано з тим, що при наявності досконалої мультиколінеарності неможливо визначити параметри лінійного рівняння регресіх.

Очевидно, що з пар колінеарних факторів один фактор слід виключити з моделі, щоб матриця $X^T X$ стала невиродженою (тобто якщо $\det(X^T X) \neq 0$), а система рівнянь для параметрів моделі мала розв'язок. Так, якщо $x_1$ – сімейні витрати на харчування, $x_2$ – витрати на комунальні послуги, $x_3$ – інші витрати, а $x_4 = x_1 + x_2 + x_3$ – загальні витрати ($c_1 = c_2 = c_3 = 1$), то один із факторів виявляється зайвим. Коефіцієнт кореляції між колінеарними факторами дорівнює -1 або 1

Матриця $X$ має порядок $(n \times (k+1))$. Однією з передумов до лінійної $k$-факторної регресії є вимога: $rang(X) = k + 1$. Якщо $rang\ (X) < k + 1$, то матриця $(X^T X)^{-1}$ буде виродженою (тобто при $\det(X^T X) = 0$). У цьому випадку в моделі існує мультиколінеарність у функціональній формі. Отже, при такій ситуації неможливо буде одержати статистичні оцінки для параметрів моделі. Однак на практиці така мультиколінеарність з'являється досить рідко і легко виявляється.

У моделях економічного змісту мультиколінеарність найчастіше проявляється у *стохастичній формі*, коли між факторами моделі існує тісний кореляційний зв'язок, який, проте, не досягає рівня функціонального.

У цьому випадку матриця $X^T X$ не буде виродженою, але її визначник набуває дуже малого значення. За формулою (2.5)

$$b = (X^T X)^{-1} X^T Y$$

та пропорційністю статистичних оцінок оберненої матриці $(X^T X)^{-1}$ елементи статистичних оцінок $b$, $cov(b \cdot b^T)$ будуть обернено пропорційними величині визначника $\det(X^T X)$. Внаслідок цього одержимо великі значення середніх квадратичних відхилень оцінок для параметрів регресійного рівняння $b_0, b_1, b_2, \ldots, b_k$.

*Причини виникнення мультиколінеарності* [11]

1. Використання малої скінченної сукупності спостережень.
2. Наявність вираженої тенденції зміни пояснювальних змінних в часі. На макроекономічні показники впливають однакові фактори. Це приводить до того, що вони відображають широкий спектр моделей однакової економічної ситуації. У період бумів або швидкого економічного зростання базові економічні показники також зростають. Так, доход, споживання, накопичення, інвестиції, ціни, зайнятість мають тенденцію до зростання в період економічної експансії і до спаду в період регресу. Наприклад, коли має місце зростання значень двох чи більше пояснювальних змінних, або ж зростання значень однієї пояснювальної змінної і зменшення іншої навіть тоді, коли економічна теорія не передбачає такої зміни.

Таким чином, причиною мультиколінеарності є сама наявність трендів у динамічних рядах (рядах однорідних статистичних величин, що показують зміну будь-якого явища в часі).

3. Наявність лагових змінних у моделі (запізнюючих змінних; змінні, значення яких відстають на один або кілька періодів часу) без чого неможливо обійтись у динамічних моделях, коли між зв'язком показників існує зрушення в часі. Наприклад, в інвестиційних

функціях лагові значення минулого рівня економічної активності вводяться як окремі змінні. У функціях споживання витрати на споживання у попередньому періоді вводяться в модель поряд з величиною поточного рівня доходу.

## 4.2. Наслідки мультиколінеарності

При виконанні основних припущень 1°- 7° оцінки параметрів регресії, що обчислені за МНК, мають найменшу дисперсію та є лінійними функціями від спостережуваних значень. Така властивість оцінок параметрів регресії зберігається при дуже високій, але недосконалій мультиколінеарності. При побудові моделі (навіть в умовах мультиколінеарності) можна отримати незміщені та стійкі оцінки. Ефект мультиколінеарності виявляється у тому, що *складно отримати значення параметрів з малою стандартною помилкою*.

Той самий ефект спостерігається при невеликій кількості спостережень або при невеликій зміні значень. Теоретично мультиколінеарність та невелика кількість змінних є однією й тією ж проблемою. Мультиколінеарність часто виникає в емпіричних дослідженнях, створюючи серйозні проблеми, тому недоцільно розглядати її тільки як порушення моделі класичної лінійної регресії. Властивість незміщеності фактично означає, що при багатократному повторенні спостережень зі сталими обсягами вибірок середнє значення оцінок буде наближатися до їх істинного значення. Проте здійснювати повторні спостереження в незмінних умовах в економіці практично неможливо.

*Теоретичні наслідки мультиколінеарності*

1. При наявності високої мультиколінеарності оцінки параметрів множинної регресійної моделі, що знайдені за допомогою МНК, залишаються незміщеними.

2. Мультиколінеарність може не впливати на ефективність оцінок параметрів моделі, проте це не означає, що дисперсія оцінки буде відносно малою порівняно із значенням параметра регресії.

3. Оцінки параметрів рівняння регресії, що отримані за допомогою МНК, та їх стандартні помилки стають нестійкими, дуже чутливими до найменших змін даних вибірки.

*Практичні наслідки мультиколінеарності*

1. Велика дисперсія і коваріація оцінок параметрів рівняння регресії, що обчислені за МНК. Це ускладнює знаходження істинних значень досліджуваних величин і розширює інтервали оцінок, що погіршуює їх точність.

2. Збільшення інтервалу довіри для параметрів рівняння регресії. Це пов'язано із збільшенням коефіцієнту кореляції, яке призводить до збільшення значень середньоквадратичних відхилень параметрів.

3. Незначущість $t$-статистики Ст'юдента. При наявності явища мультиколінеарності зменшується $t$-статистика, що може спонукати до неправильного висновку про існування впливу відповідних пояснювальних змінних на залежну змінну.

4. Ускладнюється визначення внеску кожної з пояснювальних змінних в рівняння множинної регресії.

5. Знак коефіцієнта регресії може виявиться неправильним.

*Причини невключення в модель мультиколінеарних факторів* [19]

1. Основна гіпотеза про незначущість параметрів рівняння множинної регресії може підтвердитись, але сама модель регресії при перевірці за допомогою $F$-критерію Фішера виявляється значущою, що свідчить про завищену величину коефіцієнта множинної кореляції.

2. Одержані оцінки коефіцієнтів моделі множинної регресії можуть бути невиправдано завищені або мати неправильні знаки.

3. Додавання або виключення з вихідних даних одного-двох спостережень робить сильний вплив на оцінки параметрів моделі.

4. Мультиколінеарні фактори, що включені в модель множинної регресії, здатні зробити її непридатною для подальшого застосування.

*Мультиколінеарність не є проблемою, якщо єдиною метою регресійного аналізу є прогноз* (оскільки чим більше значення коефіцієнта детермінації, тим точніше прогноз). Якщо метою аналізу є дійсне значення параметрів, то мультиколінеарність перетворюється на проблему, оскільки її наявність призводить до значних стандартних похибок оцінок параметрів.

### 4.3. Деякі ознаки наявності мультиколінеарності

Конкретних методів виявлення мультиколінеарності не існує, а прийнято застосовувати *ряд емпіричних прийомів* [23]. У більшості випадків множинний регресійний аналіз починається з розгляду матриці ($X^TX$) та кореляційної матриці факторних змінних $r_{xx}$ :

$$r_{xx}=\begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & \dots & r_{x_1x_k} \\ r_{x_2x_1} & r_{x_2x_2} & \dots & r_{x_2x_k} \\ \dots & \dots & \dots & \dots \\ r_{x_kx_1} & r_{x_kx_2} & \dots & r_{x_kx_1} \end{pmatrix} = \begin{pmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_k} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_k} \\ \dots & \dots & \dots & \dots \\ r_{x_kx_1} & r_{x_kx_2} & \dots & 1 \end{pmatrix}.$$

Коли до моделі входять більш, ніж дві пояснювальні змінні, вивчення питання про мультиколінеарність не може обмежуватись інформацією, що її містять елементи кореляційної матриці. Явище мультиколінеарності в жодному разі не зводиться лише до існування парної кореляції між незалежними змінними.

Більш загальна перевірка передбачає знаходження визначника (детермінанта) $det$ ($r_{xx}$) кореляційної матриці $r_{xx}$. Зазначимо, що величина $det\,(r_{xx})\in[0;\,1]$.

*Перша ознака*. Обчислення визначника для матриці ($X^TX$). Якщо величина визначника $\det(X^TX) = 0$, або близька до нуля, то це є незаперечним доказом наявності лінійной залежності, а тим самим існує мультиколінеарність між факторами.

При розгляді матриці коефіцієнтів парної кореляції між пояснювальними змінними з метою виявлення мультиколінеарних факторів керуються наступними правилами.

*Друга ознака*. Якщо величина визначника $\det(r_{xx}) = 0$, то існує повна мультиколінеарність. У випадку $\det(r_{xx})=1$ мультиколінеарність вітсутня. Чим ближче визначник $\det(r_{xx})$ до нуля, тим певніше можна стверджувати про наявність мультиколінеарності між факторами.

Незважаючи на те, що на величину визначника $\det(r_{xx})$ впливає дисперсія пояснювальних змінних, цей показник можна вважати точковою мірою мультиколінеарності. В міру зростання тісноти зв'язку між факторами моделі $\det(r_{xx})$ зменшується від 1 до 0.

*Третя ознака*. Якщо при обчисленні власних чисел кореляційної матриці факторних змінних $\lambda_{min}$ і $\lambda_{max}$ значення $\lambda_{min} < 10^{-5}$, то в моделі наявна мультиколінеарність. У випадку, коли $\lambda_{min} / \lambda_{max} < 10^{-5}$ також роблять висновок про наявність в моделі мультиколінеарності.

*Четверта ознака*. Наявність мультиколінеарності дозволяє виявити також перевірка парних коефіцієнтів кореляції. Якщо значення хоча б одного коефіцієнта кореляції більше 0,8, то мультиколінеарність є серйозною проблемою. Проблемним у цій ознаці є те, що високе значення парних коефіцієнтів кореляції є достатньою, але не є необхідною умовою наявності мультиколінеарності. Вона може мати місце навіть при відносно невеликих значеннях парних коефіцієнтів кореляції у більш, ніж двофакторній регресійній моделі.

*П'ята ознака*. Якщо значення коефіцієнта частинної детермінації, який обчислено для регресійних залежностей між $k$-им фактором та рештою, близьке до одиниці, то це свідчить про наявність мультиколінеарності.

*Шоста ознака*. Якщо в економетричній моделі знайдено мале значення оцінки параметра $b_k$ за високого рівня частинного коефіцієнта детермінації і водночас $F$-криткрій істотно відрізняється від нуля, то це також може свідчити про наявність мультиколінеарності.

*Сьома ознака*. «Класичною» ознакою мультиколінеарності є наявність одночасно великого значення коефіцієнта детермінації і незначущості $t$-статистики. Незначущість $t$-статистики означає, що один або більше оцінених параметрів статистично незначуще відрізняються від нуля. З іншого боку, якщо значення коефіцієнта детермінації велике, то можна з великою ймовірністю прийняти $F$-критерій, який відкидає нульову гіпотезу ($H_0$: $b_1 = b_2 = \ldots = b_k = 0$). Суперечність свідчить про наявність мультиколінеарності.

*Восьма ознака*. Якщо під час побудови економетричної моделі на основі покрокових процедур відбору найбільш інформативних змінних [13] істотно змінюються оцінки параметрів моделі за незначного підвищення (або зниження) коефіцієнтів кореляції та детермінації, то ця зміна перебуває у лінійній залежності від інших, які були введені до моделі раніше.

Існують й інші ознаки, але жодна з них не є універсальною, бо не проводить чіткої межі між тим, що треба вважати «суттєвою» мультиколінеарністю, яку необхідно враховувати, і тим, коли нею можна знехтувати.

## 4.4. Алгоритм Феррара - Глобера

Мультиколінеарність пояснювальних змінних призводить до зміщення статистичних оцінок для параметрів лінійної регресії. Постає питання про адекватність побудованої моделі. Завдяки цьому необхідно досліджувати на мультиколінеарність кожне рівняння за кожним фактором окремо. В економетричних задачах для дослідження наявності мультиколінеарності широко застосовується метод Феррара - Глобера, який має три види статистичних критеріїв, за якими перевіряється мультиколінеарність:

1) всього масиву незалежних змінних (критерій Пірсона $\chi^2$);
2) кожної незалежної змінної з рештою змінних ($F$-крітерій);
3) кожної пари незалежних змінних ($t$-критерій).

*Алгоритм методу Феррара - Глобера* [9]

*Крок 1*. Нормалізація змінних $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k$ економетричної моделі за формулою:

$$x_{ij}^* = \frac{x_{ij} - \overline{x_j}}{\sigma_{x_j}} \quad \left(i = \overline{1,n};\ j = \overline{1,k}\right), \tag{4.1}$$

де $n$ – кількість спостережень;

$k$ – кількість незалежних змінних;

$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$ - середня арифметична $j$-ої незалежної змінної;

$\sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2}{n-1}}$ – незміщене виправлене середнє квадратичне відхилення $j$-ої незалежної змінної.

При достатньо великих обсягах вибірки ($n \geq 30$) вибіркове середнє квадратичне відхилення ${\sigma_{x_j}}^{виб} = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2}{n}}$ та виправлене середнє квадратичне відхилення різняться дуже мало, тому виправлене середнє квадратичне відхилення використовують лише при обсягах вибірок $n < 30$.

*Крок 2*. Знаходження елементів кореляційної матриці:

$$r_{xx} = \frac{1}{n} X^{*T} X^* = \begin{pmatrix} 1 & r_{x_1x_2} & \cdots & r_{x_1x_k} \\ r_{x_2x_1} & 1 & \cdots & r_{x_2x_k} \\ \cdots & \cdots & \cdots & \cdots \\ r_{x_kx_1} & r_{x_kx_2} & \cdots & 1 \end{pmatrix} \tag{4.2}$$

(матриці моментів нормалізованої системи нормальних рівнянь; її елементи – парні коефіцієнти кореляції між незалежними змінними, - характеризують щільність зв’язку однієї незалежної змінної з іншою).

Тут $X* = \begin{pmatrix} {x_{11}}^* & {x_{12}}^* & & {x_{1k}}^* \\ {x_{21}}^* & {x_{22}}^* & \cdots & {x_{2k}}^* \\ {x_{31}}^* & {x_{32}}^* & & {x_{3k}}^* \\ \vdots & & \ddots & \vdots \\ {x_{n1}}^* & {x_{n2}}^* & \cdots & {x_{nk}}^* \end{pmatrix}$ - матриця нормалізованих незалежних змінних, $X^{*T}$ – матриця транспонована до матриці $X^*$, $r_{x_i x_j}$ – парні коефіцієнти кореляції.

*Крок 3.* Обчислення визначника кореляційної матриці $r_{xx}$ і значення спостережуваного критерію Пірсона $\chi^2$ («хи» квадрат) за формулою:

$$\chi^2_{\text{сп}} = -(n - 1 - (2k + 5)/6)\, ln(\det(r_{xx})) \,. \tag{4.3}$$

Висувається статистична гіпотеза $H_0$: $\det(r_{xx}) = 1$ при альтернативній гіпотезі $H_\alpha$: $\det(r_{xx}) < 1$.

Спостережуване значення критерію $\chi^2_{\text{сп}}$ порівнюється з табличним значенням $\chi^2_{\text{табл}}(\alpha, k_1)$ при $k_1 = k\,(k - 1)\,/\,2$ ступенях вільності і рівні значущості α (таблиця 3 Додатку). Критична область зображена на рисунку 4.1.

Якщо спостережуване значення статистичного критерію більше табличного, тобто $\chi^2_{\text{сп}} > \chi^2_{\text{табл}}(\alpha, k_1)$, то гіпотеза $H_0$ відхиляється, що свідчить про наявність ознаки мультиколінеарності в моделі. У протилежному випадку ця ознака відсутня.

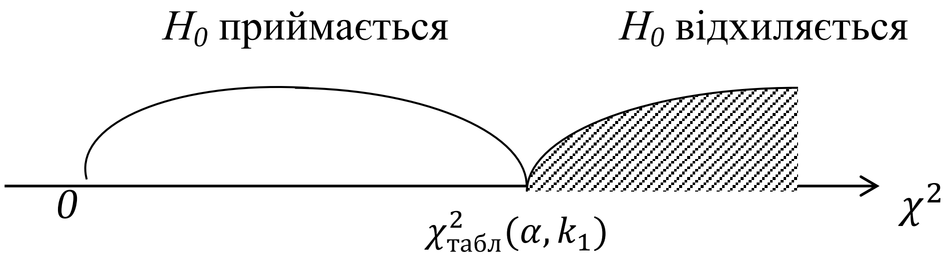



Рис.4.1

*Крок 4.* Знаходження оберненої матриці $r_{xx}^{-1} = \left(\frac{1}{n} X^{*T} X^*\right)^{-1}$. Далі будемо використовувати позначення: $D = r_{xx}^{-1}$.

*Крок 5.* Перевірка за $F$-критерієм суттєвості зв'язку кожної пояснювальної змінної з рештою змінних.

Обчислимо фактичні статистики $F$-критеріїв за формулою:

$$F_{j\ \text{факт}} = \left(d_{jj} - 1\right)\frac{n - k}{k - 1} \qquad (j = \overline{1, k}) \tag{4.4}$$

де $d_{jj}$ – діагональні елементи матриці $D$.

За визначеною статистикою $F_{j\,\text{факт}}$ для кожного фактора перевіряється на правильність статистична гіпотеза $H_0$: $d_{jj} = 1$ при альтернативній $H_\alpha$: $d_{jj} < 1$. Критична область зображена на рис. 4.1.

Фактичні значення $F_{j\,\text{факт}}$ порівнюються з табличними значеннями $F_{\text{табл}}(\alpha, k_{1,} k_2)$ при $k_1 = k - 1$, $k_2 = n - k$ ступенях вільності і рівні значущості α (таблиця 1 Додатку). Якщо

$$F_{j\text{факт}} > F_{\text{табл}}(\alpha, k_1, k_2),$$

то $j$-ий фактор буде мати суттєвий зв'язок із $k - 1$ факторами, тобто $j$-а незалежна змінна мультиколінеарна з іншими змінними. У протилежному випадку - зв'язок між цими факторами несуттєвий.

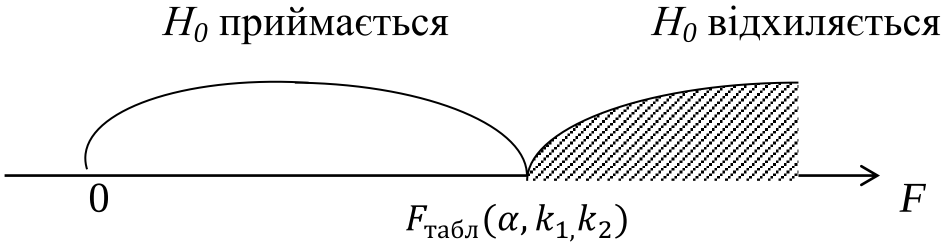



Рис. 4.2.

*Крок 6*. Обчислення часткових коефіцієнтів кореляції:

$$r_{ij} = -\frac{d_{ij}}{\sqrt{d_{ii} \cdot d_{jj}}} \qquad \left(i, j = \overline{1, k}\right), \qquad (4.5)$$

де $r_{ij}$- коефіцієнт часткової кореляції, який характеризує тісноту зв’язку між факторними ознаками $x_i$ і $x_j$ *при фіксованих значеннях інших факторів*; $d_{ij}$ − елемент матриці $D$, що міститься в $i$-*му* рядку і $j$-му стовпці; $d_{ii}$ і $d_{jj}$ – діагональні елементи цієї ж матриці.

*Часткові коефіцієнти кореляції значно менші парних* $(r_{ij} < r_{x_i x_j})$, тому на основі знання парних коефіцієнтів кореляції неможливо зробити висновок про мультиколінеарність. Часткові коефіцієнти кореляції $r_{ij}$ є статистичними прообразами теоретичних часткових коефіцієнтів, що перевіряються на статистичну значущість.

Для цього використовується $t$-статистика Ст'юдента

$$t_{ij} = |r_{ij}| \cdot \sqrt{n-m} \,/\, \sqrt{1-r_{ij}^2}\,. \qquad (4.6)$$

За її допомогою здійснюється перевірка правильності гіпотези $H_0$: $r_{ij} = 0$ при альтернативній гіпотезі $H_\alpha$: $r_{ij} \neq 0$. За заданим рівнем значущості $\alpha$ і $k_1 = n - k$ ступенях вільності за таблицею 2 Додатку знаходять $t''_{\text{табл.}}(\alpha/2,\ k_1)$. Враховуючи, що $t'_{\text{табл}}(\alpha/2, k) = -\ t''_{\text{табл.}}(\alpha/2,\ k_1)$, будують двобічну критичну область, що зображена на рисунку 4.3:

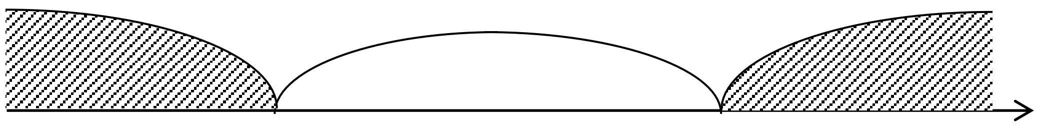



Рис. 4.3

Якщо $t_{ij} \in [t'_{\text{табл.}}(\alpha / 2, k_1);\ t''_{\text{табл.}}(\alpha / 2, k_1)]$, то $r_{ij} = 0$, інакше $r_{ij} \neq 0$. При $t_{ij} > t''_{\text{табл.}}(\alpha/2, k_1)$ між змінними $x_i$ і $x_j$ існує мультиколінеарність.

*Висновки*

1. Між незалежними змінними може існувати лінійна залежність, але вона може й не бути явищем мультиколінеарності змінних, а тому не впливатиме на кількісні оцінки розрахованих параметрів моделі.

2. При $F_{j\ \text{факт}} > F_{\text{табл}}$ фактор $x_j$ залежить від усіх інших незалежних змінних. У цьому випадку треба вирішити питання про виключення $x_j$ з переліку змінних.

3. При $t_{ij} > t''_{\text{табл.}}$ змінні $x_i$ і $x_j$ щільно пов’язані між собою.

4. Аналізуючи $F$- і $t$- критерії, робиться висновок про те, яку зі змінних треба виключити з моделі. Доцільно врахувати економіко-логіко-теоретичні міркування.

5. Якщо виконання пунктів 2 – 4 не усуває мультиколінеарність, то оцінку параметрів моделі треба обчислювати іншими методами.

## 4.5. Визначення рівня мультиколінеарності [11]

Визначення рівня мультиколінеарності за своїм змістом близьке до тестування її наявності. Якщо факторів більше, ніж два, то аналізу парних коефіцієнтів кореляції для визначення наявності мультиколінеарності вже недостатньо. У цьому випадку будуються рівняння регресії, які зв'язують кожний регресор з усіми іншими.

Розглянемо множинну лінійну кореляційну модель. Оцінимо параметри $k$ кореляційно-регресійних моделей, в яких *результуючими змінними* будуть почергово виступати факторні ознаки $x_j$ $(j = \overline{1,k})$.

$$\tilde{x}_1 = b_0^{(1)} + b_2^{(1)}x_2 + b_3^{(1)}x_3 + \cdots b_k^{(1)}x_k$$
$$\tilde{x}_2 = b_0^{(2)} + b_1^{(2)}x_1 + b_3^{(2)}x_3 + \cdots b_k^{(2)}x_k$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\tilde{x}_k = b_0^{(k)} + b_1^{(k)}x_1 + b_2^{(k)}x_2 + \cdots b_{k-1}^{(k)}x_{k-1.}$$

Коефіцієнт множинної детермінації для кожної факторної ознаки обчислюється за формулою

$$R_j^2 = 1 - 1/d_{jj}, \tag{4.7}$$

де $d_{jj}$ $(j = \overline{1,k})$ – діагональні елементи матриці $D$.

Коефіцієнт детермінації $R_j^2$ $(j = \overline{1,k})$ показує, яку частку дисперсії $j$-ої факторної ознаки пояснюють всі інші факторні ознаки, тобто характеризує вплив на факторну змінну $x_j$ усіх інших факторних ознак. Він використовується для того, щоб визначити ступінь мультиколінеарності. Для цього розраховується величина дисперсійно-інфляційного фактора *VIF* (*varience inflationary factor*) для кожної факторної ознаки:

$$VIF_j = 1/(1 - R_j^2) \qquad (j = \overline{1,k})\,. \tag{4.8}$$

За критичне значення дисперсійно-інфляційного фактора обирається $VIF_j^{\text{кр}} = 10$. Якщо $VIF_j < 10$, то це свідчить про те, що має місце недостатність зв'язку між $j$-ою факторною ознакою та всіма іншими. При $VIF_j \geq 10$ має явище мультиколінеарності між $j$-ою факторною ознакою та всіма іншими.

### 4.6. Деякі напрямки усунення мультиколінеарності

Виявлення мультиколінеарності є лише частиною справи, набагато важче її усунути, особливо при великих обсягах вибірок.

Єдиного методу усунення мультиколінеарності в економетрії не існує, усе залежить від степені мультиколінеарності. Існують такі напрямки, за допомогою яких можна якщо не усунути мультиколінеарність, то хоча б пом’якшити її дію за умови, що це не впливає на результати досліджень.

1. Відкидання однієї зі змінних мультиколінеарної пари. Це – найпростіший шлях позбутися явища мультиколінеарності. Однак на практиці вилучення якогось чинника часто суперечить логіці економічних зв’язків, що призводить до втрачання бажаних властивостей при оцінки параметрів економетричної моделі. Крім того, вилучення змінної з моделі може призвести до зміщення оцінок.

2. Перетворення незалежних змінних таким чином:

- узяти відхилення від середньої;
- замість абсолютних значень використовувати відносні (темпи зростання, приросту);
- замінити фактор $X_j^*$, який має тісний зв'язок з $X_i$, на фактор $X_j^* = X_i - X_j$; далі перевірити наявність мультиколінеарності між факторами $X_j^*$ і $X_i$;
- стандартизувати пояснювальні змінні тощо.

3. Змінити специфікацію моделі, що дає змогу уникнути явища зв’язку між пояснювальними змінними.

4. Введення додаткової пояснювальної змінної, яка суттєво корелює і впливає на досліджуваний показник. Однак нова змінна теж може корелювати з деякими пояснювальними змінними. Тому перед тим, як вводити нову змінну, треба провести додатковий аналіз.

5. Використання додаткової або первинної інформації.

6. Об’єднання інформації.

7. Збільшення кількості спостережень.

Існують також інші напрямки (використання заміщуючої змінної, введення лінійних обмежень на параметри, що зменшує кількість пояснювальних змінних та інші). Успішність застосування того чи іншого методу залежить від істотності та характеру проблеми. Якщо вказаними напрямками не вдається усунути мультиколінеарність, то для оцінювання параметрів багатовимірної моделі доцільно застосувати статистичні методи (факторний аналіз, гребенева регресія, метод головних компонентів та інші).

## 4.7. Приклад застосування алгоритму Феррара - Глобера для виявлення мультиколінеарності факторів

*Постановка задачі*. На балансовий прибуток ($Y$, тис. грош. од.) впливає низка чинників: статутний фонд ($X_1$, тис. грош. од.), кількість випущених акцій ($X_2$, тис. шт.) та обсяг продукції ($X_3$, тис. грош. од.). Статистичні дані наведені у таблиці 4.1. Визначити вплив на балансовий прибуток зазначених чинників і дослідити їх на наявність мультиколінеарності. За наявності мультиколінеарності

а) визначити її рівень;

б) використати один з наведених методів для її усунення.

***Розв'язання***

Оскільки кількість пояснювальних змінних $k = 3 > 1$, то перед тим, як знайти оцінки параметрів моделі за допомогою МНК, проведемо тестування на наявність мультиколінеарності між пояснювальними змінними. Для цього скористаємося алгоритмом Феррара - Глобера.

*1 крок.* Нормалізація статистичної інформації.

Вхідну інформацію обчислюємо за формулою (4.1). Нормалізовані дані записані в останніх трьох стовпцях таблиці 4.2.

Обчислимо дисперсії факторів:

$$\sigma_{x_1}^2 = \frac{\sum_{i=1}^{20}\left(x_{1i}-\overline{x_1}\right)^2}{n-1} \approx 85380788{,}2/19 \approx 4493725{,}69;$$

$$\sigma_{x_2}^2 = \frac{\sum_{i=1}^{20}(x_{2i}-\overline{x_1})^2}{n-1} \approx 234449304{,}2/19 \approx 12339437{,}06;$$

$$\sigma_{x_3}^2 = \frac{\sum_{i=1}^{20}(x_{3i}-\overline{x_1})^2}{n-1} \approx 14281580{,}55/19 \approx 751662{,}134.$$

Таблиця 4.1

| № | Балансовий прибуток, $Y$ | Статутний фонд, $X_1$ | Кількість випущених акцій, $X_2$ | Обсяг продукції $X_3$ |
|---|---|---|---|---|
| 1 | 5857 | 3750 | 689 | 1475 |
| 2 | 3616 | 1717 | 289 | 2788 |
| 3 | 1299 | 285 | 103 | 406 |
| 4 | 2936 | 5182 | 5070 | 1863 |
| 5 | 1767 | 476 | 427 | 1382 |
| 6 | 833 | 4869 | 3815 | 517 |
| 7 | 231 | 202 | 131 | 215 |
| 8 | 495 | 1224 | 1591 | 345 |
| 9 | 258 | 423 | 197 | 276 |
| 10 | 79 | 140 | 264 | 55 |
| 11 | 259 | 2575 | 6305 | 146 |
| 12 | 141 | 1612 | 1974 | 861 |
| 13 | 2487 | 788 | 245 | 2340 |
| 14 | 104 | 1969 | 1608 | 114 |
| 15 | 274 | 8161 | 14907 | 81 |
| 16 | 21 | 379 | 190 | 91 |
| 17 | 4562 | 2870 | 525 | 1025 |
| 18 | 2458 | 715 | 217 | 2345 |
| 19 | 687 | 79 | 32 | 387 |
| 20 | 1520 | 2450 | 2535 | 981 |

*2 крок.* Знаходження кореляційної матриці та обчислення коефіцієнтів парної кореляції. На основі нормалізованих даних за формулою (4.2) знайдемо кореляційну матрицю $r_{xx}$. Для цього засобами Excel транспонуємо матрицю $X^*$ (функція ТРАНСП), а потім знайдемо добуток матриць $(X^*)^T$ та $X^*$ (функція МУМНОЖ).

$$(X^*)^T X^* = \begin{pmatrix} 0{,}829 & -0{,}130 & -0{,}806 & \dots & -0{,}903 & 0{,}215 \\ -0{,}389 & -0{,}503 & -0{,}556 & \dots & -0{,}576 & 0{,}136 \\ 0{,}681 & 2{,}195 & -0{,}552 & \dots & -0{,}574 & 0{,}111 \end{pmatrix} \cdot \begin{pmatrix} 0{,}829 & -0{,}389 & 0{,}681 \\ -0{,}130 & -0{,}503 & 2{,}195 \\ -0{,}806 & -0{,}556 & -0{,}552 \\ \dots & \dots & \dots \\ -0{,}603 & -0{,}523 & 1{,}684 \\ -0{,}903 & -0{,}576 & -0{,}574 \\ 0{,}215 & 0{,}136 & 0{,}111 \end{pmatrix}$$

$$(X^*)^T X^* = \begin{pmatrix} 19 & 16{,}2066 & 0{,}0904 \\ 16{,}2066 & 19 & -4{,}5004 \\ 0{,}0904 & -4{,}5004 & 19 \end{pmatrix}$$

$$r_{xx} = (1/19) \begin{pmatrix} 19 & 16{,}2066 & 0{,}0904 \\ 16{,}2066 & 19 & -4{,}5004 \\ 0{,}0904 & -4{,}5004 & 19 \end{pmatrix} = \begin{pmatrix} 1 & 0{,}8530 & 0{,}0048 \\ 0{,}8530 & 1 & -0{,}2369 \\ 0{,}0048 & -0{,}2369 & 1 \end{pmatrix}$$

Перевіримо наявність мультиколінеарності серед пояснювальних змінних $X_1, X_2, X_3$.

*3 крок*. Обчислення визначника кореляційної матриці $r_{xx}$ і значення спостережуваного критерію Пірсона $\chi^2$ за формулою (4.3). За допомогою функції МОПРЕД знаходимо визначник матриці $r_{xx}$:

$$det\,(r_{xx}) = 0{,}214.$$

Оскільки $det\ (r_{xx}) \neq 0$, то між факторами відсутній лінійний зв'язок. Переконаємось також у відсутності тісного кореляційного зв’язку між пояснювальними змінними $X_1, X_2, X_3$.

Якщо $det\ (r_{xx})$ наближається до одиниці, то мультиколінеарність відсутня, а фактори, що включені у модель, незалежні.

Статистична оцінка відсутності зв’язку між факторами може бути перевірена за допомогою гіпотези про незалежність змінних, на основі яких розраховано кореляційну матрицю, тобто $H_0$: $det\ (r_{xx}) = 1$.

Спостережуване значення критерію Пірсона дорівнює

$$\chi^2_{\text{сп}} = -\,(n - 1 - (2k + 5) / 6)\ ln\ (det\ (r_{xx}))) =$$
$$= -\,(20 - 1 - (2 \cdot 3 + 5) / 6)\ ln\ 0{,}214 \approx 26{,}437.$$

Знайдемо за таблицею 3 Додатку табличне значення $\chi^2_{\text{табл}}$ при рівні значущості $\alpha$ = 0,05 і $k_1 = k\,(k-1)/2 = 3\cdot(3-1)/2 = 3$ ступенями вільності:

$$\chi^2_{\text{табл}}(\alpha, k_1) = \chi^2_{\text{табл}}(0{,}05; 3) = 7{,}81.$$

Оскільки $\chi^2_{\text{сп}} > \chi^2_{\text{табл}}$, то гіпотеза $H_0$ відхиляється, визначник $det(r_{xx})$ суттєво відрізняється від одиниці, що, в свою чергу, свідчить про наявність мультиколінеарності пояснювальних змінних.

Таблиця 4.2

| $i$ | $X_1$ | $X_2$ | $X_3$ | $(X_{1i}-X_{1сер})^2$ | $(X_{2i}-X_{2сер})^2$ | $(X_{3i}-X_{3сер})^2$ | $X_{1i}^*$ | $X_{2i}^*$ | $X_{3i}^*$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3750 | 689 | 1475 | 3085994,89 | 1867868,89 | 348513,12 | 0,829 | -0,389 | 0,681 |
| 2 | 1717 | 289 | 2788 | 76341,69 | 3121228,89 | 3622741,22 | -0,130 | -0,503 | 2,195 |
| 3 | 285 | 103 | 406 | 2918288,89 | 3813037,29 | 229105,82 | -0,806 | -0,556 | -0,552 |
| 4 | 5182 | 5070 | 1863 | 10167807,69 | 9086004,49 | 957168,72 | 1,504 | 0,858 | 1,128 |
| 5 | 476 | 427 | 1382 | 2302199,29 | 2652663,69 | 247357,02 | -0,716 | -0,464 | 0,574 |
| 6 | 4869 | 3815 | 517 | 8269650,49 | 3095136,49 | 135166,52 | 1,357 | 0,501 | -0,424 |
| 7 | 202 | 131 | 215 | 3208755,69 | 3704470,09 | 448431,12 | -0,845 | -0,548 | -0,772 |
| 8 | 1224 | 1591 | 345 | 591822,49 | 215946,09 | 291222,12 | -0,363 | -0,132 | -0,622 |
| 9 | 423 | 197 | 276 | 2465842,09 | 3454765,69 | 370454,82 | -0,741 | -0,529 | -0,702 |
| 10 | 140 | 264 | 55 | 3434720,89 | 3210188,89 | 688319,12 | -0,874 | -0,510 | -0,957 |
| 11 | 2575 | 6305 | 146 | 338374,89 | 18056550,49 | 545603,82 | 0,274 | 1,210 | -0,852 |
| 12 | 1612 | 1974 | 861 | 145389,69 | 6674,89 | 559,32 | -0,180 | -0,023 | -0,027 |
| 13 | 788 | 245 | 2340 | 1452748,09 | 3278634,49 | 2118043,62 | -0,569 | -0,515 | 1,679 |
| 14 | 1969 | 1608 | 114 | 590,49 | 200435,29 | 593901,42 | -0,011 | -0,127 | -0,889 |
| 15 | 8161 | 14907 | 81 | 38040523,29 | 165155911,7 | 645853,32 | 2,910 | 3,658 | -0,927 |
| 16 | 379 | 190 | 91 | 2605964,49 | 3480836,49 | 629880,32 | -0,762 | -0,531 | -0,915 |
| 17 | 2870 | 525 | 1025 | 768602,89 | 2343042,49 | 19698,12 | 0,414 | -0,436 | 0,162 |
| 18 | 715 | 217 | 2345 | 1634050,89 | 3380817,69 | 2132622,12 | -0,603 | -0,523 | 1,684 |
| 19 | 79 | 32 | 387 | 3664544,49 | 4095361,69 | 247655,52 | -0,903 | -0,576 | -0,574 |
| 20 | 2450 | 2535 | 981 | 208574,89 | 229728,49 | 9283,32 | 0,215 | 0,136 | 0,111 |
| **Σ** | **39866** | **41114** | **17693** | **85380788,2** | **234449304,2** | **14281580,55** | | | |
| **Сер. знач** | **1993,3** | **2055,7** | **884,65** | **4269039,41** | **11722465,21** | **714079,0275** | | | |
| **$\sigma^2$** | **4493725,7** | **12339437,1** | **751662** | | | | | | |
| **$\sigma$** | **2119,841** | **3512,75349** | **866,985** | | | | | | |

*4 крок.* Знаходження оберненої матриці $D = (r_{xx})^{-1}$.

За допомогою функції МОБР знайдемо матрицю $D$:

$$D = \begin{pmatrix} 4{,}4029 & -3{,}9841 & -0{,}9646 \\ -3{,}9841 & 4{,}6646 & 1{,}1238 \\ -0{,}9646 & 1{,}1238 & 1{,}2708 \end{pmatrix}$$

*5 крок.* Перевіримо за $F$-критерієм суттєвість зв'язку кожної пояснювальної змінної з рештою змінних. Обчислимо значення $F$-критеріїв за формулою (4.4):

$$F_1 = (d_{11} - 1)\frac{20-3}{3-1} = (4{,}4029 - 1)\cdot\frac{17}{2} \approx 28{,}925;$$

$$F_2 = (d_{22} - 1)\frac{20-3}{3-1} = (4{,}6646 - 1)\cdot\frac{17}{2} \approx 31{,}1487;$$

$$F_3 = (d_{33} - 1)\frac{20-3}{3-1} = (1{,}2708 - 1)\cdot\frac{17}{2} \approx 2{,}3016.$$

Знайдемо за таблицею 1 Додатку табличне значення $F$-критерію при рівні значущості $\alpha = 0{,}05$ та ступенями вільності $k_1 = k - 1 = 2$ і $k_2 = n - k = 17$ ($n = 20$ – обсяг вибірки; $k = 3$ – число факторів):

$$F_{\text{табл}}(\alpha, k_1, k_2) = F_{\text{табл}}(0{,}05;\ 2,\ 17) = 3{,}59.$$

*Висновок.*

Оскільки $F_1 > F_{\text{табл}}$, $F_2 > F_{\text{табл}}$, то має місце суттєва залежність першої пояснювальної змінної від другої та третьої (статутного фонду від кількості випущених акцій та обсягу продукції), а також другої пояснювальної змінної від першої та третьої (кількості випущених акцій від статутного фонду та обсягу продукції). З нерівності $F_3 < F_{\text{табл}}$ випливає, що третя змінна (обсяг продукції) не корелює із змінними, що визначають статутний фонд та кількість випущених акцій.

*6 крок.* Обчислення часткових коефіцієнтів кореляції за формулою (4.5).

$$r_{12} = -\frac{d_{12}}{\sqrt{d_{11}\cdot d_{22}}} \approx -\frac{-3{,}9841}{\sqrt{4{,}4029\cdot 4{,}6646}} \approx 0{,}879;$$

$$r_{13} = -\frac{d_{13}}{\sqrt{d_{11} \cdot d_{33}}} \approx -\frac{-0{,}9646}{\sqrt{4{,}4029 \cdot 1{,}2708}} \approx 0{,}4078;$$

$$r_{23} = -\frac{d_{23}}{\sqrt{d_{22} \cdot d_{33}}} \approx -\frac{1{,}1238}{\sqrt{4{,}6646 \cdot 1{,}2708}} \approx -0{,}4616.$$

Перевіримо статистичну значущість часткових коефіцієнтів кореляції за $t$-критерієм Ст'юдента. Для цього висуваємо гіпотезу $H_0$: $r_{ij} = 0$ при альтернативній гіпотезі $H_\alpha$: $r_{ij} \neq 0$. За формулою (4.6)

$$t_{12} = 0{,}879 \cdot \frac{\sqrt{20-3}}{\sqrt{1-0{,}879^2}} \approx 7{,}61;$$

$$t_{13} = 0{,}408 \cdot \frac{\sqrt{20-3}}{\sqrt{1-0{,}408^2}} \approx 1{,}84;$$

$$t_{23} = |-0{,}462| \cdot \frac{\sqrt{20-3}}{\sqrt{1-0{,}462^2}} \approx 2{,}15.$$

При рівні значущості $\alpha = 0{,}05$ і $k_1 = n - k = 17$ ступенях вільності за таблицею 2 Додатку знаходимо табличне значення статистики

$$t''_{\text{табл.}}(\alpha / 2, k_1) = t''_{\text{табл.}}(0{,}025, 17) = 2{,}11.$$

Оскільки $t_{12} > t_{\text{табл}}$, $t_{23} > t_{\text{табл}}$, $t_{13} < t_{\text{табл}}$ , то з імовірністю 0,95 нульова гіпотеза про рівність нулю частинних коефіцієнтів кореляції $r_{12}$ і $r_{23}$ відхиляється, а для частинного коефіцієнта $r_{23}$ приймається. Таким чином, між змінними $X_1$ і $X_2$ (статутним фондом і кількістю випущених акцій), $X_2$ і $X_3$ (кількістю випущених акцій і обсягом продукції) існує тісний кореляційний зв'язок, що свідчить про наявність мультиколінеарності

*Визначення рівня мультиколінеарності.*

Обчислимо коефіцієнти детермінації для кожної множинної моделі, яка описує залежність однієї факторної ознаки від усіх інших.

За формулою (4.7)

$$R_1^2 = 1 - 1/d_{11} \approx 1 - 1/4{,}403 \approx 0{,}77;$$

$$R_2^2 = 1 - 1/d_{22} \approx 1 - 1/4{,}66 \approx 0{,}79;$$

$$R_3^2 = 1 - 1/d_{33} \approx 1 - 1/1{,}27 \approx 0{,}21.$$

*Висновки*. Зміна кількості випущених акцій ($X_2$) та обсяг продукції ($X_3$) пояснює 77% дисперсії величини, що характеризує статутний фонд ($X_1$). Зміна статутного фонду ($X_1$) та обсягу продукції ($X_3$) пояснює 79% дисперсії кількості випущених акцій ($X_2$). Зміна статутного фонду ($X_1$) та кількості випущених акцій ($X_2$) пояснює 21% дисперсії обсягу продукції ($X_3$).

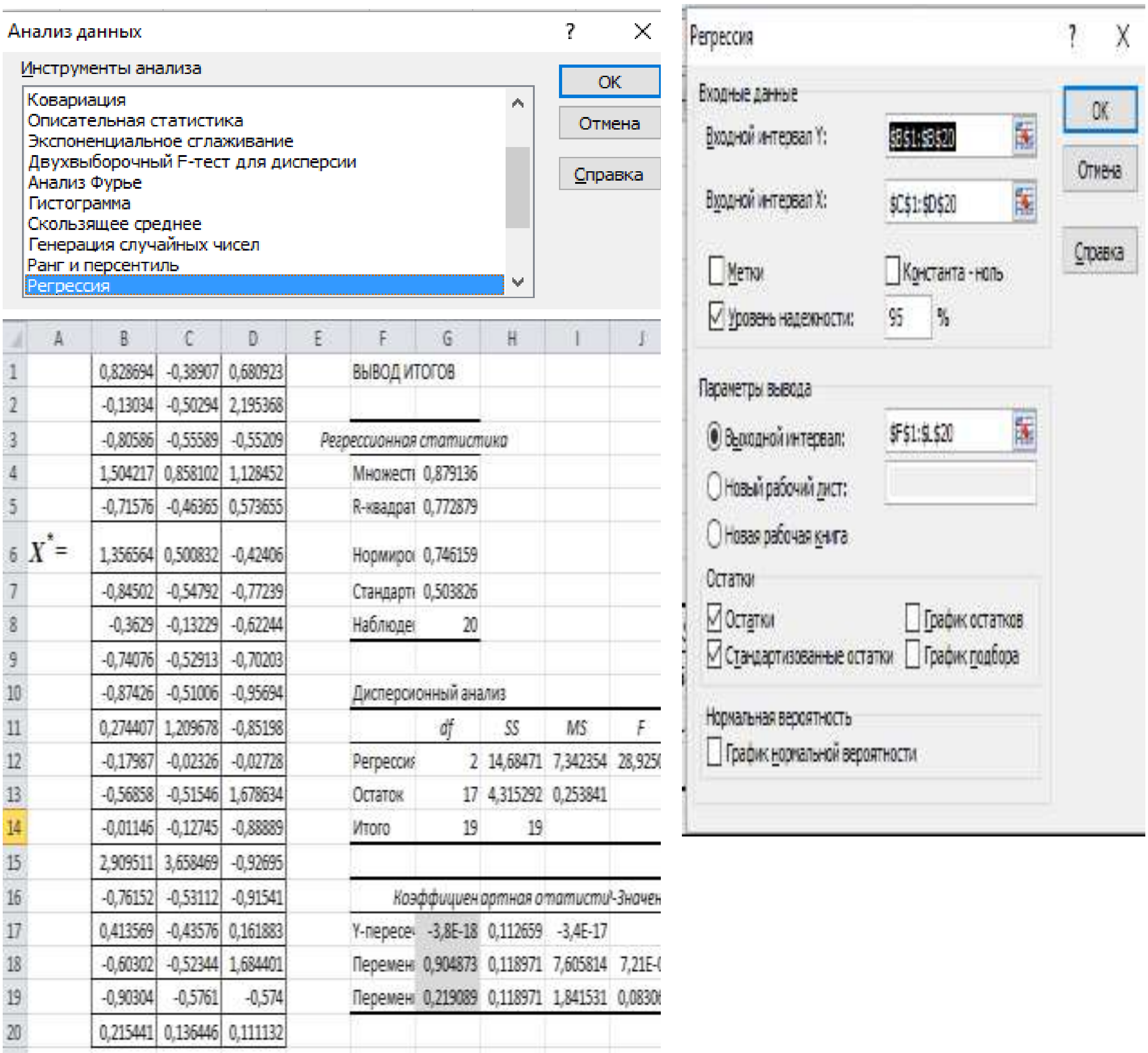



Рис. 4.4

*Обчислення значень дисперсійно-інфляційного фактору $VIF_j$ для факторних ознак $X_j$* ($j = 1, 2, 3$).

Побудуємо лінійну двофакторну кореляційно-регресійну модель, результуючою змінною якої є факторна ознака $X_1$ (величина статутного фонду). Невідомі параметри моделі визначаємо за допомогою МНК. Для цього використаємо засоби Excel:

*Данные → Анализ данных → Регрессия* (Вывод итогов) (рис. 4.4). Тут у полі *Входной интервал Y* розміщується діапазон значень $X_1^*$, а у полі *Входной интервал X* – діапазон значень $X_2^*, X_3^*$ .

Вибіркова кореляційно-регресійна модель має вигляд:

$$\tilde{x}_1 = 0{,}9049 \cdot x_2 + 0{,}2191 \cdot x_3.$$

Коефіцієнт множинної детермінації для цієї моделі $R_1^2 \approx 0{,}77$.

Побудуємо лінійну двофакторну кореляційно-регресійну модель, результуючою змінною якої є факторна ознака $X_2$ (кількість випущених акцій). Невідомі параметри моделі визначаємо за допомогою МНК. Для цього використаємо засоби Excel:

*Данные → Анализ данных → Регрессия* (Вывод итогов). У полі *Входной интервал Y* розміщується діапазон значень $X_2^*$, а у полі *Входной интервал X* – діапазон значень $X_1^*, X_3^*$ (розмістивши їх поруч на сусідніх стовпцях).

Вибіркова кореляційно-регресійна модель має вигляд:

$$\tilde{x}_2 = 0{,}8541 \cdot x_1 - 0{,}2409 \cdot x_3.$$

Коефіцієнт множинної детермінації для цієї моделі $R_2^2 \approx 0{,}79$.

Аналогічно побудуємо лінійну двофакторну модель, результуючою змінною якої є факторна ознака $X_3$ (обсяг продукції):

$$\tilde{x}_3 = 0{,}7591 \cdot x_1 - 0{,}8844 \cdot x_2.$$

Коефіцієнт множинної детермінації для цієї моделі $R_3^2 \approx 0{,}21$.

За формулою (4.8) обчислимо значення дисперсійно-інфляційного фактору $VIF_j$ для кожної факторної ознаки $X_j$:

$$VIF_1 \approx 1 / (1 - 0{,}77) \approx 4{,}40 < 10;$$

$VIF_2 \approx 1 / (1 - 0{,}79) \approx 4{,}67 < 10;$ $VIF_3 \approx 1 / (1 - 0{,}21) \approx 1{,}27 < 10.$

Для всіх факторних ознак $VIF_j < 10$, що вказує на недостатність зв'язку між $j$-ю ознакою та іншими ознаками, що залишилися.

*Визначення змінних, між якими відсутній зв'язок і які можна включити до моделі як незалежні змінні*

Нормалізуємо вектор $Y$ за формулою: $y_i^* = (y_i - \overline{y})/\sigma_y$,

де виправлена вибіркова дисперсія $\sigma_y^2 = \frac{\sum_{i=1}^{20}(y_i - \bar{y})^2}{n-1}$.

Вектор коефіцієнтів парної кореляції має вигляд:

$$r_{yx} = \frac{1}{n-1} X^{*T} y^*.$$

Виконуємо розрахунки за допомогою функцій ТРАНСП та МУМНОЖ.

| $i$ | $y_i$ | $y_i^*$ |
|---|---|---|
| 1 | 5857 | 2,6069 |
| 2 | 3616 | 1,2678 |
| 3 | 1299 | -0,1166 |
| 4 | 2936 | 0,8615 |
| 5 | 1767 | 0,1630 |
| 6 | 833 | -0,3951 |
| 7 | 231 | -0,7548 |
| 8 | 495 | -0,5970 |
| 9 | 258 | -0,7387 |
| 10 | 79 | -0,8456 |
| 11 | 259 | -0,7381 |
| 12 | 141 | -0,8086 |
| 13 | 2487 | 0,5932 |
| 14 | 104 | -0,8307 |
| 15 | 274 | -0,7291 |
| 16 | 21 | -0,8803 |
| 17 | 4562 | 1,8331 |
| 18 | 2458 | 0,5759 |
| 19 | 687 | -0,4823 |
| 20 | 1520 | 0,0154 |
| **Σ** | **29884** | |
| **середнє** | **1494,2** | |
| **$б^2$** | 2800870,5 | |
| **б** | 1673,5801 | |

$$(X^*)^T y^* = \begin{pmatrix} 3{,}887 \\ -3{,}901 \\ 13{,}550 \end{pmatrix}, \qquad r_{y^*x} = \begin{pmatrix} 0{,}2046 \\ -0{,}2053 \\ 0{,}7132 \end{pmatrix}$$

Аналіз елементів вектора $r_{y^*x}$ вказує, що суттєвіше впливає на $y^*$ пояснювальна змінна $X_3$, оскільки має найбільший парний коефіцієнт кореляції (0,7132). Змінні $X_1$ і $X_2$ майже однаково впливають на $y^*$ та корелюють між собою. Доцільно не включати змінну $X_2$ до моделі як пояснювальний фактор.

Розглянемо таку двофакторну економетричну модель:

$$y_i^* = b_0 + b_1 x_{1i} + b_3 x_{3i} \ (i = \overline{1,n}). \qquad (4.9)$$

Переконаємося у тому, що пояснювальні змінні цієї моделі не колінеарні між собою. Кореляційна матриця має вигляд (*крок* 2, елементи $r_{x1x3} = r_{x3x1} = 0{,}0048$ матриці $r_{xx}$):

$$r_{xx} = \begin{pmatrix} 1 & 0{,}0048 \\ 0{,}0048 & 1 \end{pmatrix}.$$

Обчислимо визначник цієї матриці:

$$\det(r_{xx}) = 0{,}999977.$$

Спостережуване значення критерію Пірсона дорівнює

$$\chi^2_{\text{сп}} = -(n - 1 - (2k + 5) / 6) \ln(\det(r_{xx})) =$$

$$= -(20 - 1 - (2 \cdot 2 + 5) / 6) \ln 0{,}999977 \approx 0{,}00039.$$

Знайдемо за таблицею 3 Додатку табличне значення $\chi^2_{\text{табл}}$ при рівні значущості $\alpha = 0{,}05$ і $k_1 = k(k-1)/2 = 2 \cdot (2-1)/2 = 1$ ступені вільності:

$$\chi^2_{\text{табл}}(\alpha, k_1) = \chi^2_{\text{табл}}(0{,}05; 1) = 3{,}84.$$

Оскільки $\chi^2_{\text{сп}} < \chi^2_{\text{табл}}$, то це свідчить про несуттєву відмінність визначника кореляційної матриці від одиниці. Можна стверджувати, що пояснювальні змінні $X_1$ і $X_3$ незалежні.

Порівняємо оцінки параметрів моделей, обчислених без та з урахуванням мультиколінеарності. За МНК знайдемо оцінки параметрів вибіркової моделі

$$\hat{y} = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i}$$

без урахування зв'язків між пояснювальними змінними. Для цього використаємо функцію ЛИНЕЙН категорії *Статистические Мастер функции $f_x$*. В результаті одержимо таку таблицю значень:

| | | | | |
|---|---|---|---|---|
| коэф.регр | **0,923157** | **-0,46271** | **0,813739** | **6,692387** |
| s коэф | 0,27129 | 0,128282 | 0,206528 | 353,462 |
| R2 | **0,751317** | 909,4663 | #Н/Д | #Н/Д |
| F | **16,11296** | 16 | #Н/Д | #Н/Д |
| Sspe,Ssoct | 39982476 | 13234063 | #Н/Д | #Н/Д |

Застосовуємо верхній рядок для запису рівняння зв'язку:

$$\hat{y} = 6{,}692 + 0{,}8137 x_{1i} - 0{,}4627 x_{2i} + 0{,}923 x_{3i}.$$

Для цієї моделі коефіцієнт множинної детермінації $R^2$ = 0,75; фактичне значення критерію Фішера $F_{факт}$ = 16,11.

Знайдемо за таблицею 1 Додатку табличне значення $F$-критерію при рівні значущості $\alpha$ = 0,05 та ступенями вільності $k_1 = m - 1 = 3$ і $k_2 = n - m = 16$ ($n$ = 20 – обсяг вибірки; $m$ = 4 – число параметрів):

$$F_{табл}(\alpha, k_1, k_2) = F_{табл}(0{,}05;\ 3,\ 16) = 3{,}24.$$

Оскільки $F_{факт} > F_{табл}$, то це підтверджує суттєвість зв'язку між залежною та пояснювальними змінними. Модель є адекватною.

Для знайдених статистичних оцінок параметрів маємо (другий рядок наведеної вище таблиці) такі виправлені середньоквадратичні похибки:

$$S_{b_0} = 353{,}462;\ S_{b_1} = 0{,}206528;\ S_{b_2} = 0{,}128282; S_{b_3} = 0{,}27129.$$

Перевіримо значущість оцінок параметрів за критерієм Ст'юдента.

Обчислимо спостережувані значення статистичних критеріїв за формулою:

$$t_{b_i} = \frac{b_i - \beta_i}{S_{b_i}} \qquad (i = \overline{0, k}).$$

При заданому рівні значущості α здійснюється перевірка правдивості статистичної гіпотези $H_0$: $\beta_i = 0$ при альтернативній гіпотезі $H_\alpha$: $\beta_i \neq 0$. Спостережувані значення статистичних критеріїв:

$$t_{b_0} = \frac{b_0-\beta_0}{S_{b_0}} = \frac{6{,}692-0}{353{,}462} \approx 0{,}02; \quad t_{b_1} = \frac{b_1-\beta_1}{S_{b_1}} = \frac{0{,}8137-0}{353{,}462} \approx 3{,}94;$$

$$t_{b_2} = \frac{b_2-\beta_2}{S_{b_2}} = \frac{-0{,}4627-0}{0{,}128282} \approx -3{,}61; \quad t_{b_3} = \frac{b_3-\beta_3}{S_{b_3}} = \frac{0{,}923-0}{0{,}27129} \approx 3{,}40.$$

При рівні значущості $\alpha = 0{,}05$ і $k_1 = n - m = 16$ ступенях вільності за таблицею 2 Додатку знаходимо табличне значення статистики

$$t''_{\text{табл.}}(\alpha / 2, k_1) = t''_{\text{табл.}}(0{,}025, 16) = 2{,}12,$$

$$t''_{\text{табл.}}(0{,}025, 16) = - t'_{\text{табл.}}(0{,}025,16) = 2{,}12.$$

Оскільки $t_{b_0} \in[-2{,}12;\ 2{,}12]$; $t_{b_1} \notin[-2{,}12;\ 2{,}12]$; $t_{b_2} \notin[-2{,}12;\ 2{,}12]$; $t_{b_3} \notin[-2{,}12;\ 2{,}12]$, то параметри моделі $b_1$, $b_2$ , $b_3$ є статистично значущими, а параметр $b_0$ несуттєво відрізняється від нуля.

Розглянемо економетричну модель (4.9), з якої виключено змінну $X_2$, що корелює із змінною $X_1$ та $X_3$.

За МНК знайдемо оцінки параметрів моделі

$$\hat{y} = b_0^* + b_1^* \cdot x_{1i} + b_3^* \cdot x_{3i}.$$

Для цього на вкладці *Данные→Анализ* данных вибираємо функцію *Регрессия* (Вывод итогов). Знаходимо статистичні оцінки параметрів (рис. 4.5, стовпець *Коэффициенты*):

$$\hat{y} = -38{,}670 + 0{,}159x_{1i} + 1{,}375x_{3i}.$$

Для цієї моделі коефіцієнт множинної детермінації $R^2 = 0{,}5491$; фактичне значення критерію Фішера $F_{\text{факт}} = 10{,}35$.

Знайдемо за таблицею 1 Додатку табличне значення $F$-критерію при рівні значущості $\alpha = 0{,}05$ та ступенями вільності $k_1 = m - 1 = 2$ і $k_2 = n - m = 17$ ($n = 20$ – обсяг вибірки; $m = 3$ – число параметрів):

$$F_{\text{табл}}(\alpha, k_1, k_2) = F_{\text{табл}}(0{,}05;\ 2,\ 17) = 3{,}59.$$

Оскільки $F_{\text{факт}} > F_{\text{табл}}$, то це підтверджує суттєвість зв’язку між залежною та пояснювальними змінними. Модель є адекватною.

ВЫВОД ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный | 0,74101614 |
| R-квадрат | 0,54910492 |
| Нормированны | 0,49605844 |
| Стандартная | 1188,05515 |
| Наблюдения | 20 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость* |
|---|---|---|---|---|---|
| Регрессия | 2 | 29221463,66 | 14610731,83 | 10,35139 | 0,0011472 |
| Остаток | 17 | 23995075,54 | 1411475,032 | | |
| Итого | 19 | 53216539,2 | | | |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* |
|---|---|---|---|---|---|
| Y-пересечение | -38,670 | 461,442 | -0,084 | 0,934 | -1012,228 |
| Переменная X 1 | 0,159 | 0,129 | 1,235 | 0,233 | -0,112 |
| Переменная X 3 | 1,375 | 0,314 | 4,373 | 0,000 | 0,712 |

Рис. 4.5

Для статистичних оцінок параметрів маємо середньоквадратичні похибки: $S_{b_0^*} = 461{,}4425$; $S_{b_1^*} = 0{,}1286$; $S_{b_3^*} = 0{,}3144$.

Спостережувані значення статистичних критеріїв Ст'юдента можна обчислити безпосередньо, або скориставшись даними, що зображені у стовпці *Стандартная ошибка*:

$$t_{b_0^*} \approx 0{,}084;\ \ t_{b_1^*} \approx 1{,}235;\ t_{b_3^*} \approx 4{,}373.$$

При рівні значущості $\alpha = 0{,}05$ і $k_1 = n - m = 17$ ступенях вільності за таблицею 2 Додатку знаходимо табличне значення статистики

$$t''_{\text{табл.}}(\alpha / 2, k_1) = t''_{\text{табл.}}(0{,}025, 17) = - t'_{\text{табл.}}(0{,}025,17) = 2{,}11.$$

Оскільки $t_{b_0^*} \in [-2{,}11; 2{,}11]$; $t_{b_1^*} \in [-2{,}11; 2{,}11]$; $t_{b_3^*} \notin [-2{,}12; 2{,}12]$, то гіпотеза $H_0$: $b_3^* = 0$ відхиляється, а гіпотези $H_0$: $b_0^* = 0$ і $H_0$: $b_1^* = 0$ приймаються.

Отже, виключення змінної $X_2$ привело до звільнення від мультиколінеарності, але не покращило якість знайдених оцінок параметрів двофакторної регресії.

# РОЗДІЛ 5. УЗАГАЛЬНЕНА ЛІНІЙНА МОДЕЛЬ. ГЕТЕРОСКЕДАСТИЧНІСТЬ

Під час моделювання реальних економічних процесів мають місце ситуації, коли умови класичної лінійної моделі регресії виявляються порушеними. У багатьох економетричних дослідженнях може очікуватись, що умова проо сталу дисперсію випадкової змінної не зберігатиметься. Це можна зрозуміти, якщо врахувати фактори, вплив яких абсорбується значенням помилки. Згадаємо, що випадкова величина $\varepsilon$ виражає вплив на залежну змінну помилок в її вимірюванні та неврахованиї факторів. У обох випадках є підстави для зміни дисперсії $\varepsilon$. З іншого боку, техніка вибірки та інші методи збору даних постійно вдосконалюються і тому похибки вимірювання можуть зменшуватися. У такому разі $\sigma_{\varepsilon_i}^2$ також зменшуватиметься. Але важливіше те, що багато із неврахованих змінних можуть змінюватись в однакому з $x$ напрямку, викликаючи, таким чином, збільшення відхилення спостережень від лінії регресії.

Розглянемо, наприклад, парну лінійну регресійну модель [19]

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\ ,$$

де $y_i$ – заощадження $i$-го домогосподарства, $x_i$ – доход $i$-го домогосподарства.

З рис. 5.1 а) і б) видно, що із зростанням доходу заощадження також зростають. Але на рис. 5.1 а) дисперсія заощаджень залишається однаковой за усіх рівнів доходу, тоді як на рис. 5.1 б) вона зростає разом з доходом. Очевидно, типовішою є ситуація, що зображена на рис. 5.1 б). Сім'ї з більшим доходом, як правило, показують більшу варіацію у своїй поведінці заощаджень, ніж сім'ї з низьким доходом. Сім'ї з високим доходом схильні дотримуватись певного стандарту життя, і коли їхній доход падає, вони швидше скоротять свої заощадження, ніж споживання. З іншого боку, сім'ї з низьким доходом заощаджують із певною метою (наприклад, щоб сплатити розстрочку чи борг), тому їхні заощадження більш регулярні. Це означає, що при високих доходах вони будуть високими, тоді як при низьких доходах – малими. Тому умова про сталу дисперсію не витримується при оцінюванні функції заощаджень сімейного бюджету.

Компанії з більшими прибутками проводять ризикованішу дивідендну політику порівняно з компаніями, які мають менші прибутки. Банки, які мають у розпорядженні обладнання для обробки у місячних та квартальних звітах своїх клієнтів, помиляються менше, ніж банки, які не мають подібних засобів обслуговування.

Розглянемо випадок, коли не виконується умова 3° основних припущень кореляційно-регресійного аналізу про те, що випадкові збурення моделі мають сталу дисперсію та не корельовані між собою. Для лінійної множинної моделі ця умова означає, що коваріаційна матриця збурень має вигляд:

$$cov\,(\varepsilon\cdot\varepsilon^T) = M(\varepsilon\cdot\varepsilon^T) = \sigma_\varepsilon^2 E, \tag{5.1}$$

де $E$ – одинична матриця $n$-го порядку.

У тих випадках, коли існуючі статистичні дані достатньо однорідні, припущення (5.1) цілком виправданно. Однак у інших ситуаціях воно може виявитись неприйнятним.

Якщо дисперсія залишків в економетричних моделях залишається сталою для кожного спостереження, тобто має місце умова (5.1), то таке явище називається *гомоскедастичністю* (*homo* – однаковий, *skedasis* – дисперсія). Воно має місце, коли спостереження відносяться до однорідних об’єктів (підприємства, фірми тощо).

Однак, є явища, у яких значення змінних значно відрізняються в різних спостереженнях, тобто при $i, j = \overline{1, n}$

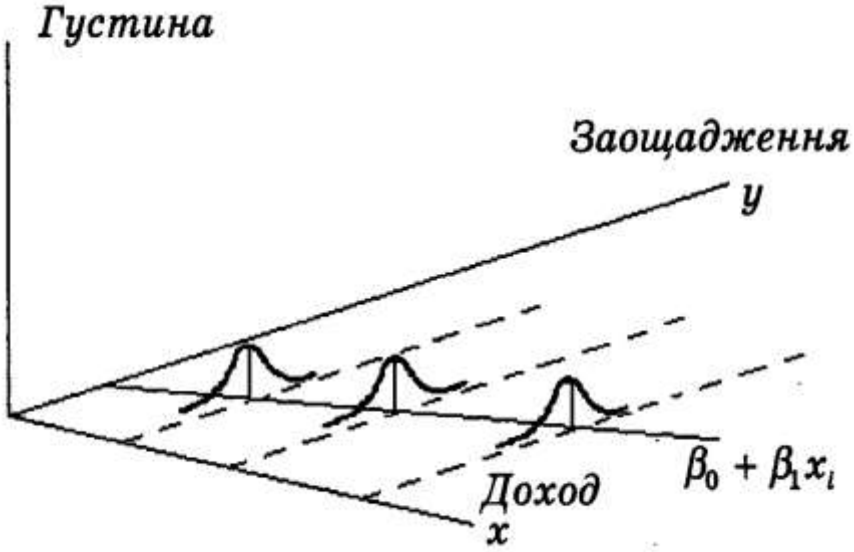
$$cov(\varepsilon_i \cdot \varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma_\varepsilon^2 \neq const, & i = j \end{cases}$$

*Гетероскедастичністю* (*hetero* – різний, *skedasis* – дисперсія) називають явище, при якому дисперсія залишків змінюється для кожного спостереження або групи спостережень, тобто

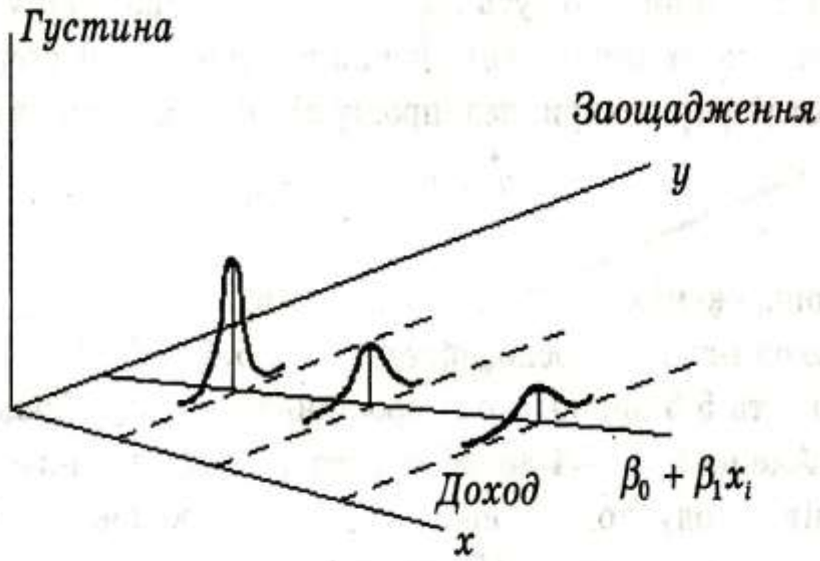
$$cov\,(\varepsilon\cdot\varepsilon^T) = M(\varepsilon\cdot\varepsilon^T) = \sigma_\varepsilon^2 S,$$

де $S$ – симетрична, додатно-визначена матриця $n$-го порядку.

Терміни гомоскедастичність та гетероскедастичність запропоновані відомим вченим-статистиком О.О.Чупровим.




а) Випадок гомоскедастичності




б) Випадок гетероскедастичності

Рис. 5.1 [19]

## 5.1. Класифікація моделей з порушенням передумов використання звичайного МНК [9]

Моделі, для яких не виконуються умови 1° - 7° класичного кореляційно-регресійного аналізу можна розділити на три групи.

*Перша група*. Моделі, для яких виконуються такі умови стосовно компонент випадкового вектора збурень $\varepsilon$:

1) вони мають нульові математичні сподівання:

$$M(\varepsilon_i) = 0, \quad i = \overline{1, n};$$

2) між собою є попарно некорельовані:

$$cov(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma_{\varepsilon_i}^2 \neq const, & i = j, \ i = \overline{1, n}. \end{cases}$$

Коваріаційна матриця випадкового вектора $\boldsymbol{\varepsilon}$ у цьому випадку має вигляд:

$$cov(\varepsilon \cdot \varepsilon^T) = M(\varepsilon \cdot \varepsilon^T) =$$

$$M\left(\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix} \cdot (\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_n)\right) = M\begin{pmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \dots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2\varepsilon_n \\ \dots & \dots & \dots & \dots \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \dots & \varepsilon_n^2 \end{pmatrix} =$$

$$= \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_n) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_n) \\ \dots & \dots & \dots & \dots \\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \dots & M(\varepsilon_n^2) \end{pmatrix} = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon_2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{\varepsilon_n}^2 \end{pmatrix}$$

Таким чином,

$$cov(\varepsilon \cdot \varepsilon^T) = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon_2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{\varepsilon_n}^2 \end{pmatrix}$$

*Друга група*. Моделі, для яких виконуються такі умови стосовно компонент випадкового вектора збурень $\varepsilon$:

1) вони мають нульові математичні сподівання:

$$M(\varepsilon_i) = 0, \ i = \overline{1, n};$$

2) вони є попарно корельованими:

$$cov(\varepsilon_i, \varepsilon_j) = \begin{cases} k_{ij}, & i \neq j, \\ \sigma_{\varepsilon_i}^2 = \sigma_{\varepsilon_j}^2 = const, & i = j, \ i = \overline{1, n}, \end{cases}$$

де $k_{ij} = M(\varepsilon_i \cdot \varepsilon_j) - M(\varepsilon_i) \cdot M(\varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) \neq 0$.

Не зважаючи на те, що дисперсії випадкових відхилень $\varepsilon_i$, $\varepsilon_j$ є сталими величинами, між ними має місце кореляційний зв'язок.

Коваріаційна матриця має вигляд

$$cov(\varepsilon \cdot \varepsilon^T) = \begin{pmatrix} \sigma_\varepsilon^2 & k_{12} & k_{13} & \dots & k_{1n} \\ k_{21} & \sigma_\varepsilon^2 & k_{23} & \dots & k_{2n} \\ k_{31} & k_{32} & \sigma_\varepsilon^2 & \dots & k_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & k_{n3} & \dots & \sigma_\varepsilon^2 \end{pmatrix}$$

Оскільки виконується властивість $k_{ij} = k_{ji}$ (матриця є симетричною), то в цих моделях, не зважаючи на виконання умови гомоскедастичності (сталість дисперсій залишків), використання звичайного МНК є недоцільним внаслідок існування коваріаційних моментів між випадковими залишками.

*Третя група*. Моделі, для яких виконуються такі умови стосовно компонент випадкового вектора збурень $\varepsilon$:

1) вони мають нульові математичні сподівання:

$$M(\varepsilon_i) = 0, \ i = \overline{1, n};$$

2) елементи $\varepsilon$ попарно корельовані:

$$cov(\varepsilon_i, \varepsilon_j) = \begin{cases} k_{ij}, & i \neq j \\ \sigma_{\varepsilon_i}^2 \neq const, & i = j, \ i = \overline{1, n}. \end{cases}$$

Для всіх трьох груп лінійних моделей з порушенням передумов застосування МНК точкові статистичні оцінки $b_i$ для теоретичних параметрів $\beta_i$ будуть *незміщеними*, але *не ефективними*, що призведе до зниження ймовірності одержання доброякісної оцінки.

Для моделі першої групи статистична оцінка параметрів здійснюється використанням *зваженого МНК*, для моделей другої та третьої груп – *узагальненого МНК* (*УМНК*).

## 5.2. Узагальнена лінійна модель множинної регресії

Умова гомоскедастичності є головною для лінійної класичної моделі і записується так:

$$cov(\varepsilon \cdot \varepsilon^{\mathrm{T}}) = \mathrm{M}(\varepsilon \cdot \varepsilon^{\mathrm{T}}) = \sigma_\varepsilon^2 E.$$

Використання звичайного МНК для лінійних моделей з властивістю випадкового вектора $\boldsymbol{\varepsilon}$, коли

$$cov(\varepsilon \cdot \varepsilon^T) = \begin{pmatrix} \sigma_1^2 & k_{12} & k_{13} & \dots & k_{1n} \\ k_{21} & \sigma_2^2 & k_{23} & \dots & k_{2n} \\ k_{31} & k_{32} & \sigma_3^2 & \dots & k_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & k_{n3} & \dots & \sigma_n^2 \end{pmatrix} = S$$

є неможливим для визначення статистичних оцінок, як це було здійснено для лінійної класичної моделі. У такому випадку використовують узагальнений метод найменших квадратів (УМНК), що розглядається у п.5.3.

Узагальнена лінійна модель множинної регресії (Generalized Linear Multiple Regression model) має вигляд

$$Y = X \cdot \beta + \varepsilon, \tag{5.2}$$

де $Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{12} & & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & x_{31} & x_{32} & & x_{3k} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$ – невипадкова матриця;

$\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)^{\mathrm{T}} -$ вектор параметрів моделі,

$\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_n)^{\mathrm{T}} -$ випадковий вектор.

Економетрична модель (5.2) задовольняє таким умовам:

1. $M(\varepsilon) = 0$;
2. $cov(\varepsilon \cdot \varepsilon^T) = M(\varepsilon \cdot \varepsilon^T) = S$, де $S$ - додатно-визначена матриця;
3. Факторні ознаки $x_1, x_2, \dots, x_k$ не залежні між собою;
   $rang(X) = k + 1 < n$.

Узагальнена модель відрізняється від класичної тільки *виглядом коваріаційної матриці*: замість $cov(\varepsilon \cdot \varepsilon^T) = \sigma^2 E$ для класичної моделі маємо $cov(\varepsilon \cdot \varepsilon^T) = S$ для узагальненої. Це означає, що на відміну від класичної, в узагальненій моделі коваріація і дисперсія факторів можуть бути довільними. У цьому полягає сутність узагальнення регресійної моделі.

## 5.3.Узагальнений метод найменших квадратів

При виявленні гетероскедастичності для її вилучення початкову модель трансформують так, щоб залишки мали постійну дисперсію. Невідомі параметри трансформованої моделі розраховують за допомогою МНК. Трансформація моделі проводиться залежно від специфічної форми гетероскедастичності,тобто від форми залежності між дисперсією $\sigma_\varepsilon^2$ та значеннями незалежних змінних: $\sigma_\varepsilon^2 = f(x_i)$. Економетрична модель, що має гетероскедастичність, є узагальненою моделлю. Розглянемо УМНК, яким будемо користуватися для оцінювання її параметрів. Нехай задана економетрична модель (5.2).

Розглянемо такі випадки гетероскедастичності:

а) дисперсії залишків пропорційні $i$-им значенням $j$-их ознак $x_{ij}$, тобто $\sigma_{\varepsilon_i}^2 = \sigma^2 x_{ij}$;

б) дисперсії залишків пропорційні квадратам $i$-их значень $j$-их ознак $x_{ij}^2$, тобто $\sigma_{\varepsilon_i}^2 = \sigma^2 x_{ij}^2$;

в) відомі значення дисперсії залишків $\sigma_{\varepsilon_i}^2$.

$\sigma^2$ – скінчена постійна дисперсія нових змінених значень випадкових величин $\varepsilon_i/\sqrt{x_{ij}}$ , $\varepsilon_i/x_{ij}$, $\varepsilon_i/\sigma_{\varepsilon_i}$ відповідно.

Випадки а) і б) застосовуються у разі чистої гетероскедастичності. Випадок в) відноситься до ідеальної ситуації, яка на практиці нездійснена, а тому його наведено лише з ілюстративною метою.

Задача полягає в знаходженні оцінок елементів вектора $B$ в моделі (5.2). Для цього використовується матриця $S$, за допомогою якої відбувається трансформація вихідної інформації. Ця ідея покладена в основу УМНК (методу Ейткена, [1]).

Матриця $S$ – додатно-визначена матриця, тому вона може бути подана як добуток двох невироджених матриць $PP^T$, тобто:

$$S = P \cdot P^T. \tag{5.3}$$

Помножимо у рівності (5.3) обидві частини зліва на обернену матрицю $P^{-1}$, а справа – на обернену транспоновану матрицю $(P^T)^{-1}$ (матриці не комутативні, тому порядок множення має значення). В результаті отримаємо:

$$P^{-1} \cdot S \cdot (P^T)^{-1} = P^{-1} \cdot P \cdot P^T \cdot (P^T)^{-1}.$$

Оскільки $P^{-1} \cdot P = P^T \cdot (P^T)^{-1} = E$, то

$$P^{-1} \cdot S \cdot (P^T)^{-1} = E.$$

Помножимо обидві частини рівняння (5.3) на обернену матрицю $P^{-1}$. Трансформована модель буде мати такий вигляд:

$$Y^* = B\,X^* + e^*, \tag{5.4}$$

де $Y^* = P^{-1} \cdot Y$, $X^* = P^{-1} \cdot X$, $e^* = P^{-1} \cdot e$.

Доведено, що модель (5.4) задовольняє умові незмінності дисперсії залишків. Оцінимо параметри трансформованої моделі на основі МНК:

$$B = (X^{*T} X^{*})^{-1} X^{*T} Y^* = (X^T S^{-1} X)^{-1} X^T S^{-1} Y\ ,$$

тобто

$$B = (X^T S^{-1} X)^{-1} X^T S^{-1} Y \tag{5.5}$$

Оцінка параметрів $B$, яку знайдено за допомогою формули (5.5), є оцінкою УМНК (методу Ейткена).

Матриця $S$ визначається за формулою

$$S = \begin{pmatrix} 1/\lambda_1 & 0 & 0 & \dots & 0 \\ 0 & 1/\lambda_2 & 0 & \dots & 0 \\ 0 & 0 & 1/\lambda_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1/\lambda_{n-c} \end{pmatrix}.$$

Обернена матриця має вигляд:

$$S^{-1} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_{n-c} \end{pmatrix}$$

Визначення числа $c$ дається далі (п. 5.7). Значення $\lambda_i$ у матриці $S^{-1}$ обчислюються залежно від випадку гетероскедастичності а) – в):

а) $\lambda_i = 1/x_{ij}$; б) $\lambda_i = 1/x_{ij}^2$; в) $\lambda_i = 1/\sigma_{\varepsilon_i}^2$ .

Модель УМНК іноді специфікується у вигляді

$$Y = X\,B + e,$$

$$M(e) = 0,$$

$$M(e \cdot e^T) = V,$$

де квадратна матриця $n$-го порядку $V = \sigma_e^2 S$ є симетричною додатно-визначеною.

Вираз для оцінки параметрів, згідно з методом Ейткена, має вигляд:

$$B = (X^T V^{-1} X)^{-1} X^T V^{-1} Y.$$

Дисперсії визначаються за формулами:

$$\sigma_y^2 = \frac{Y^T S^{-1} Y}{n-1}; \quad \sigma_x^2 = \frac{B^T X^T S^{-1} Y}{m-1}; \quad \sigma_e^2 = \frac{e^T S^{-1} e}{n-m},$$

де $n$ – кількість спостережень, $m$ – число параметрів моделі.

*Різниця між звичайним та узагальненим МНК* [23]

За звичайним МНК невідомі параметри знаходяться мінімізацією суми квадратів відхилень фактичних значень від теоретичних.

Для простої лінійної регресії:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_1)^2 \to min.$$

В УМНК мінімізується вираз мінімізується зважена сума квадратів відхилень з вагами, обернено пропорційними до середньоквадратичних відхилень $\sigma_i$.

$$\sum_{i=1}^{n}\frac{1}{\sigma_i^2}e_i^2=\sum_{i=1}^{n}\frac{1}{\sigma_i^2}(y_i-b_0{}^*-b_1{}^*x_1)^2\rightarrow min. \qquad (5.6)$$

Спостереження з більшими значеннями $\sigma_i$ отримають у цьому методі відповідно менші значення ваги, і навпаки, спостереження з меншими значеннями $\sigma_i$ отримають пропорційно більшу вагу при мінімізації суми квадратів відхилень. Вираз (5.6) відомий як зважені найменші квадрати (ЗНК), а оцінки параметрів, отримані за цим методом, - як ЗНК-оцінки. У контексті гетероскедастичності можна користуватися обома термінами: і ЗНК, і УНК.

## 5.4. Прогноз

У випадку, коли параметри економетричної моделі оцінюються за допомогою УМНК, проблема прогнозування потребує спеціального дослідження. Так, в економетричній моделі (5.2) задача зводиться до того, щоб передбачити значення залежної змінної $Y_0$ для заданого вектора $X_0$. Можна записати

$$Y_0 = X_0\,B + e_0\,,$$

де $e_0$ – невідоме значення відхилень у прогнозний період.

Нехай для $e_0$ виконуються умови:

$$M(e) = 0,$$

$$M(e\cdot e^T) = \sigma_e{}^2 S,$$

$$W = M(e_0 e) = \begin{pmatrix} M(e_0e_1) \\ M(e_0e_2) \\ \dots \\ M(e_0e_n) \end{pmatrix},$$

де $W$ - вектор коваріації поточних і прогнозних значень залишків.

Лінійний прогноз має вигляд:

$$Y_{np} = c^T\,Y,$$

де $c$ – $n$-вимірний вектор, який має мінімізувати дисперсію прогнозу:

$$\sigma_{\text{пр}}^2 = M\big[(Y_{\text{пр}} - Y_0)\big]\,.$$

Розв’язавши задачу на екстремум, знайдемо *c*. Найкращий лінійний незміщений прогноз за моделлю, оцінки якої знайдені за методом Ейткена, обчислюється за формулою:

$$Y_{\text{пр}}=X_0\,B + W^T\,V^{-1}\,e,$$

де $e\ =Y - X{\cdot}B$ – вектор залишків, що відповідає оцінці параметрів моделі за МНК.

## 5.5. Тестування наявності гетероскедастичності [11]

Інколи на підставі знань про характер статистичних даних появу проблеми гетероскедастичності можна передбачити і спробувати її усунути ще на етапі специфікації кореляційно-регресійної моделі, для цього провівши глибокий аналіз досліджуваної проблеми. Проте значно частіше цю проблему доводиться вирішувати вже після побудови регресійної моделі. Виявлення гетероскедастичності у кожному разі є досить складним завданням, оскільки для знання дисперсії відхилень $\sigma^2(\varepsilon_i) = \sigma_i^2$ потрібно знати закон розподілу випадкової величини $\varepsilon$, що відповідає вибраному значенню факторної ознаки $x_i$. Часто на практиці для кожного конкретного значення $x_i$ визначають лише одне значення $y_i$, що не дає можливості оцінити дисперсію випадкових величин $\varepsilon$ для даного $x_i$.

Як і у випадку мультиколінерності, єдиних правил виявлення наявності гетероскедастичності немає, а є тільки різноманітні тести, зокрема, графічний аналіз відхилень та аналітичні методи тестування.

### 5.5.1. Графічний аналіз

Графічний метод є досить наочним та простим методом тестування припущення про наявність гетероскедастичності. Дослідник не завжди володіє необхідним для аналізу проблеми емпіричним матеріалом. Крім того, його висновки щодо наявності або відсутності гетероскедастичності носять суб’єктивний характер, і в цих умовах використовуються графіки. У цьому разі по осі абсцис відкладають значення факторної ознаки $x_j$ (або значення лінійної комбінації факторних ознак $\hat{y}_i = b_0 + b_1 x_1 + ... + b_k x_k$ ), а по осі ординат – або відхилення $e_i$, або їх квадрати $e_i^2$.

Спочатку треба зробити аналіз регресійної моделі на базі припущення про відсутність гетероскедастичності. Потім за допомогою графіків проаналізувати, чи квадрати залишків $e_i^2$ систематично залежать від упорядкованих значень факторної ознаки (або оцінених середніх значень результуючої змінної), чи не залежать. Не зважаючи на те, що квадрати залишків $\mathrm{e}_\mathrm{i}^2$ – це тільки оцінки невідомих $\varepsilon_i^2$, вони можуть успішно використовуватися, особливо при великих обсягах вибірки.

Досліджуючи квадрати залишків, можна отримати, наприклад, такі види графіків (рис. 5.2 а) - г)). На рис. 5.2 (а) – г)) квадрати залишків розсіяні навколо оціненої лінії регресії. Потрібно з’ясувати, чи оцінене середнє значення *y* систематично пов’язано з квадратом залишків.

На рис. 5.2 а) всі квадрати залишків $\mathrm{e}_\mathrm{i}^2$ містяться в середині смуги постійної ширини, що паралельна осі абсцис. Це свідчить про незалежність дисперсій параметрів регресій від значень незалежної змінної та їх постійність, тобто наявне явище явище гомоскедастичності.

На рис. 5.2 б) – г) спостерігаються деякі систематичні зміни у співвідношеннях між значеннями незалежної змінної $x_i$ і квадратами залишків $\mathrm{e}_\mathrm{i}^2$. На рис. 5.2 б) — квадратична залежність, на рис. 5.2 в) - лінійний зв’язок, а на рис. 5.2 г) - гіперболічна залежність.

Графічний аналіз випадкових відхилень допомагає не тільки виявити гетероскедастичність, а й зробити висновок про саму форму зв’язку, що особливо корисно під час трансформування наявних даних для побудови моделі з гомоскедастичністю відхилень. Графічний аналіз найдоцільніше проводити при великій кількості факторних ознак.

### 5.5.2. Аналітичні методи тестування гетероскедаксичності [18]

Як зазначалося раніше, немає однозначного методу визначення гетероскедастичності. Існують тільки апробовані тести, за допомогою яких можна виявити наявність гетероскедастичності. Використання їх дає позитивні наслідки. Для тестування використовують такі три групи методів.

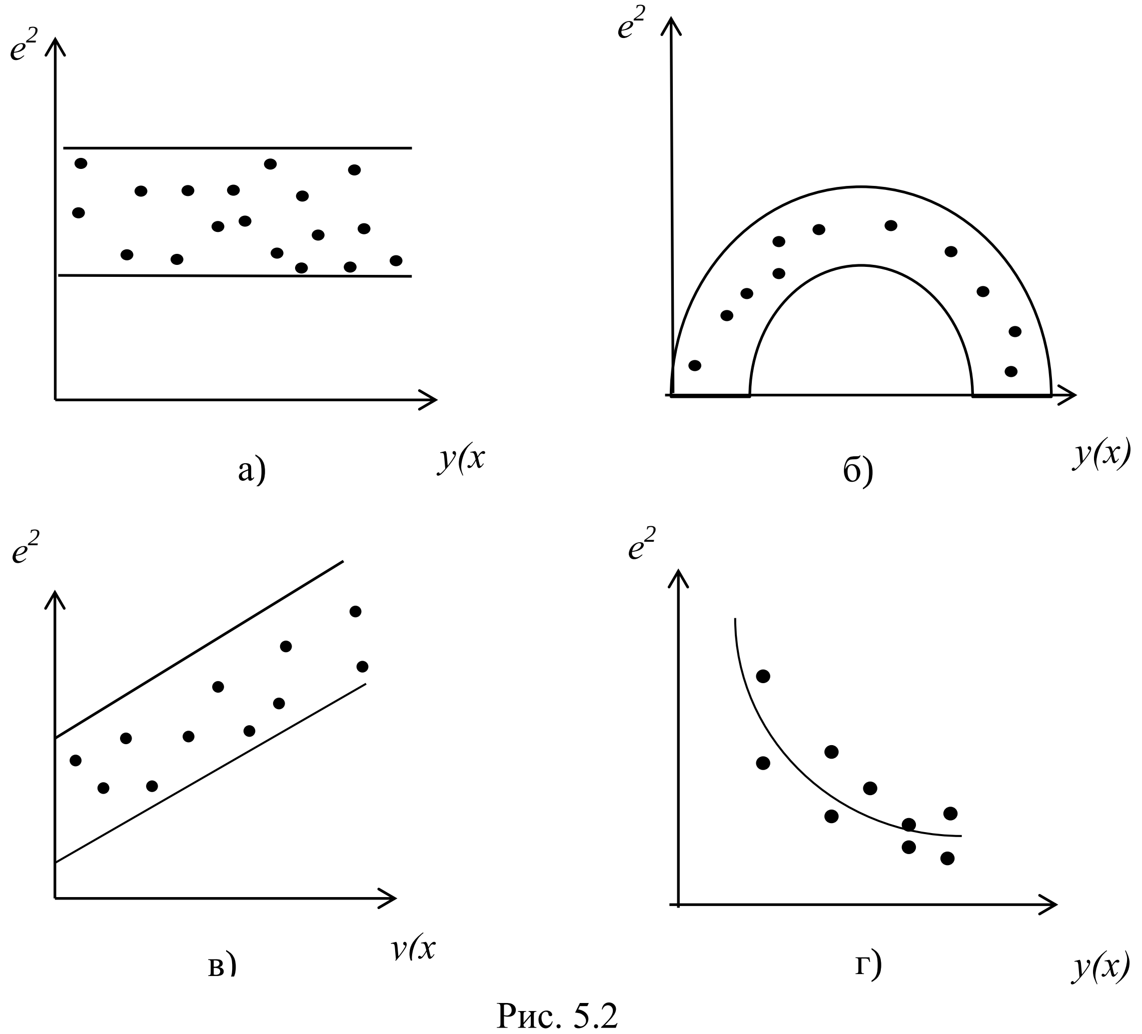



Рис. 5.2

*Методи першої групи* (зокрема, критерій Бартлетта) дають можливість проводити тестування гетероскедастичності, яка задається довільною неперервною функцією $\sigma^2(e_i)$.

Для *методів другої групи* (зокрема, тест рангової кореляції Спірмена та тест Голдфельда - Квандта) функція $\sigma^2(e_i)$ має бути монотонною.

У *методах третьої групи* (зокрема, тест Парка, тест Глейзера, тест Годфрея) функцію $\sigma^2(e_i)$ вважаються відомою (задано форму функціональної залежності та кількість параметрів).

Тести Глейзера, Гольдфельда - Квандта можуть виявити присутність ознаки гетероскедастичності в моделі лише у випадку порушення умови $\sigma_{\varepsilon_i}^2 = const$, тобто, якщо дисперсії залишків $\varepsilon_i$ не сталі величини. При виконанні умови $\sigma_{\varepsilon_i}^2 = const$ разом з умовою $cov(\varepsilon_i,\ \varepsilon_j) \neq 0$ виявити це порушення умови гомоскедастичності тестами Глейзера та Гольдфельда - Квандта неможливо. У цьому випадку йдеться річ про лінійну модель з порушенням ознаки гомоскедастичності, яка належить до другої групи. Така ситуація виникає при дослідженні моделей із ознакою автокореляції, які будуть розглянуті у наступному розділі.

Розглянемо найбільш поширені підходи визначення існування гетероскедастичності:

- тест Глейзера;
- тест Гольдфельда - Квандта.

## 5.6. Тест Глейзера

За цим тестом оцінюється кореляційно-регресійна модель залежності модулів відхилень $|e_i|$ (які тісно пов'язані з середнім квадратичним відхиленням $\sigma_{\varepsilon_i}$) від значень факторних ознак $x_i$.

Залежність між $|e_i|$ і $x_i$ моделюють за допомогою такої кореляційно-регресійної моделі:

$$|e_i| = \alpha + \beta\, x_i^p + v_i. \quad (i = \overline{1,n}), \tag{5.7}$$

де $p$ може побудувати значень …-2, -1, -0,5, 0,5, 1, 2, …, а випадкові відхилення $v_i$ мають нормований нормальний закон розподілу ймовірностей ($v_i \sim N(0, 1)$).

Оскільки фактична форма залежності невідома, то можна підбирати різні види залежності (5.7), зокрема:

$$|e| = \alpha + \beta \frac{1}{x} + v;$$

$$|e| = \alpha + \beta \frac{1}{\sqrt{x}} + v;$$

$$|e| = \alpha + \beta \sqrt{x} + v; \tag{5.8}$$

$$|e| = \alpha + \beta x + v;$$

$$|e| = \alpha + \beta x^2 + v;$$

$$|e| = \sqrt{\alpha + \beta x} + v;$$

$$|e| = \sqrt{\alpha + \beta x^2} + v.$$

Про наявність гетероскедастичності свідчить статистична значущість параметрів $\alpha$ і $\beta$, а також високий коефіцієнт кореляції між змінними моделі. Статистичну значущість параметрів $\alpha$ і $\beta$ моделі визначають за допомогою $t$-статистики Ст'юдента перевіркою нульових гіпотез.

| | |
|---|---|
| Гіпотези | Нульові гіпотези $H_0$:<br>- вільний член α моделі дорівнює нулю: $\alpha = 0$;<br>- коефіцієнт регресії $\beta$ моделі дорівнює нулю: $\beta = 0$ |
| | Альтернативна гіпотеза $H_\alpha$:<br>- вільний член α моделі не дорівнює нулю: $\alpha \neq 0$;<br>- коефіцієнт регресії $\beta$ моделі не дорівнює нулю: $\beta \neq 0$ |

Якщо для декількох кореляційно-регресійних моделей (5.8) коефіцієнт регресії $\beta$ є статистично значущим, то «найкращу» з цих моделей вибирають, порівнюючи їхні коефіцієнти кореляції та середньоквадратичні відхилення параметрів $\alpha$ і $\beta$. В цьому випадку при визначенні форми регресії рекомендується орієнтуватися на коефіцієнт, що більш сприйнятливий в розумінні подальших обчислень.

При $\alpha = 0$ і $\beta \neq 0$ має місце явище *чистої гетероскедастичності*, а при $\alpha \neq 0$ і $\beta \neq 0$ - наявна *змішана гетероскедастичність*.

У разі дослідження множинної кореляційно-регресійної моделі, якщо розглядається залежність (5.7), то замість значення факторної ознаки $x_{ij}$ застосовують лінійну комбінацію значень факторних ознак $x_{ij}$, тобто будують кореляційно-регресійну залежність між модулем відхилення $|e_i|$ та теоретичним значенням результуючої змінної.

У тесті Глейзера для відхилень $v_i$ може порушуватися умова гомоскедастичності, однак у багатьох випадках запропоновані моделі є придатними для визначення гетероскедастичності. Як свідчить практика, цей тест статистично обгрунтований, а тому має широке застосування. У разі множинної лінійної регресії значення $e_i$ моделюються по кожному фактору, включеному до моделі, і в подальшому послідовно виконуються операції з виявлення ознак гетероскедастичності.

*Послідовність застосування тесту* [9]

*Крок* 1. За звичайним МНК визначають статистичні оцінки $b_0$ та $b_1$ у випадку парної регресії та статистичні оцінки $b_0$ та $b_j$ $(j = \overline{1,k})$ у випадку множинної регресії.

*Крок* 2. За одержаними статистичними оцінками $b_0$ , $b_1$, або $b_0$, $b_j$ $(j = \overline{1,k})$ будується модель функції регресії. З одержаної моделі визначаються відхилення $e_i$.

*Крок* 3. Побудова моделі залежності відхилень $|\, e_i \,|$ від обраного фактора $x_i$ у вигляді (5.18).

*Крок* 4. Для вибраного значення $p$ лінеаризованої моделі, використовуючи МНК, визначаються статистичні оцінки $b_0$ та $b_1$ для параметрів $\beta_0$ та $\beta_1$ та $S_{b_1}$.

*Крок* 5. Перевірка на статистичну значущість всіх параметрів моделі при заданому рівні значущості α за допомогою статистичного $F$-критерію Фішера.

*Крок* 6. Перевірка на статистичну значущість коефіцієнта $\beta$. Для цього за вибраним рівнем значущості α перевіряється правдивість гіпотези $H_0$: $\beta = 0$ за альтернативної гіпотези $H_\alpha$: $\beta \neq 0$. Перевірка здійснюється за допомогою статистичного критерію

$$t = \frac{\beta^* - \beta}{S_{\beta^*}},$$

який є випадковою величиною із законом розподілу Ст'юдента та ступенем вільності $k_1 = n - k - 1$.

*Крок* 7. Формулювання альтернативної гіпотези дає можливість побудови двобічної критичної області. Для цього за вибраним рівнем значущості α і числом ступенів вільності $k_1 = n - k - 1$ (для парної регресії $k_1 = n - 2$) за таблицею 2 Додатку знаходимо значення $t''_{\text{табл}}\left(\frac{\alpha}{2}; k_1\right)$. Враховуючи, що $t'_{\text{табл}}\left(\frac{\alpha}{2}; k_1\right) = -t''_{\text{табл}}\left(\frac{\alpha}{2}; k_1\right)$, будуємо двобічну критичну область (рис 5.3).


Рис. 5.3

Після обчислення спостережуваного значення критерію за формулою

$$t_\beta = \beta / S_\beta ,$$

визначається можливість прийняття однієї з гіпотез:

- якщо $t_\beta \in (-\infty;\ t'_{\text{табл}}(\alpha/2\,; k_1)] \cup [t''_{\text{табл}}(\alpha/2\,; k_1); \infty)$, то параметр $\beta$ є статистично значущим і не дорівнює нулю ($\beta \neq 0$), що дає можливість зробити висновок про наявність в досліджуваній моделі ознаки гетероскедастичності;
- якщо $t_\beta \in [t'_{\text{табл}}(\alpha/2\,; k_1); t''_{\text{табл}}(\alpha/2\,; k_1)]$, то гіпотеза $H_0$: $\beta = 0$ виявляється справедливою, що свідчить про малозначущість коефіцієнта $\beta$ і відсутність в моделі явища гетероскедастичності.

## 5.7. Параметричний тест Гольдфельда - Квандта

Якщо сукупність спостережень невелика, то розглянутий вище тест не використовується. У цьому випадку Гольдфельд і Квандт запропонували параметричний тест.

| | |
|---|---|
| Використовується | до великих обсягів вибірок |
| Припущення | 1) незалежність випадкових величин $\varepsilon_i$;<br>2) нормальний розподіл випадкових величин $\varepsilon_i$;<br>3) дисперсії залишків пропорційні квадратам $i$-их значень $j$-их ознак $x_{ij}^2$, тобто $\sigma_{e_i}^2 = \sigma^2 x_{ij}^2$. |
| Гіпотези | Нульова гіпотеза $H_0$:<br>випадкова величина $\varepsilon_i$ гомоскедастична |
| | Альтернативна гіпотеза $H_1$:<br>випадкова величина $\varepsilon_i$ гетероскедастична (із зростаючою дисперсією) |

Тест Гольдфельда-Квандта складається за таким алгоритмом [16].

*Крок* 1. Ранжування спостережень незалежних змінних $x_i$ у порядку зростання або спаду величин. У випадку багатофакторних регресій, коли має місце декілька незалежних змінних, вибирається одна з них і для неї проводиться ранжування, тобто впорядковуємо решту змінних відповідно до значень вибраної змінної. Якщо відразу важко вибрати таку змінну для впорядкування, то треба почергово провести впорядковування за кожною змінною окремо і в кожному випадку застосувати тест Гольдфельда-Квандта.

*Крок* 2. Виключення з подальшого аналізу $c$ спостережень, які містяться в центрі упорядкованого ряду загальної кількості спостережень. Згідно з експериментальними розрахунками, кількість виключених спостережень пропонується приймати за формулою:

$$c = [(4/15)\cdot n],$$

де $n$ – кількість елементів вектора, $[q]$ – ціла частина числа $q$.

Доцільно керуватися таким правилом: якщо $n > 30$, то $c = n / 4$, тобто відкидається чверть спостережень.

*Крок* 3. Залишок ($n - c$) спостережень поділяється на дві підвибірки (нижня та верхня) однакового розміру $(n - c) / 2$, причому перша з них включає малі значення факторних ознак $x_i$ (у випадку ранжування за зростанням), а друга – великі значення факторних ознак $x_i$. При формуванні двох підвибірок даних $x_i$ треба слідкувати за тим, щоб виконувалась умова

$$n_1 = n_2 = (n-c)/2 > m,$$

де $m$ – загальна кількість оцінюваних параметрів у моделі. Якщо $n_1 \neq n_2$, то відкидається перше або останнє спостереження сукупності.

*Крок* 4. Побудова двох економетричних моделей за допомогою МНК за двома утвореними сукупностями спостережень. Знаходження видів їх регресій та сум квадратів залишків $S_1$ та $S_2$ відповідно за першою та другою моделями

$$S_1 = \sum_{i=1}^{\frac{n-c}{2}} \left(y_i^{(1)} - \hat{y}_i^{(1)}\right)^2 = \sum_{i=1}^{\frac{n-c}{2}} \left(e_i^{(1)}\right)^2 = \left(e^{(1)}\right)^T \left(e^{(1)}\right)$$

$$S_2 = \sum_{i=(n+c)/2}^{n} (y_i^{(2)} - \hat{y}_i^{(2)})^2 = \sum_{i=(n+c)/2}^{n} \left(e_i^{(2)}\right)^2 = \left(e^{(2)}\right)^T \left(e^{(2)}\right),$$

де $n$ – загальна кількість спостережень;

$c$ – кількість центральних неврахованих спостережень;

$y_i$ – експериментальні значення;

$\hat{y}_i$ – теоретичні (знайдені за допомогою рівняння регресії) значення;

$e_i^{(1)} = (y_i^{(1)} - \hat{y}_i^{(1)})^2$ – залишки за першою моделлю;

$e_i^{(2)} = (y_i^{(2)} - \hat{y}_i^{(2)})^2$ - залишки за другою моделлю.

*Крок* 5. Обчислення значення емпіричного $F$-критерію:

$$F_{\text{факт}} = S_2/S_1, \tag{5.9}$$

який в разі виконання гіпотези про гомоскедастичність $H_0$ наближено відповідає $F$-критерію з ступенями вільності $k_1 = k_2 = (n - c - 2m) / 2$.

Дисперсії $D_1$ і $D_2$ за першою і другою вибірками дорівнюють

$$D_1=S_1/k_1, \quad D_2=S_2/k_1.$$

Із табл.1 Додатку при заданому рівні значущості $\alpha$ та кількістю ступенів вільності $k_1 = k_2 = (n - c - 2m) / 2$ знаходиться критичне значення $F$-статистики. Якщо $F_{\text{факт}} < F_{\text{табл.}}$, то нульова гіпотеза про відсутність гетероскедаксичності приймається. Якщо $F_{\text{факт}} > F_{\text{табл.}}$, то приймається альтернативна гіпотеза $H_1$, тобто має місце гетероскедаксичність.

Тест Гольдфельда-Квандта можна використовувати при припущенні про обернену пропорційну залежність між середньо-квадратичним відхиленнями $\sigma_i$ і значеннями факторних ознак $x_i$, тобто коли дисперсія випадкових величин зменшується із зростанням значень факторних ознак $x_i$. При цьому емпіричне значення $F$-статистики обчислюється за формулою: $F_{\text{факт}} = S_1/S_2$.

## 5.8. Приклад дослідження наявності гетероскедастичності на основі тесту Гольдфельда - Квандта

*Постановка задачі.* На основі даних таблиці 5.1 побудувати економетричну модель, що характеризує залежність витрат на харчування ($Y$, грош. од.) від загальних витрат ($X$, грош. од.) сімей. Перевірити гіпотезу про гомоскедастичність дисперсії залишків. Під час дослідження використати параметричний тест Гольдфельда-Квандта. Оцінити параметри моделі на основі метода Ейткена.

***Розв'язання***

1) *Побудова економетричної моделі, яка характеризує залежність витрат на харчування від загальних витрат сімей.*

Ідентифікація змінних:

$Y$ – витрати на харчування сімей (залежна змінна);

$X$ – загальні витрати сімей (незалежна змінна).

Специфікація. Вибіркова модель має вигляд: $\hat{y} = b_0 + b_1 x$.

Таблиця 5.1

| $i$ | $X$ | $Y$ |
|---|---|---|
| 1 | 45 | 5,30 |
| 2 | 45 | 5,20 |
| 3 | 46 | 5,08 |
| 4 | 47 | 5,20 |
| 5 | 47 | 5,10 |
| 6 | 48 | 5,32 |
| 7 | 98 | 5,40 |
| 8 | 50 | 5,50 |
| 9 | 50 | 5,20 |
| 10 | 52 | 5,50 |
| 11 | 94 | 6,10 |
| 12 | 49 | 5,45 |
| 13 | 102 | 5,82 |
| 14 | 110 | 6,04 |
| 15 | 115 | 5,70 |
| 16 | 130 | 6,99 |
| 17 | 125 | 6,10 |
| 18 | 120 | 6,94 |

2). *Перевірка на наявність гетероскедастчності за допомогою параметричного тесту Гольдфельда – Квандта.*

*Крок* 1. Ранжування спостережень незалежних змінних у порядку зростання.

Впорядковуємо сукупність спостережень $X$ від меншого значення до більшого (таблиця 5.2).

*Крок* 2. Виключення з подальшого аналізу

$$c = [(4 / 15) \cdot n] = [(4 / 15) \cdot 18] = [4{,}8] = 4$$

спостережень, які містяться в центрі упорядкованого ряду загальної кількості спостережень.

*Крок* 3. Залишок 18 – 4 = 14 спостережень поділяється на дві підвибірки (нижня та верхня) однакового розміру $n_1 = n_2 = 14/2 = 7$.

Оскільки $n = 18$, $c = 4$, $m = 2$ (модель має два параметра $b_0$ і $b_1$), то виконується умова $(n – c) / 2 > m$.

Таблиця 5.2

| *i* | *X* | *Y* |
|---|---|---|
| 1 | 45 | 5,30 |
| 2 | 45 | 5,20 |
| 3 | 46 | 5,08 |
| 4 | 47 | 5,20 |
| 5 | 47 | 5,10 |
| 6 | 48 | 5,32 |
| 7 | 49 | 5,45 |
| 8 | 50 | 5,50 |
| 9 | 50 | 5,20 |
| 10 | 52 | 5,50 |
| 11 | 94 | 6,10 |
| 12 | 98 | 5,40 |
| 13 | 102 | 5,82 |
| 14 | 110 | 6,04 |
| 15 | 115 | 5,70 |
| 16 | 120 | 6,94 |
| 17 | 125 | 6,10 |
| 18 | 130 | 6,99 |

*Крок* 4. Побудова двох економетричних моделей за допомогою МНК за двома утвореними сукупностями спостережень.

*Параметри моделі оцінимо, скориставшись засобами Excel*:

а) виділяємо масив ($5 \times m$) = ($5 \times 2$);

б) використовуємо засіб *Мастер функцій*:

$f_x$ → *Статистические* → ЛИНЕЙН;

в) у вікні, що з'явилося, послідовно у відповідних полях вводимо елементи масиву *Y*, потім масиву *X*, у полі «константа» ставимо 1, у полі «Статистика» теж ставимо 1.

г) після натискання ОК клацаємо на рядок формул;

д) одночасно натискаємо клавіші Ctrl+Shift+Enter.

Таблиця 5.3

| Перша сукупність | | | Друга сукупність | | |
|---|---|---|---|---|---|
| Коеф. регресії | **0,045532** | **3,108723** | Коеф. регресії | **0,040506** | **1,512129** |
| | 0,033317 | 1,557083 | | 0,014373 | 1,650118 |
| | 0,271946 | 0,122092 | | 0,613657 | 0,413953 |
| | 1,867624 | 5 | | 7,941868 | 5 |
| | 0,02784 | **0,074532** | | 1,360898 | **0,856787** |

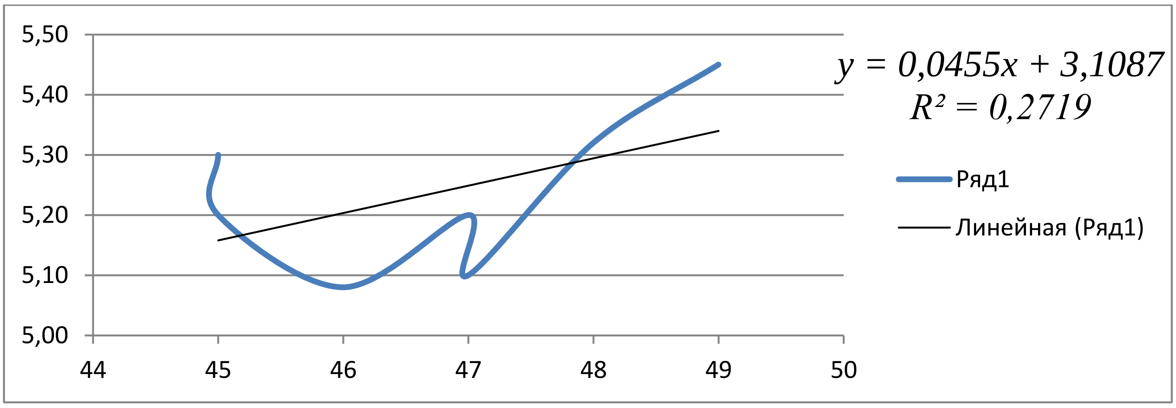
На основі першої сукупності отримаємо таку модель:

$$\hat{y} = 3{,}108723 + 0{,}045532\ x,$$

на основі другої –

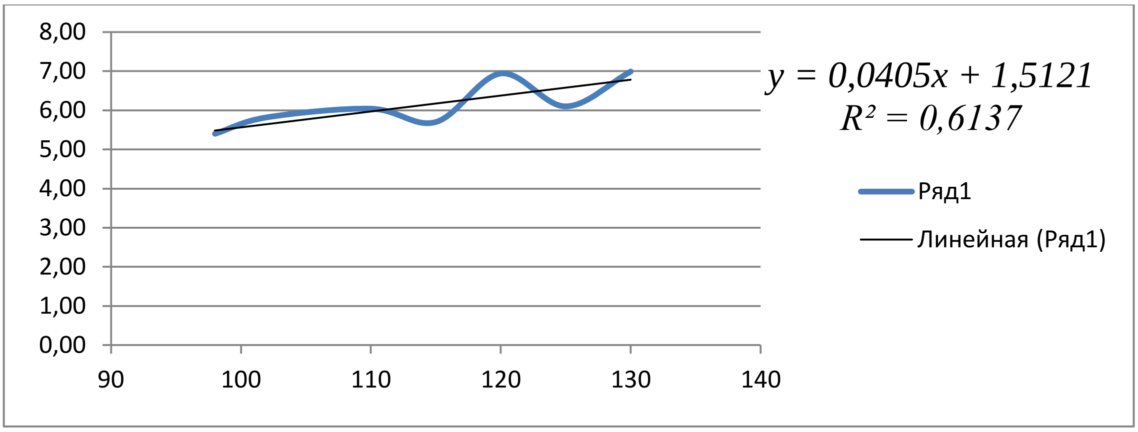
$$\hat{y} = 1{,}512129 + 0{,}040506\ x.$$

Графік функці для для першої сукупності має вигляд:



,

для другої сукупності:




На підставі наведених даних знайдемо суму квадратів залишків. Її можна дістати за допомогою таблиці 5.4, або зі стандартної програми ЛИНЕЙН, значення суми розташовано у другому стовпчику п’ятого рядка таблиці 5.3.

Таблиця 5.4

| $i$ | $X$ | $Y$ | $\hat{Y}$ | $e = Y - \hat{Y}$ | $e^2$ |
|---|---|---|---|---|---|
| 1 | 45 | 5,30 | 5,1577 | 0,1423 | 0,0203 |
| 2 | 45 | 5,20 | 5,1577 | 0,0423 | 0,0018 |
| 3 | 46 | 5,08 | 5,2032 | -0,1232 | 0,0152 |
| 4 | 47 | 5,20 | 5,2487 | -0,0487 | 0,0024 |
| 5 | 47 | 5,10 | 5,2487 | -0,1487 | 0,0221 |
| 6 | 48 | 5,32 | 5,2943 | 0,0257 | 0,0007 |
| 7 | 49 | 5,45 | 5,3398 | 0,1102 | 0,0121 |
| | 50 | 5,50 | | **Сума $S_1$=** | **0,07453** |
| | 50 | 5,20 | | | |
| | 52 | 5,50 | | | |
| | 94 | 6,10 | | | |
| 1 | 98 | 5,40 | 5,481753 | -0,0818 | 0,0067 |
| 2 | 102 | 5,82 | 5,643779 | 0,1762 | 0,0311 |
| 3 | 110 | 6,04 | 5,96783 | 0,0722 | 0,0052 |
| 4 | 115 | 5,70 | 6,170362 | -0,4704 | 0,2212 |
| 5 | 120 | 6,94 | 6,372894 | 0,5671 | 0,3216 |
| 6 | 125 | 6,10 | 6,575425 | -0,4754 | 0,2260 |
| 7 | 130 | 6,99 | 6,777957 | 0,2120 | 0,0450 |
| | | | | **Сума $S_2$=** | **0,856787** |

*Крок* 5. Щоб перевірити нульову гіпотезу $H_0$ про наявність гомоскедастичності, розрахуємо за формулою (5.9) фактичне значення $F$-статистики:

$$F_{\text{факт}} = \frac{S_2}{S_1} \approx \frac{0{,}856787}{0{,}07453} \approx 11{,}5.$$

Знайдемо критиче значення $F$-статистики із заданим рівнем значущості $\alpha = 0{,}05$ та кількістю ступенів вільності

$k_1 = k_2 = (n - c - 2m) / 2 = (18 - 4 - 2 \cdot 2) / 2 = 5.$

За таблицею 1 Додатку $F_{\text{табл.}}(\alpha,\ k_1, k_2) = F_{\text{табл.}}(0{,}05;\ 5; 5) = 5{,}05$

Оскільки $F_{\text{факт}} > F_{\text{табл}}$, то з ймовірністю 0,95 нульову гіпотезу $H_0$ відхиляємо, тобто має місце явище гетероскедастичності.

В економетричній моделі для вихідних даних таблиці 5.2 виявлено гетероскедастичність, а тому для оцінювання параметрів моделі не можна застосовувати МНК.

3). *Використання методу Ейткена для оцінювання параметрів моделі*.

За наявністю гетероскедастичності оцінку параметрів моделі виконуємо узагальненим МНК (методом Ейткена) за формулою:

$$B = (X^T S^{-1} X)^{-1} X^T S^{-1} Y.$$

Матриця $X^T$ має вигляд:

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 45 & 45 & 46 & 47 & 47 & 48 & 49 & 98 & 102 & 110 & 115 & 120 & 125 & 130 \end{pmatrix}$$

Обернена матриця $S^{-1}$ є діагональною. Згідно з припущенням про те, що в матриці $S$ діагональні елементи $\lambda_i$ пропорційні $i$-им елементам $x_i$ вектора пояснювальної змінної, будемо мати $\lambda_i = 1/x_i$:

$$\lambda_1 = 1/x_1 = 1/45 \approx 0{,}022, \ \ldots, \ \lambda_{14} = 1/x_{14} = 1/130 \approx 0{,}008.$$

Обернена матриця $S^{-1}$ має вигляд:

$$S^{-1} = \begin{pmatrix} 0{,}02 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0{,}02 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0{,}02 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0{,}02 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & 0{,}00 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0{,}00 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0{,}00 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0{,}00 \end{pmatrix}$$

$$X^T S^{-1} = \begin{pmatrix} 0{,}0222 & 0{,}0222 & 0{,}0217 & \ldots & 0{,}08 \\ 1 & 1 & 1 & \ldots & 1 \end{pmatrix}; \qquad X^T S^{-1} X = \begin{pmatrix} 0{,}212 & 14 \\ 14 & 1127 \end{pmatrix};$$

$$\left(X^T S^{-1} X\right)^{-1} = \begin{pmatrix} 26{,}4 & -0{,}33 \\ -0{,}328 & 0{,}005 \end{pmatrix}; \qquad X^T S^{-1} Y = \begin{pmatrix} 1{,}16 \\ 79{,}6 \end{pmatrix};$$

$$B = \begin{pmatrix} 26,4 & -0,33 \\ -0,328 & 0,005 \end{pmatrix} \cdot \begin{pmatrix} 1,16 \\ 79,6 \end{pmatrix} = \begin{pmatrix} 4,558 \\ 0,0138 \end{pmatrix}.$$

Отже, шукані параметри $b_0 = 4,558$, $b_1 = 0,0138$ .

Таблиця 5.5

| $i$ | $Y_i$ | $\hat{Y}_i$ | $(Y_i - \bar{Y})^2$ | $(\hat{Y}_i - \bar{Y})^2$ |
|---|---|---|---|---|
| 1 | 5,30 | 5,1577 | 0,150988 | 0,28187 |
| 2 | 5,20 | 5,1577 | 0,238702 | 0,28187 |
| 3 | 5,08 | 5,2032 | 0,370359 | 0,235594 |
| 4 | 5,20 | 5,2487 | 0,238702 | 0,193466 |
| 5 | 5,10 | 5,2487 | 0,346416 | 0,193466 |
| 6 | 5,32 | 5,2943 | 0,135845 | 0,155485 |
| 7 | 5,45 | 5,3398 | 0,056916 | 0,12165 |
| 8 | 5,40 | 5,4818 | 0,083273 | 0,042774 |
| 9 | 5,82 | 5,6438 | 0,017273 | 0,002006 |
| 10 | 6,04 | 5,9678 | 0,123502 | 0,077985 |
| 11 | 5,70 | 6,1704 | 0,000131 | 0,232122 |
| 12 | 6,94 | 6,3729 | 1,566073 | 0,468297 |
| 13 | 6,10 | 6,5754 | 0,169273 | 0,78651 |
| 14 | 6,99 | 6,7780 | 1,693716 | 1,186762 |
| ***Σ*** | 79,64 | 79,64 | ***5,191171*** | ***4,259852*** |
| Сер.зн | **5,69** | 5,69 | 0,370798 | 0,304275 |

Економетрична модель затрат на харчування сімей має вигляд:

$$\hat{Y} = 4,558 + 0,0138 \cdot X.$$

Обчислимо коефіцієнт детермінації моделі за формулою

$$R^2 = \sum_{i=1}^{14} (\hat{Y}_i - \bar{Y})^2 / \sum_{i=1}^{14} (Y_i - \bar{Y})^2 = 4,259852/5,191171 \approx 0,82$$

(використали дані таблиці 5.5, де $\bar{Y} = \sum_{i=1}^{14} y_i = 5,69$).

Вектор залишків та транспонований до нього вектор мають відповідно вигляд:

$$e = \begin{pmatrix} 0{,}1423 \\ 0{,}0423 \\ -0{,}1232 \\ -0{,}0487 \\ -0{,}1487 \\ 0{,}0257 \\ 0{,}1102 \\ -0{,}0818 \\ 0{,}1762 \\ 0{,}0722 \\ -0{,}4704 \\ 0{,}5671 \\ -0{,}4754 \\ 0{,}2120 \end{pmatrix}$$

$e^T$ = ( 0,1423 0,0423 -0,1232 …. -0,4754 0,2120 );

$e^T S^{-1}$=( 0,00316 0,00094 …. -0,00380 0,00163);

$e^T S^{-1} e$ = 0,00878.

Залишкова дисперсія за методом Ейткена:

$$\sigma_e^2 = \frac{1}{n-m} \cdot e^T S^{-1} e = \frac{1}{14-2} 0{,}00878 \approx 0{,}0007.$$

*Висновки.*

Збільшення загальних витрат сімей на одну грош. од. призводить до зростання витрат на харчування на 0,0138 грош. од.

Коефіцієнт детермінації показує, що 82% загальної зміни витрат на харчування сімей пояснюється зміною їх загальних витрат, у той час, як на інші фактори доводиться лише 18%. Коефіцієнт детермінації (один з найефективніших показників оцінки адекватності регресійної моделі) свідчить про досить тісний зв'язок між витратами на харчування та загальними витратами сімей.

Лінійний коефіцієнт кореляції 0,906 свідчить про досить тісний зв'язок між витратами на харчування та загальними витратами сімей.

Залишкова дисперсія $\sigma_e^2 \approx 0{,}0007$ вказує на те, що розрахункові значення витрат на харчування дуже близькі до фактичних.

4) *Для порівняння результатів* побудуємо економетричну модель з 18 спостереженнями за допомогою МНК. За допомогою функції ЛИНЕЙН (Мастер функции $f_x$, категорія Статистическая) одержимо такі результати:

| | | |
|---|---|---|
| Коеф. | **0,014365** | **4,567598** |
| | 0,002318 | 0,192465 |
| | 0,705899 | 0,322515 |
| | 38,40308 | 16 |
| | 3,994541 | ***1,664259*** |

Економетрична модель, параметри якої оцінені за допомогою МНК, має вигляд:

$$y = 4{,}567598 + 0{,}014365\,x.$$

Залишкова дисперсія дорівнює $\sigma_e^2 \approx 1{,}664259 \,/\, 16 \approx 0{,}104$. Порівняння характеристик цієї моделі із моделлю, для оцінок параметрів якої використовувався метод Ейткена, свідчить про те, що у даному випадку метод Ейткена дає більш ефективні оцінки.

*Примітка*. Слід підкреслити важливу особливість: якщо за умов гетероскедастичності розрахувати невідомі параметри МНК, то дисперсії цих параметрів уже не можна розрахувати за відомими формулами. В цій ситуації звичайні процедури тестування мають сумнівну, якщо не найнижчу, цінність. Знання дисперсій $\sigma_i^2$ – взагалі рідкість. Тому на практиці найпоширенішим випадком розв’язку гетероскедастичних регресійних моделей залишається їх трансформація. Припускалось, що випадкова змінна $\varepsilon$ має нормальний розподіл. Припущення про нормальність необхідне для перевірки на значимість оцінок параметрів за статистичними тестами та для побудови інтервалів довіри. Якщо це припущення порушується, то оцінки параметрів залишаються найкращими, але не можна визначити їхню статистичну надійність за допомогою класичних тестів значущості ($t$-, $F$- і тощо). На практиці намагаються довести припущення про нормальність на апріорних підставах, вважаючи, наприклад, що випадкова величина головним чином абсорбує вплив численних неважливих факторів та нестійких елементів людської поведінки.

# РОЗДІЛ 6. АВТОКОРЕЛЯЦІЯ У КОРЕЛЯЦІЙНО-РЕГРЕСІЙНОМУ АНАЛІЗІ

## 6.1. Суть автокореляції

Важливим припущенням побудови якісної кореляційно-регресійної моделі є незалежність значень випадкової величини $\varepsilon_i$ від значень цієї ж величини $\varepsilon_j$ у всіх інших спостереженнях. Відсутність кореляції між будь-якими значеннями випадкової величини гарантує припущення 2° кореляційно-регресійного аналізу $cov(\varepsilon_i, \varepsilon_j) = 0$ $(i \neq j)$.

*Автокореляція* [9] - *це кореляція між значеннями результуючої змінної, яка виникає унаслідок залежності значень випадкової величини в різних спостереженнях.*

Отже, причиною появи автокореляції є кореляція між значеннями випадкових величин $\varepsilon_i$ $(i = \overline{1, n})$, яка називається *автокореляцією випадкових відхилень*. Автокореляцію (кореляцію між значеннями випадкової змінної) та автокореляцію випадкових відхилень називають одним терміном: автокореляція.

У кореляційно-регресійному аналізі автокореляція випадкових відхилень здебільшого трапляється під час використання даних динамічних рядів. При аналізі часових рядів часто приходиться враховувати залежність спостережень у різні моменти часу. У таких ситуаціях припущення про некорельованість збурень не виконується. Це зв'язано з тим, що випадковий член $\varepsilon$ піддається впливові тих змінних, які не включені в рівняння регресії, але впливають на залежну змінну $Y$. Однак не виключено, що значення неврахованої змінної в якомусь спостереженні корелює з її значенням у попередніх спостереженнях. Природно очікувати, що і значення $\varepsilon$ у різних спостереженнях будут корелювати між собою.

Автокореляція – це статистичний взаємозв'язок значень послідовних елементів часового чи просторового ряду даних (кореляція всередині вибірки, «між своїми»).

У кореляційно-регресійному аналізі автокореляція випадкових відхилень здебільшого трапляється під час використання даних динамічних рядів. Через це замість символа $i$ порядкового номера спостереження використовують символ $t$, що позначає момент або інтервал часу спостереження, а обсяг вибірки позначимо $T\,(t=\overline{1,T})$.

Розрізняють поняття автокореляції та серійної кореляції.

*Серійною кореляцією* називається залежність між значеннями випадкових відхилень *двох різних* вибірок залишків $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ та $u_2, u_3, \ldots, u_{p+1}$. У посібнику це явище не досліджується.

Далі буде розглядатись власне *автокореляція* - залежність між значеннями випадкових відхилень *однієї вибірки* залишків $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ та $\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_{p+1}$.

## 6.2. Автокореляція першого порядку

Для часових рядів лінійну регресію записують у вигляді

$$y_t = \beta_0 + \beta_1 x_{t_1} + \beta_2 x_{t_2} + \cdots + \beta_k x_{t_k} + \varepsilon_t \quad (t=\overline{1,\mathrm{T}}). \quad (6.1)$$

Якщо випадкові відхилення в різні моменти часу корелюють між собою і не виконується умова 2° кореляційно-регресійного аналізу, то між значеннями випадкового члена в різні моменти часу існує залежність, яку у загальному випадку можна записати у вигляді

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \cdots + \rho_s\varepsilon_{t-s} + u_t\ . \quad (6.2)$$

При виконанні рекурентних співвідношень (6.2) говорять, що послідовність $\varepsilon_t$ утворює авторегресійний процес $s$-го порядку (авторегресійний, оскільки $\varepsilon_t$ визначається значеннями цієї ж самої величини з запізнюванням, а $s$-го порядку, тому що максимальне запізнювання дорівнює $s$). Коефіцієнти $\rho_i$ $(i=\overline{1,s})$ називаються $i$-ми коефіцієнтами автокореляції. Їх значення невідомі, а тому підлягають оцінюванню. Можна перевірити виконання умов теореми Гаусса - Маркова для значень випадкового члена $\varepsilon_t$ в лінійній регресії (6.1) і для значень випадкового члена $u_t$ в рівнянні (6.2).

Розглянемо парну лінійну регресію

$$y_t = \beta_0 + \beta_1 x_t \ + \varepsilon_t \quad (t = \overline{1, \text{T}}). \tag{6.3}$$

Кореляція між сусідніми членами часового ряду (тобто коли $cov(\varepsilon_t, \varepsilon_{t-1}) \neq 0$) називається *автокореляцією першого порядку*. Вона записується у вигляді:

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t \qquad (t = \overline{1, \text{T}})\,. \tag{6.4}$$

Величина $\rho$ є коефіцієнтом автокореляції залишків, що характеризує рівень взаємозв’язку кожного наступного значення з попереднім, тобто коваріацію залишків. Якщо коефіцієнт автокореляції $\rho > 0$ (($cov(\varepsilon_t, \varepsilon_{t-1}) > 0$), то автокореляція називається додатною, а якщо $\rho < 0$ ($cov(\varepsilon_t, \varepsilon_{t-1}) < 0$) – від'ємною. При $\rho = 0$ автокореляція відсутня. На практиці найчастіше має місце додатна автокореляція. Від′ємна автокореляція має місце досить рідко, наприклад, при перетворенні нелінійної форми моделі до лінійної. Додатна автокореляція спричинена напрямленою постійною дією деяких не врахованих у моделі факторів. При від'ємній автокореляції після додатного відхилення виникає від’ємне і навпаки.

Використовуючи рекурентні співвідношення (6.4), можна виразити значення $\varepsilon_t$ через значення $\varepsilon_0$ (яке не визначене) і $u_1, u_2, \ldots, u_t, \ldots$ . Дійсно,

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t = \rho\, (\rho\, \varepsilon_{t-2} + u_{t-1}) + u_t = u_t + \rho\, u_{t-1} + \rho^2 \varepsilon_{t-2} = \ldots =$$
$$= u_t + \rho\, u_{t-1} + \rho^2 u_{t-2} + \ldots + \rho^{t-1} u_1 + \rho^t \varepsilon_0,$$

тобто

$$\varepsilon_t = \sum_{i=0}^{t-1} \rho^i u_{t-i} + \rho^t \varepsilon_0.$$

Щодо невизначеної випадкової величини $\varepsilon_0$ припускаємо, що вона має такий самий закон розподілу, що й величини $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t$.

Можна показати, що

$$\sigma_\varepsilon^2 = \frac{\sigma_u^2}{(1-\rho^2)},$$

звідки випливає умова на коефіцієнт автокореляції:

$$-1 < \rho < 1.$$

Для обчислення коваріацій помножимо рівність (6.4) на $\varepsilon_{t-1}$. Користуючись незалежністю залишків $\varepsilon_{t-1}$ і $u_t$, знаходимо що $\rho = \rho_{\varepsilon_t \varepsilon_{t-1}}$. З цієї рівносі випливає така інтерпретація коефіцієнта автоколяції $\rho$: $\rho$ є в точності коефіцієнт кореляції між двома сусідніми відхиленнями, причому будь-які два сусідніх відхилення корелюють однаково ($\rho$ = const).

Якщо є залежність між значеннями випадкових відхилень однієї вибірки із запізненням у два лаги (зрушення залишків на 2 номера щодо періоду $t$), тобто $cov(\varepsilon_{t-1}, \varepsilon_t) \neq 0$, $cov(\varepsilon_{t-2}, \varepsilon_t) \neq 0$ автокореляцію називають *автокореляцією другого порядку* тощо.

## 6.3. Причини появи явища автокореляції залишків

1. *Неправильне визначення економічних змінних*. До регресійної моделі не включено чинники, що відіграють суттєву роль в економетричній моделі. Корельованість залишків пов’язана дією якогось суттевого фактора, що не включений в модель і який може невипадковим чином впливати на величину відхилень залишків, наприклад, сезонні умови. Якщо рівняння регресії буде розширено за рахунок такого фактора, це може виявитися найкращим розв’язанням проблеми автокореляції.

*2. Невдала специфікація моделі.* Наприклад, якщо залежність описується лінійною моделлю, а насправді зв’язок $y$ з пояснюючими змінними $x_i$ має нелінійний характер. Автокореляція може бути присутньою в одному рівнянні регресії й відсутньою – в іншому. Необхідно будувати альтернативні моделі.

3. При дослідженні економічного явища *числові дані отримано з великими похибками.*

4. *Інерція.* Деяким економічним показникам (наприклад, інфляція, безробіття, валовий внутрішній продукт тощо) притаманна певна

інертність, що пов'язана з хвилеподібним (циклічним) розвитком ділової активності. Економічний підйом призводить до зростання зайнятості, скорочення інфляції, збільшення ВВП тощо. Це зростання продовжується доти, доки зміна кон'юнктури ринку і низки економічних характеристик не спричинить уповільнення зростання, потім припинення зростання, а після цього і спадання цих показників. У будь-якому разі такому циклічному розвитку економічних процесів властива деяка інертність.

5. *Ефект павутини.* У багатьох сферах економіки економічні показники реагують на зміну економічних умов із запізненням (часовим лагом). Наприклад, пропозиція сільськогосподарської продукції реагує на зміну ціни із запізненням, що дорівнює періоду дозрівання урожаю. Висока ціна сільськогосподарської продукції минулого року здебільшого спричинить її перевиробництво у поточному році, а отже, ціна на неї знизиться. У цій ситуаціїне можна припускати випадковості відхилень один від одного.

6. *Згладжування даних*. Дані за деякий тривалий часовий період здебільшого одержують усереднюванням даних за складовими його підінтервалів. Це може призвести до деякого згладжування коливань, які були всередині даного періоду, що, своєю чергою, може слугувати причиною автокореляції.

7. *Перетворення початкової специфікації моделі до лінійної форми*. Автокореляція залишків викликає більш суттєву проблему у випадку, коли інтервал між спостереженнями досить малий. При збільшенні інтервалу між спостереженнями вплив неврахованих змінних буде зменшуватися.

Автокореляція в цілому виявляє тим більш суттєву проблему, чим менший інтервал між спостереженнями. Очевидно, чим більший цей інтервал, тим менш правдоподібно, що при переході від одного спостереження до іншого характер впливу неврахованих змінних буде зберігатися.

## 6.4. Наслідки автокореляції

Якщо параметри моделі оцінені за класичним МНК і у залишків існує автокореляція, то мають місце такі погані наслідки [11]:

1. *Оцінки параметрів* лінійної моделі $b_j$ , знайдені за МНК, будуть лінійними, незміщеними, обгрунтованими, але *перестають бути ефективними* (тобто вибіркові дисперсії вектора оцінок $B$ можуть бути невиправдано великими). Оцінка $B$ не буде найкращою. Неефективність оцінок параметрів економетричної моделі, як правило, призводить до неефективних прогнозів, тобто прогнозні значення матимуть велику вибіркову дисперсію.

2. *Дисперсії оцінок параметрів є зміщеними.* Переважно дисперсії є заниженими, що спричиняє збільшення фактичних $t$-статистик Ст'юдента. Це також може призвести до визнання статистично значущими тих факторних ознак, які насправді не мають такої властивості.

3. *Оцінка непоясненої дисперсії* за кореляційно-регресійною моделлю з $k$ факторами

$$\boldsymbol{S^2 = \frac{\sum_{t=1}^{T} \varepsilon_t^2}{T - k - 1}}$$

є *зміщеною оцінкою істинного значення дисперсії* $\sigma^2$ (переважно занижує істинне значення).

4. *Статистичні критерії*, які отримані для класичної лінійної моделі, *не можуть бути використані для дисперсійного аналізу*, бо їх розрахунок не враховує наявності коваріації залишків.

Висновки із $t$- і $F$-статистик про значущість коефіцієнтів регресії і коефіцієнта детермінації можуть бути неправильними. Унаслідок цього прогнозні якості побудованої кореляційно-регресійної моделі погіршуються.

Таким чином, наслідки автокореляції деякою мірою збігаються з наслідками гетероскедастичності.

## 6.5. Тестування автокореляції

Через те, що під час побудови кореляційно-регресійної моделі за допомогою МНК істинні значення випадкових величин невідомі, то висновки про їх незалежність роблять на підставі оцінок залишків $e_i$, що отримують із вибіркової кореляційно-регресійної моделі.

Розглянемо деякі методи тестування автокореляції [11],[14].

### 6.5.1. Графічний метод

Використовують декілька варіантів графічного визначення автокореляції. Один із них передбачає побудову так званих послідовно-часових графіків, які на площині представляють відхилення залишків $e_t$ в моменти часу $t$ (рис. 6.1) У цьому разі по осі абсцис відкладають або моменти часу отримання статистичних даних, або порядкові номери спостережень, а по осі ординат – відхилення залишків $e_{t-1.}$ На рис. 6.1 а) – г) зображені деякі зв’язки між відхиленнями залишків, тобто між ними має місце наявна автокореляція. Відсутність систематичності зміни відхилень на рис. 6.1 д) свідчить про відсутність автокореляції.

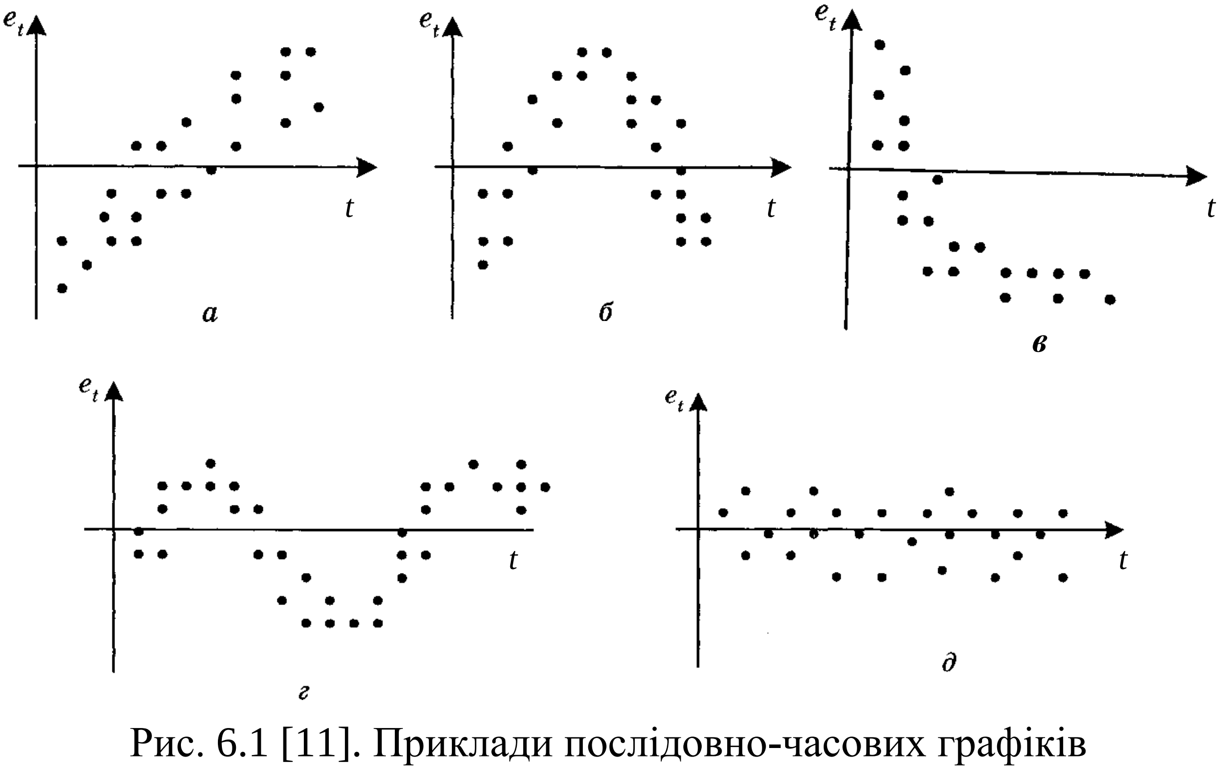



Рис. 6.1 [11]. Приклади послідовно-часових графіків

На рис. 6.1 б) відхилення спочатку здебільшого від’ємні, потім додатні, потім знову від’ємні. Це свідчить про наявність деякої залежності між відхиленнями. У цьому прикладі наявна додатна автокореляція випадкових відхилень. Вона стає більш наочною, якщо рисунок 6.1 б) доповнити графіком залежності $e_t$ від $e_{t-1}$, який у цьому разі орієнтовно матиме вигляд, як на рисунку 6.2. Здебільшого точки на цьому графіку розташовані в I і III четвертях декартової системи координат, що підтверджує додатну залежність між сусідніми значення випадкових відхилень.

В сучасних економетричних пакетах аналітичний вираз регресійної моделі доповнюється графічним представленням результатів. Зіставивши графіки кореляційно-регресійної моделі та випадкових відхилень залишків, можна висунути гіпотезу про наявність або відсутність автокореляції випадкових величин. Якщо ці графіки перетинаються рідко, то можна припустити наявність додатної автокореляції.

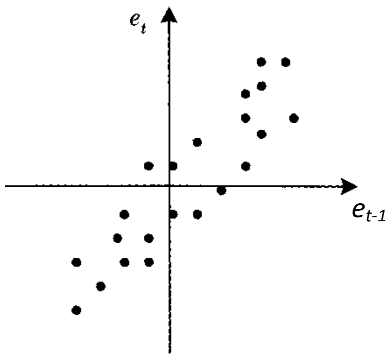



Рис. 6.2 [11]. Залежність поточних значень $e_t$ випадкових відхилень від попередніх $e_{t-1}$

### 6.5.2. Метод рядів

Цей метод достатньо простий для застосування. Потрібно послідовно визначити знаки випадкових відхилень $e_t$.

*Рядом* називають неперервну послідовність однакових знаків значень випадкових відхилень. Кількість знаків у ряді називають *довжиною ряду*.

Візуальний аналіз розподілу знаків свідчить про невипадковий характер зв'язку між випадковими відхиленнями. Якщо кількість рядів невелика порівняно з кількістю спостережень *T,* то ймовірна додатна автокореляція. Якщо рядів дуже багато, то ймовірна наявність від'ємної автокореляції.

Для докладнішого аналізу рядів використовують таку процедуру. Нехай $T$ - обсяг вибірки; $n_1$ - загальна кількість знаків " +" ("плюс") при $T$ спостереженнях (кількість додатних відхилень $e_t$); $n_2$ - загальна кількість знаків ("мінус") при $T$ спостереженнях (кількість від'ємних відхилень $e_t$); $l$ - кількість рядів.

При достатньо великій кількості спостережень ($T \geq 30$) і відсутності автокореляції величина $n_3$ має асимптотично нормальний розподіл з математичним сподіванням

$$M(l) = \frac{2n_1n_2}{n_1 + n_2} + 1 \qquad (6.5)$$

та дисперсією

$$D(l) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}\,. \qquad (6.6)$$

Для того, щоб протестувати наявність автокореляції, треба:

1) сформулювати нульову гіпотезу $H_0$: *автокореляції між випадковими відхиленнями немає*;

2) перевірити, чи розраховане значення $n_3$ має властивості (6.5) та (6.6), тобто належить до інтервалу

$$M(l) - u_{\alpha/2}^{\text{табл}}\sqrt{D(l)} < l < M(l) + u_{\alpha/2}^{\text{табл}}\sqrt{D(l)}\ , \qquad (6.7)$$

де $u_{\alpha/2}^{\text{табл}}$ - табличне значення нормального розподілу (таблиця 7 Додатку), що визначається з умови $\Phi\left(u_{\alpha/2}^{\text{табл}}\right) = (1-\alpha)/2$ ($\Phi(u)$ - функція Муавра - Лапласа) при заданому рівні значущості α.

3) при виконанні умови (6.7) нульова гіпотеза про відсутність автокореляції приймається, тобто з імовірністю $p = 1 - \alpha$ можна стверджувати, що автокореляції випадкових відхилень немає.

Якщо умова (6.7) не виконується, то нульова гіпотеза про відсутність автокореляції відхиляється, тобто з імовірністю $p = 1 - \alpha$ можна стверджувати, що наявна автокореляція випадкових відхилень.

*Тест Сведа − Ейзенхарта*

Для того, щоб протестувати автокореляцію при невеликій кількості спостережень, перевіряють дві нульові гіпотези:

$\mathrm{H}_0^+$- наявна додатна автокореляція;

$\mathrm{H}_0^-$- наявна від'ємна автокореляція.

Для перевірки цих нульових гіпотез, Свед і Ейзенхарт розробили таблиці критичних значень кількості рядів при *T* спостереженнях (таблиця 4 Додатку). Згідно з тестом Сведа − Ейзенхарта, потрібно визначити відповідно нижнє і верхнє значення $k_1^{\text{табл}}$ та $k_2^{\text{табл}}$, якщо ступені вільності $v_1 = n_1$ і $v_2 = n_2$ та рівень значущості α.

Якщо $k_1^{\text{табл}} < n_3 < k_2^{\text{табл}}$, то з ймовірністю $p = 1 - \alpha$ можна стверджувати, що автокореляція відсутня.

Якщо $n_3 < k_1^{\text{табл}}$, то з ймовірністю $p = 1 - \alpha$ приймаємо нульову гіпотезу $\mathrm{H}_0^+$, тобто наявна додатна автокореляція випадкових відхилень. Якщо $k_2^{\text{табл}} < n_3$, то з ймовірністю $p = 1 - \alpha$ приймається нульова гіпотеза $\mathrm{H}_0^-$, тобто від'ємна автокореляція випадкових відхилень наявна.

### 6.5.3. Критерій Дарбіна−Уотсона (*DW*-критерій)

Найвідомішим критерієм тестування автокореляції першого порядку є критерій Дарбіна – Уотсона [16]. Щоб перевірити умови про статистичну незалежність випадкових відхилень $e_t$ між собою, знаходять коефіцієнт кореляції між *сусідніми* значеннями $e_t$. Тест складається з декількох кроків і включає зони невизначеності.

*Крок* 1. Розраховується значення *d*-статистики за формулою

$$DW = \sum_{t=2}^{n}(e_t - e_{t-1})^2 \Big/ \sum_{t=1}^{n} e_t^2. \qquad (6.8)$$

Доведено, що значення $d$-статистики перебуває у межах $0 \leq DW \leq 4$.

Щоб визначити точніше, яке значення $DW$ свідчить про відсутність атокореляції, а яке про її наявність, побудована таблиця критичних точок розподілу Дарбіна - Уотсона (таблиця 5 Додатку).

*Крок* 2. Задається рівень значущості $\alpha$. За табл. 5 Додатку при заданому рівні значущості α, кількості факторів $k$ і кількості спостережень $n$ знаходять два значення $DW_1$ і $DW_2$ (в таблицях їх позначають відповідно $DW_L$ і $DW_U$). Можливі такі випадки (рис. 6.3):

- при $0 < DW < DW_1$ має місце наявна додатна автокореляція;
- при $DW_1 \leq DW \leq DW_2$ або $4 - DW_2 \leq DW \leq 4 - DW_1$ не можна зробити висновки ані про наявність, ані про відсутність автокореляції ($DW$ потрапляє в зону невизначеності);
- при $4 - DW_1 \leq DW \leq 4$ має місце наявна від'ємна автокореляція;
- при $DW_2 \leq DW \leq 4 - DW_2$ автокореляція відсутня.

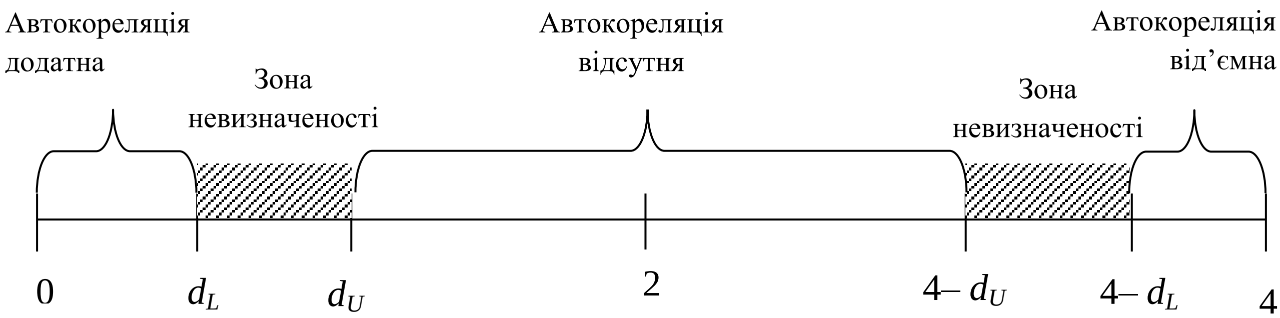



Рис. 6.3. Зони авто кореляційного зв'язку за $DW$-критерієм

*Умови застосування критерію Дарбіна -Уотсона*

Під час використання $DW$-критерію потрібно враховувати умови:

1. Критерій $DW$ застосовують лише для тих моделей, які містять вільний член кореляційно-регресійної моделі.
2. Припускається, що випадкові величини залишків $\varepsilon_t$ визначають за такою ітераційною схемою:

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t\,, \qquad -1 < \rho < 1, \tag{6.9}$$

де $\rho$ - коефіцієнт автокореляції; $v_t$ - випадкова величина, для якої виконуються всі припущення класичного регресійного аналізу.

Модель (6.9) називають *авторегресійною моделлю Маркова першого порядку AR*(1).Вона описує *автокореляцію першого порядку.*

3. Статистичні дані повинні мати однакову періодичність, тобто не має бути пропусків у спостереженнях.
4. $DW$-критерій не застосовують для кореляційно-регресійних моделей, що містять у переліку екзогенних змінних лагове значення результуючої змінної, тобто для таких авторегресійних моделей:

$$y_t = \beta_0 + \beta_1 x_{t_1} + \beta_2 x_{t_2} + \cdots + \beta_k x_{t_k} + \lambda\, y_{t-1} + \varepsilon_t \quad (t = \overline{1, \mathrm{T}}).$$

5. Довжина вибірки не повинна бути менше 15.

*Примітка*. Можна показати, що величина статистики $DW$ наближається до $2(1-\rho)$, тобто

$$DW \approx 2(1-\rho)$$

Зрозумілий і змістовний сенс статистики $DW$: якщо між залишками $e_t$ та $e_{t\text{-}1}$ є достатньо висока додатна кореляція, то в певному сенсі $e_t$ та $e_{t\text{-}1}$ наближені один до одного і величина статистики $DW$ мала. Це погоджується з наведеною наближеною формулою: якщо коефіцієнт автокореляції першого порядку $\rho$ наближається до одиниці, то величина статистики $DW$ наближається до нуля. Відсутність кореляції означає, що статистика $DW$ наближається до 2.

### 6.5.4. Критерій фон Неймана

Для виявлення автокореляції використовується формула:

$$Q = Q_{факт} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}} \cdot \frac{n}{n-1} \qquad (6.10)$$

Звідси $Q = \frac{n}{n-1} DW$. Якщо число спостережень $n \to \infty,$ то фактичне значення $Q \to DW$. При $Q_{факт} < Q_{табл}$ маємо існування автокореляції.

## 6.6. Нециклічний коефіцієнт автокореляції

Нециклічний коефіцієнт автокореляції відображає ступінь взаємозв'язку залишків кожного наступного значення з попереднім:

I ряд – $e_1, e_2, e_3, \ldots, e_{n\text{-}1}$;

II ряд – $e_2, e_3, e_4, \ldots, e_n$.

Він обчислюється за формулою:

$$r^* = \frac{\sum_{t=1}^{n}(e_t e_{t-1}) - \frac{1}{n-1}\sum_{t=1}^{n} e_t \sum_{t=2}^{n} e_{t-1}}{\sqrt{\left[\sum_{t=1}^{n} e_t^2 - \frac{1}{n-1}\left(\sum_{t=1}^{n} e_t\right)^2\right]\left[\sum_{t=2}^{n} e_{t-1}^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n} e_{t-1}\right)^2\right]}}$$

Коефіцієнт $r^*$ може набувати значень в інтервалі (-1; +1). Його від'ємні значення свідчать про від'ємну автокореляцію, додатні – про додатну. Значення, що містяться в деякій критичній області біля нуля, свідчать про відсутність автокореляції, тобто стверджують нульову гіпотезу про відсутність автокореляції залишків. Оскільки ймовірнісний розподіл $r^*$ встановити важко, то на практиці замість $r^*$ обчислюють циклічний коефіцієнт автокореляції $r°$.

## 6.7. Циклічний коефіцієнт автокореляції

Циклічний коефіцієнт автокореляції відображає ступінь взаємозв'язку рядів [17]:

I ряд – $e_1, e_2, e_3, \ldots, e_{n-1}, e_n$;

II ряд – $e_2, e_3, e_4, \ldots, e_n, e_1$.

Циклічний коефіцієнт обчислюється за формулою:

$$r^0 = \frac{\sum_{t=1}^{n}(e_t e_{t-1}) + e_n e_1 - \frac{1}{n}\left(\sum_{t=1}^{n} e_t\right)^2}{\sum_{t=1}^{n} {e_t}^2 - \frac{1}{n}\left(\sum_{t=1}^{n} e_t\right)^2}.$$

Для досить довгих рядів вплив циклічних членів на величину коефіцієнта $r°$ незначний, тому можна вважати, що ймовірнісний розподіл $r^*$ наближається до розподілу $r$. Якщо останній член ряду дорівнює першому, тобто $e_1 = e_n$, то нециклічний коефіцієнт автокореляції дорівнює циклічному.

Очевидно, що коли залишки не містять тренду, то припущення про рівність $e_1 = e_n$ недалеке від реальності і циклічний коефіцієнт автокореляції наближається до нециклічного.

Фактичне значення циклічного коефіцієнта автокореляції порівнюється з табличним $DW$-критерієм для рівня значущості $\alpha$ і довжини ряду $n$. Якщо $r_{\text{факт}} > r_{\text{табл}}$, то існує автокореляція. При $\sum_{t=1}^{n} e_t \approx \sum_{t=2}^{n} e_{t-1} \approx 0$ циклічний коефіцієнт автокореляції можна обчислити за формулою:

$$r^0 = \frac{\sum_{t=2}^{n}(e_t e_{t-1})}{\sum_{t=1}^{n} {e_t}^2} \ . \tag{6.11}$$

На практиці замість (6.11) використовують таку формулу:

$$r^0 = \frac{\frac{1}{n-1}\sum_{t=2}^{n}(e_t e_{t-1})}{\frac{1}{n}\sum_{t=1}^{n} {e_t}^2}.$$

## 6.8. Методи оцінки параметрів моделі з автокореляцією

Для оцінювання параметрів моделі з автокорельованими залишками можна застосовувати такі методи [17],[18]:

1) Ейткена;

2) перетворення вхідної інформації;

3) Кочрена - Оркатта;

4) Дарбіна.

Перші два методи доцільно застосовувати, коли залишки описуються авторегресійною моделлю першого порядку. Ітераційні методи Кочрена - Оркатта і Дарбіна можна застосовувати для оцінки параметрів економетричної моделі й тоді, коли залишки описуються авторегресійною моделлю вищого порядку.

### 6.8.1. Метод Ейткена

Розглядається економетрична модель (6.3):

$$y_t = \beta_0 + \beta_1 x_t \quad + \varepsilon_t \quad (t = \overline{1, \mathrm{T}}).$$

Припустимо, що між випадковими величинами залишків є лінійна залежність (6.4):

$$\varepsilon_t = \rho\, \varepsilon_{t\text{-}1} + u_t \qquad (t = \overline{1, \mathrm{T}})\ .$$

Для усунення автокореляції залишків $\varepsilon_t$ треба перетворити основну модель так, щоб вона мала залишки $u_t$. Оскільки $u_t = \varepsilon_t - \rho \cdot \varepsilon_{t\text{-}1}$, то для перетворення треба записати модель для попереднього періоду:

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

З рівняння (6.3) віднімемо останнє рівняння, яке помножено на $\rho$. Отримаємо таку економетричну модель:

$$y_t \ -\rho\, y_{t\text{-}1} = \beta_0\ \ (1-\rho) + \beta_1\, (x_t - \rho\, x_{t\text{-}1}) + (\ \varepsilon_{\,t} - \rho\ \varepsilon_{\,t\text{-}1}).$$

Після того, як вихідні дані перетворені, тобто $y_t - \rho\ y_{t\text{-}1}$, $x_t - \rho\ x_{t\text{-}1}$, то для оцінювання параметрів можна застосувати МНК. Для перетворення можна використати перші різниці $y_t - y_{t\text{-}1}$, $x_t - x_{t\text{-}1}$, коли коефіцієнт автокореляції $\rho$ наближається до одиниці. Якщо $\rho = 1$, то у перетвореній моделі відсутній вільний член (як виняток може бути ситуація, коли вихідна модель містить часовий тренд).

Параметр $\rho$ характеризує силу зв'язку величин залишків у період $t$ з величинами залишків у період $t - 1$. Його значення можна знайти на основі залишків, якщо обчислити циклічний коефіцієнт кореляції $r^0$. На практиці для розрахунку коефіцієнта автокореляції $\rho$ використовується одне з таких співвідношень:

$$\rho \approx r^0 \approx \frac{\sum_{t=2}^{n}(e_t e_{t-1})}{\frac{1}{n}\sum_{t=1}^{n} {e_t}^2},$$

$$\rho \approx \frac{\sum_{t=2}^{n}(e_t e_{t-1})}{\sum_{t=1}^{n} {e_t}^2} \cdot \frac{n}{n-1} + \frac{k+1}{n}. \qquad (6.12)$$

Тут $n$ – число спостережень, $k$ – число факторів.

Для оцінювання параметрів економетричної моделі, що має автокореляцію залишків, можна застосувати метод Ейткена, який базується на скоригованій вихідній інформації з урахуванням коваріації залишків.

На основі методу Ейткена маємо:

$$B = (X^T V^{-1} X)^{-1} X^T V^{-1} Y,$$

або

$$B = (X^T S^{-1} X)^{-1} X^T S^{-1} Y,$$

де $B$ – вектор оцінок параметрів економетричної моделі;

$X$ – матриця незалежних змінних;

$X^T$ – матриця, транспонована до матриці $X$;

$S^{-1}$– матриця, обернена до матриці кореляції залишків;

$V^{-1}$ – матриця, обернена до матриці $V$;

$V = \sigma_\varepsilon^2 \cdot S$, де $\sigma_\varepsilon^2$ – залишкова дисперсія;

$Y$ – вектор залежних змінних.

Матриця коваріації залишків $S$ має вигляд:

$$S = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & 1 \end{pmatrix}$$

Матриця $S$ – симетрична, елемент матриці $\rho^s$ виражає коефіцієнт автокореляції $s$-го порядку для залишків $\varepsilon_t$. Коефіцієнт автокореляції нульового порядку дорівнює 1.

Оскільки коваріація залишків $\rho^s$ при $s > 2$ часто наближається до нуля, то обернену матрицю коваріації залишків доцільно використовувати у вигляді:

$$S^{-1} = \frac{1}{1-\rho^2} \cdot \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

### 6.8.2. Метод перетворення вхідної інформації

Метод використовується при авторегресійній моделі першого порядку і дає оцінку параметрів моделі за допомогою двокрокової процедури.

*Крок*1. Перетворення вхідної інформації за допомогою матриці $T$.

*Крок* 2. Застосування класичного МНК для оцінки перетворених даних. Для цього треба знайти матрицю перетворення $T$, щоб модель

$$TY = TXB + Te,$$

мала скалярну дисперсійну матрицю

$$M(T \cdot e \cdot e^T \cdot T^T) = \sigma_e^2 \cdot E.$$

Перетворення вхідної інформації виконується за допомогою такої нижньотрикутної матриці $n$-го порядку:

$$T_1 = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Іноді для перетворення вхідної інформації використовується матриця $T_2$ розміром $(n\text{-}1)\times n$, яка отримується з матриці $T_1$ викреслюванням першого рядка:

$$T_2 = \begin{pmatrix} -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Застосування МНК до даних $T_1Y$ і $T_1X$ дає таку саму оцінку параметрів моделі, як і метод Ейткена, а для даних $T_2Y$ і $T_2X$ воно забезпечує порівняно добру апроксимацію.

У загальному випадку, коли немає інформації ні про порядок авторегресійної моделі, ні про значення параметрів у ній (через те неможливо застосувати ні метод Ейткена, ні метод перетворення вихідної інформації), використовують наближені методи Кочрена - Оркатта і Дарбіна.

### 6.8.3. Метод Кочрена - Оркатта

Метод використовується тоді, коли залишки задовольняють автокореляційній моделі вищого порядку. Він є ітеративним методом оцінювання параметрів економетричної моделі, коли мінімізується сума квадратів залишків [17].

Розглядається економетрична модель (6.3)

$$y_t = \beta_0 + \beta_1 x_t \; + \varepsilon_t \quad (t = \overline{1, \mathrm{T}}).$$

Припустимо, що між випадковими величинами залишків є лінійна залежність (6.4):

$$\varepsilon_t = \rho\, \varepsilon_{t\text{-}1} + u_t \qquad\qquad (t = \overline{1, \mathrm{T}}).$$

Сума квадратів залишків визначається формулою:

$$\sum_{t=1}^{n} u_t^2 = \sum_{t=2}^{n} [(y_t - \rho y_{t-1}) - \beta_0 (1-\rho) - \beta_1 (x_t - \rho x_{t-1})\,]^2 . \quad (6.13)$$

Безпосередня мінімізація функції (6.13) приводить до системи нелінійних рівнянь, тому аналітичний вираз оцінок параметрів $b_0$, $b_1$ і коефіцієнта автокореляції $\rho$ дістати важко. Використовується така ітераційна процедура послідовних наближень.

*Крок* 1. Довільно вибирається значення коефіцієнта автокореляції $\rho$, наприклад, $\rho = r_1^0$. Підставивши його значення в формулу (6.13), обчислюють $b_0^{(1)}$ і $b_1^{(1)}$.

*Крок* 2. Поклавши $b_0 = b_0^{(1)}$ і $b_1 = b_1^{(1)}$, підставляють їх у формулу (6.13) і знаходять коефіцієнт $\rho = r_2^0$.

*Крок 3*. Підставивши у формулу (6.13) значення $\rho = r_2^0$, знайдемо $b_0 = b_0^{(2)}$ і $b_1 = b_1^{(2)}$.

*Крок 4*. Значення $b_0^{(2)}$ і $b_1^{(2)}$ використовуються для мінімізації суми квадратів залишків (6.13) за невідомим коефіцієнтом $\rho = r_3^0$. Процедура триває доти, доки наступні значення параметрів $b_0$, $b_1$ і коефіцієнта $\rho$ не будуть відрізнятись менш, чим на задану величину.

Як і інші подібні процедури, цей ітеративний метод має дві проблеми: збіжності та характеру знайденого мінімуму – локальний чи глобальний. Проведені дослідження за цими двома проблемами показали, що в результаті застосування методу Кочрена – Оркатта завжди знаходимо глобальний оптимум і алгоритм забезпечує порівняно добру збіжність.

### 6.8.4. Метод Дарбіна

Метод Дарбіна використовується у випадку авторегресійної моделі вищого порядку. Дарбін запропонував просту двокрокову процедуру, що дозволяє одержати статистичні оцінки, яким будуть притаманні ті ж властивості, як і оцінкам, одержаним за МНК [18].

*Крок* 1. Підставимо значення залишків (6.4), які підпорядковані авторегресійній моделі першого порядку

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

у рівняння економетричної моделі (6.3):

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (t = \overline{1, T}).$$

В результаті отримаємо

$$y_t = \beta_0 + \beta_1 x_t + \rho\, \varepsilon_{t-1} + u_t = \beta_0 + \beta_1 x_t + \rho\, (y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + u_t =$$
$$= \beta_0 (1 - \rho) + \beta_1 x_t + \rho\, y_{t-1} - \beta_1 \rho\, x_{t-1} + u_t.$$

Отже,

$$y_t = \beta_0 (1 - \rho) + \beta_1 x_t + \rho\, y_{t-1} - \beta_1 \rho\, x_{t-1} + u_t.$$

де залишок $u_t$ має скалярну матрицю дисперсій.

Згідно з МНК, визначаються параметри цієї моделі і коефіцієнт $\rho$. У результаті обчислень отримуємо $\rho = r'$.

*Крок* 2. Значення $\rho = r'$ використовується для перетворення змінних $(y_t - r'\, y_{t-1})$ і $(x_t - r'\, x_{t-1})$, а МНК застосовується до перетворених даних. Коефіцієнт при $(x_t - r'\, x_{t-1})$ є оцінкою параметра $\beta_1$, а вільний член, поділений на $(1 - r')$, оцінює параметр $\beta_0$.

Метод Дарбіна можна поширити на випадок кількох незалежних змінних і для автокореляції вищих порядків.

Метод Кочрена - Оркатта і розглянута процедура Дарбіна за наявності автокореляції залишків асимптотично ефективніші, ніж МНК. Але при цьому постають два важливі запитання:

1) чи будуть ці методи ефективнішими, ніж МНК, для малих вибіркових сукупностей?
2) якою – однаковою чи різною – буде ефективність застосування вказаних методів для малих вибірок?

Чисельний аналіз, виконаний американським економістом Ц.Гриліхесом (1930–1999) та індійським математиком К.Р.Рао (1920) за допомогою методу Монте-Карло, дав відповідь на ці запитання:

- МНК дає менш ефективні оцінки порівняно з іншими методами, якщо сукупність спостережень $n$ = 20 одиниць, а $|\rho| > 0{,}3$.
- Якщо $|\rho| < 0{,}3$, то зниження ефективності оцінок за МНК порівняно зі складнішими процедурами невелике.
- Метод Дарбіна забезпечує найкращу оцінку для більш широкого кола параметрів порівняно з іншими методами.
- Нелінійний метод оцінювання параметрів не дає відчутних переваг порівняно з двокроковою процедурою Дарбіна.

## 6.9. Приклад оцінювання параметрів моделі з автокорельованими залишками

*Постановка задачі*. На основі статистичних даних таблиці 6.1 провести такі дослідження:

1) За допомогою МНК побудувати кореляційно-регресійну модель, яка описує залежність рентабельності підприємства ($y_t$, %) від виробництва товару ($x_t$, млн. од.) за 1996 – 2015 роки;

2) Провести тестування наявності автокореляції випадкових відхилень цієї моделі, користуючись такими методами:

а) графічним; б) методом рядів; в) тестом Сведа – Ейзенхарта;

г) критерієм Дарбіна – Уотсона; д) критерієм Неймана.

3) Оцінити параметри моделі методом перетворення вхідної інформації та узагальненим методом найменших квадратів.

***Розв'язання.*** 1) За допомогою МНК на підставі даних табл. 6.1 побудуємо парну лінійну кореляційно-регресійну модель

$$y = \beta_0 + \beta_1 x + \varepsilon.$$

За допомогою функції ЛИНЕЙН одержимо такі значення:

| | |
|---|---|
| **0,558667** | **5,448113** |
| 0,029157 | 2,746334 |
| 0,953263 | 3,606009 |
| 367,1304 | 18 |
| 4773,909 | 234,0595 |

Вибіркова кореляційно-регресійна модель має такий вигляд:

$$\hat{y} = 5{,}4481 + 0{,}5587\,x. \tag{6.14}$$

Таблиця 6.1

| Роки | $t$ | $x_t$ | $y_t$ |
|---|---|---|---|
| 1996 | 1 | 44,80 | 37,86 |
| 1997 | 2 | 51,20 | 38,91 |
| 1998 | 3 | 52,40 | 38,91 |
| 1999 | 4 | 66,80 | 43,39 |
| 2000 | 5 | 84,40 | 50,23 |
| 2001 | 6 | 81,20 | 48,12 |
| 2002 | 7 | 86,00 | 51,01 |
| 2003 | 8 | 82,00 | 49,01 |
| 2004 | 9 | 73,20 | 45,34 |
| 2005 | 10 | 85,20 | 49,89 |
| 2006 | 11 | 85,20 | 50,79 |
| 2007 | 12 | 85,60 | 50,30 |
| 2008 | 13 | 81,20 | 49,06 |
| 2009 | 14 | 86,40 | 51,08 |
| 2010 | 15 | 104,40 | 61,20 |
| 2011 | 16 | 110,00 | 64,51 |
| 2012 | 17 | 121,60 | 72,23 |
| 2013 | 18 | 130,00 | 79,45 |
| 2014 | 19 | 144,40 | 91,52 |
| 2015 | 20 | 144,80 | 92,20 |

2) *Тестування моделі на наявність автокореляції випадкових відхилень.*

а) Тестування автокореляції за *графічним методом*.

На підставі парної лінійної кореляційно-регресійної моделі (6.14) розрахуємо значення випадкових відхилень $e_t$ теоретичних значень рентабельності підприємства від відповідних фактичних значень (табл. 6.2).

Таблиця 6.2

| $t$ | $y_t$ | $\hat{y}_t$ | $e_t$ | $e_{t-1}$ |
|---|---|---|---|---|
| 1 | 37,86 | 30,4764 | 7,3836 | |
| 2 | 38,91 | 34,0519 | 4,8581 | 7,3836 |
| 3 | 38,91 | 34,7223 | 4,1877 | 4,8581 |
| 4 | 43,39 | 42,7671 | 0,6229 | 4,1877 |
| 5 | 50,23 | 52,5996 | -2,3696 | 0,6229 |
| 6 | 48,12 | 50,8119 | -2,6919 | -2,3696 |
| 7 | 51,01 | 53,4935 | -2,4835 | -2,6919 |
| 8 | 49,01 | 51,2588 | -2,2488 | -2,4835 |
| 9 | 45,34 | 46,3425 | -1,0025 | -2,2488 |
| 10 | 49,89 | 53,0466 | -3,1566 | -1,0025 |
| 11 | 50,79 | 53,0466 | -2,2566 | -3,1566 |
| 12 | 50,30 | 53,2700 | -2,9700 | -2,2566 |
| 13 | 49,06 | 50,8119 | -1,7519 | -2,9700 |
| 14 | 51,08 | 53,7170 | -2,6370 | -1,7519 |
| 15 | 61,20 | 63,7730 | -2,5730 | -2,6370 |
| 16 | 64,51 | 66,9015 | -2,3915 | -2,5730 |
| 17 | 72,23 | 73,3820 | -1,1520 | -2,3915 |
| 18 | 79,45 | 78,0748 | 1,3752 | -1,1520 |
| 19 | 91,52 | 86,1196 | 5,4004 | 1,3752 |
| 20 | 92,20 | 86,3431 | 5,8569 | 5,4004 |
| Сума | 1115,01 | 1115,01 | 0,0000 | -5,8569 |

Зобразимо послідовно-часовий графік, який подає випадкові відхилення $e_t$ в моменти часу $t$. На рис. 6.4 можна побачити, що має місце 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 залишків для вибірки.

На рис. 6.5 зображений графік залежності випадкових відхилень $e_t$ від $e_{t-1}$. Здебільшого точки на цьому графіку розташовані в I та III чвертях, що підтверджує наявність додатної залежності між суміжними значеннями випадкових величин.

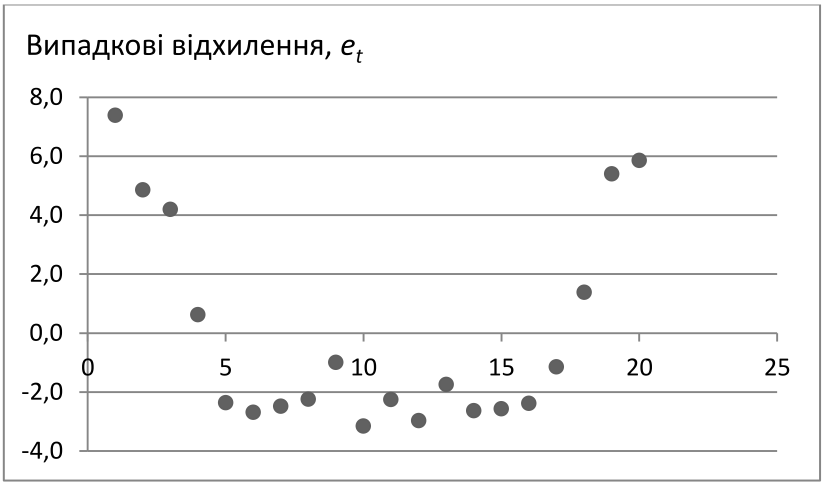



Рис. 6.4. Послідовно-часовий графік випадкових відхилень

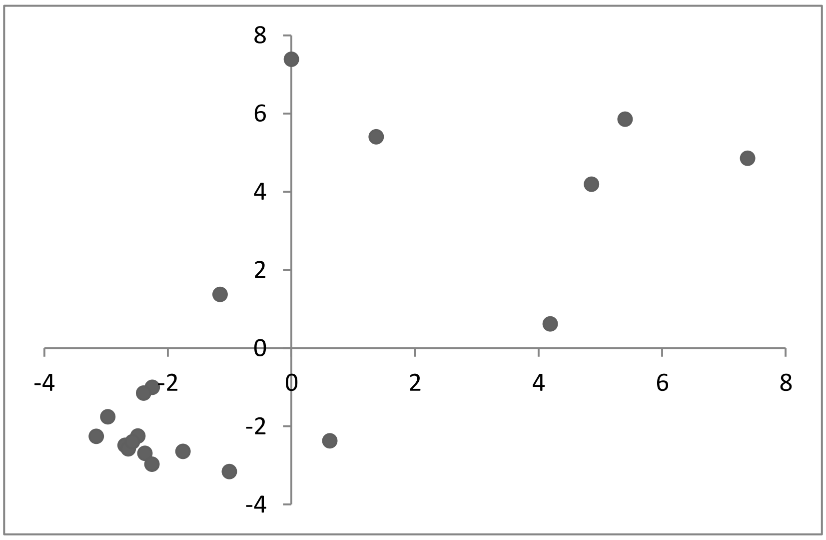



Рис. 6.5. Залежність поточних значень $e_t$ випадкових відхилень від попередніх $e_{t-1}$

б). Тестування автокореляції за *методом рядів*.

Подамо послідовність знаків випадкових відхилень залишків $e_t = y_t - \hat{y}_t$ (табл. 6.2), обчислених на підставі вибіркової кореляційно-регресійної моделі (6.14):

$$+ + + + - - - - - - - - - - - - - \;\; + + +.$$

Спостерігається послідовність із 4 "+" ("плюсів"), 13 ("мінусів") та 3 "+" ("плюсів") при 20 спостереженнях. У цьому разі є 3 ряди довжиною 4, 13 та 3 елементи відповідно. Кількість додатних відхилень $n_1 = 7$, кількість відємних відхилень $n_2 = 13$. Оскільки кількість рядів $l = 3$ є невеликою порівняно з кількістю спостережень $T = 20$, то наявна додатна автокореляція випадкових відхилень.

Перевіримо наявність автокореляції методом побудови інтервалу довіри для величини $l$. Математичне сподівання та дисперсія величини $n_3$ *відповідно дорівнюють*

$$\mathrm{M(l)} = \frac{2\mathrm{n}_1\mathrm{n}_2}{\mathrm{n}_1 + \mathrm{n}_2} + 1 = \frac{2 \cdot 7 \cdot 13}{7 + 13} + 1 = 10{,}1;$$

$$\mathrm{D(l)} = \frac{2\mathrm{n}_1\mathrm{n}_2 \cdot (2\mathrm{n}_1\mathrm{n}_2 - \mathrm{n}_1 - \mathrm{n}_2)}{(\mathrm{n}_1 + \mathrm{n}_2)^2 \cdot (\mathrm{n}_1 + \mathrm{n}_2 - 1)} = \frac{182 \cdot 162}{20^2 \cdot 19} = 3{,}8795.$$

При заданому рівні значущості $\alpha = 0{,}05$ з умови

$$\Phi\left(\boldsymbol{u}_{\boldsymbol{\alpha}/2}^{\text{табл}}\right) = (\mathbf{1} - \boldsymbol{\alpha})/\mathbf{2} = \mathbf{0{,}475}$$

за табл. 7 Додатку знаходимо $\boldsymbol{u}_{\mathbf{0{,}025}}^{\text{табл}} = 1{,}96$.

Довірчий інтервал має вигляд:

$$[10{,}1 - 1{,}96\sqrt{3{,}8795};\ 10{,}1 + 1{,}96\sqrt{3{,}8795}] = [6{,}2396;\ 13{,}9604].$$

Оскільки значення кількості рядів $l = 3$ не входить у довірчий інтервал, то з ймовірністю 0,95 можна стверджувати наявність автокореляції випадкових відхилень.

в) Тестування автокореляції за допомогою *тесту Сведа - Ейзенхарта.*

Для перевірки цих нульових гіпотез, визначимо відповідно нижнє і верхнє табличні значення $k_1^{\text{табл}} = 5$ та $k_2^{\text{табл}} = 13$, якщо кількість ступенів вільності $v_1 = n_1 = 7$, $v_2 = n_2 = 13$, рівень значущості $\alpha = 0{,}05$.

Оскільки $l < k_1^{\text{табл}}$, то з імовірністю 0,95 приймається нульова гіпотеза $\mathrm{H}_0^+$ про наявність додатної автокореляції випадкових величин.

г) Тестування автокореляції за *критерієм Дарбіна –Уотсона*.

Запишемо необхідні розрахунки для знаходження фактичного значення *DW*-критерію в табл. 6.3.

Таблиця 6.3

| $t$ | $e_t$ | $e_{t-1}$ | $e_t^2$ | $(e_t - e_{t-1})^2$ |
|---|---|---|---|---|
| 1 | 7,3836 | 0,0000 | 54,5176 | 0,0000 |
| 2 | 4,8581 | 7,3836 | 23,6014 | 6,3780 |
| 3 | 4,1877 | 4,8581 | 17,5371 | 0,4494 |
| 4 | 0,6229 | 4,1877 | 0,3880 | 12,7078 |
| 5 | -2,3696 | 0,6229 | 5,6151 | 8,9553 |
| 6 | -2,6919 | -2,3696 | 7,2462 | 0,1039 |
| 7 | -2,4835 | -2,6919 | 6,1677 | 0,0434 |
| 8 | -2,2488 | -2,4835 | 5,0572 | 0,0551 |
| 9 | -1,0025 | -2,2488 | 1,0051 | 1,5532 |
| 10 | -3,1566 | -1,0025 | 9,9638 | 4,6397 |
| 11 | -2,2566 | -3,1566 | 5,0920 | 0,8100 |
| 12 | -2,9700 | -2,2566 | 8,8210 | 0,5090 |
| 13 | -1,7519 | -2,9700 | 3,0691 | 1,4839 |
| 14 | -2,6370 | -1,7519 | 6,9535 | 0,7833 |
| 15 | -2,5730 | -2,6370 | 6,6201 | 0,0041 |
| 16 | -2,3915 | -2,5730 | 5,7193 | 0,0329 |
| 17 | -1,1520 | -2,3915 | 1,3272 | 1,5363 |
| 18 | 1,3752 | -1,1520 | 1,8911 | 6,3867 |
| 19 | 5,4004 | 1,3752 | 29,1638 | 16,2022 |
| 20 | 5,8569 | 5,4004 | 34,3031 | 0,2084 |
| Сума | 0,0000 | -5,857 | **234,0595** | **62,8427** |

За формулою (6.8) фактичне значення *DW*-критерію

$$DW = \frac{62{,}8427}{234{,}0595} = 0{,}2685.$$

За табл. 5 Додатку при рівні значущості α = 0,05, кількості параметрів моделі $m$ = 2 та кількості спостережень $n$ = 20 знаходимо критичні точки $d_L$ = 1,20 та $d_U$ =1,41. Розраховане значення потрапляє в інтервал [0; $d_L$), тому з ймовірністю 0,95 можна стверджувати, що у вибірковій сукупності наявна додатна автокореляція.

д) Тестування автокореляції за *критерєм Неймана.*

Обчислимо фактичне значення критерію Неймана

$$Q = Q_{факт} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\ 2}} \cdot \frac{n}{n-1} = 0{,}2685 \cdot \frac{20}{20-1} \approx 0{,}2826$$

За табл. 6 Додатку знаходимо критичне значення $Q_{табл} = 0{,}00323$. Оскільки $Q_{табл.} < Q_{факт}$, то у вибірковій сукупності існує додатна автокореляція.

3) *Оцінювання параметрів економетричної моделі з автокорельованими залишками.*

а) *Метод перетворень вхідної інформації*.

Сформуємо матрицю $T_1$ для перетворення вхідних даних. Для обчислення коефіцієнта автокореляції скористуємося даними табл. 6.4 та формулою (6.12):

$$\rho \approx \frac{\sum_{t=2}^{n}(e_t e_{t-1})}{\sum_{t=1}^{n} e_t^{\,2}} \cdot \frac{n}{n-1} + \frac{k+1}{n} = \frac{158{,}2278}{234{,}0595} \cdot \frac{20}{19} + \frac{2}{20} \approx 0{,}8116.$$

Тоді матриця перетворення $T_1$ буде мати вигляд:

$$T_1 = \begin{pmatrix} 0{,}5842 & 0 & 0 & \dots & 0 & 0 & 0 \\ -0{,}8116 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -0{,}8116 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & -0{,}8116 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & -0{,}8116 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -0{,}8116 & 1 \end{pmatrix}$$

Перетворимо змінні $Y$ і $X$ на основі матриці $T_1$:

$$T_1Y = \begin{pmatrix} 22{,}1186 \\ 8{,}1830 \\ \cdots\cdots\cdots \\ 27{,}0388 \\ 17{,}9228 \end{pmatrix} \quad T_1X = \begin{pmatrix} 0{,}5842 & 26{,}1731 \\ 0{,}1884 & 14{,}8405 \\ \cdots\cdots & \cdots\cdots\cdots \\ 0{,}1884 & 38{,}8927 \\ 0{,}1884 & 27{,}6057 \end{pmatrix}$$

Для перетворених даних скористаємося МНК:

$$B= (X^T X)^{-1} X^TY \rightarrow B= (X^{*T} X^*)^{-1} X^{*T}Y^*.$$

Таблиця 6.4

| $t$ | $e_t$ | $e_{t-1}$ | $e_t e_{t-1}$ | $e_t^2$ |
|---|---|---|---|---|
| 1 | 7,3836 | 0,0000 | 0,0000 | 54,5176 |
| 2 | 4,8581 | 7,3836 | 35,8705 | 23,6014 |
| 3 | 4,1877 | 4,8581 | 20,3445 | 17,5371 |
| 4 | 0,6229 | 4,1877 | 2,6086 | 0,3880 |
| 5 | -2,3696 | 0,6229 | -1,4761 | 5,6151 |
| 6 | -2,6919 | -2,3696 | 6,3787 | 7,2462 |
| 7 | -2,4835 | -2,6919 | 6,6852 | 6,1677 |
| 8 | -2,2488 | -2,4835 | 5,5849 | 5,0572 |
| 9 | -1,0025 | -2,2488 | 2,2545 | 1,0051 |
| 10 | -3,1566 | -1,0025 | 3,1646 | 9,9638 |
| 11 | -2,2566 | -3,1566 | 7,1229 | 5,0920 |
| 12 | -2,9700 | -2,2566 | 6,7020 | 8,8210 |
| 13 | -1,7519 | -2,9700 | 5,2031 | 3,0691 |
| 14 | -2,6370 | -1,7519 | 4,6196 | 6,9535 |
| 15 | -2,5730 | -2,6370 | 6,7848 | 6,6201 |
| 16 | -2,3915 | -2,5730 | 6,1532 | 5,7193 |
| 17 | -1,1520 | -2,3915 | 2,7551 | 1,3272 |
| 18 | 1,3752 | -1,1520 | -1,5842 | 1,8911 |
| 19 | 5,4004 | 1,3752 | 7,4264 | 29,1638 |
| 20 | 5,8569 | 5,4004 | 31,6293 | 34,3031 |
| Сума | 0,0000 | -5,8569 | **158,2278** | **234,0595** |

Позначимо $T_1 Y_t = Y^*$, $T_1 X_t = X^*$. Отримаємо

$$X^{*T} X^* = \begin{pmatrix} 1{,}015746 & 92{,}9135 \\ 92{,}9135 & 11146{,}33 \end{pmatrix} \qquad (X^{*T} X^*)^{-1} = \begin{pmatrix} 4{,}145275 & -0{,}0346 \\ -0{,}03455 & 0{,}0004 \end{pmatrix}$$

$$X^{*T} Y^* = \begin{pmatrix} 59{,}4662 \\ 6855{,}908 \end{pmatrix} \qquad B = \begin{pmatrix} 9{,}603574 \\ 0{,}535029 \end{pmatrix}$$

Вибіркова економетрична модель має вигляд:

$$\hat{y}_t = 9{,}6036 + 0{,}535\ x_t.$$

б) *Метод Ейткена*.

Для оцінки параметрів економетричної моделі з автокорельованими залишками скористаємося формулою:

$$B = (X^T S^{-1} X)^{-1} X^T S^{-1} Y,$$

де $S^{-1}$ – матриця, обернена до дисперсійно-коваріаційной матриці $S$:

$$S = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{14} & \rho^{19} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{15} & \rho^{18} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{16} & \rho^{17} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{17} & \rho^{16} & \rho^{15} & \rho^{14} & \dots & \rho & \rho^2 \\ \rho^{18} & \rho^{17} & \rho^{16} & \rho^{15} & \dots & 1 & \rho \\ \rho^{19} & \rho^{18} & \rho^{17} & \rho^{16} & \dots & \rho & 1 \end{pmatrix}.$$

Для сформування дисперсійно-коваріаційной матриці $S$, необхідно визначити величину коефіцієнта автокореляції $\rho$, яка характеризує взаємозв'язок між послідовними членами ряду залишків. Нехай залишки описуються автокореляційною моделлю першого порядку. За формулою (6.12)

$$\rho \approx \frac{\sum_{t=2}^{n}(e_t e_{t-1})}{\sum_{t=1}^{n} e_t{}^2} \cdot \frac{n}{n-1} + \frac{k+1}{n} \approx \frac{158{,}2278}{234{,}0595} \cdot \frac{20}{19} + \frac{2}{20} \approx 0{,}8116.$$

Отже, матриця $S$ має такий вигляд:

S=

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 | 0,0663 | 0,0538 | 0,0437 | 0,0354 | 0,0288 | 0,0233 | 0,0189 |
| 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 | 0,0663 | 0,0538 | 0,0437 | 0,0354 | 0,0288 | 0,0233 |
| 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 | 0,0663 | 0,0538 | 0,0437 | 0,0354 | 0,0288 |
| 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 | 0,0663 | 0,0538 | 0,0437 | 0,0354 |
| 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 | 0,0663 | 0,0538 | 0,0437 |
| 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 | 0,0663 | 0,0538 |
| 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 | 0,0663 |
| 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 | 0,0817 |
| 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 | 0,1006 |
| 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 | 0,1240 |
| 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 | 0,1528 |
| 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 | 0,1882 |
| 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 | 0,2319 |
| 0,0663 | 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 | 0,2858 |
| 0,0538 | 0,0663 | 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 | 0,3521 |
| 0,0437 | 0,0538 | 0,0663 | 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 | 0,4339 |
| 0,0354 | 0,0437 | 0,0538 | 0,0663 | 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 | 0,5346 |
| 0,0288 | 0,0354 | 0,0437 | 0,0538 | 0,0663 | 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 | 0,6587 |
| 0,0233 | 0,0288 | 0,0354 | 0,0437 | 0,0538 | 0,0663 | 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 | 0,8116 |
| 0,0189 | 0,0233 | 0,0288 | 0,0354 | 0,0437 | 0,0538 | 0,0663 | 0,0817 | 0,1006 | 0,1240 | 0,1528 | 0,1882 | 0,2319 | 0,2858 | 0,3521 | 0,4339 | 0,5346 | 0,6587 | 0,8116 | 1 |

$S^{-1} =$

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2,929857 | -2,37786 | 1,41E-16 | 6,73E-16 | -1,3E-16 | -2,5E-16 | 1,07E-16 | -5,8E-17 | -2,4E-16 | 9,53E-17 | 2,89E-17 | 4,34E-17 | -5,8E-17 | -5,2E-18 | -5E-17 | 1,33E-17 | 2,64E-17 | 2,33E-18 | 0 | 6,94E-18 |
| -2,37786 | 4,859713 | -2,37786 | -1,8E-15 | 4,4E-16 | 1,63E-16 | 1,1E-16 | -3,8E-17 | 2,71E-16 | -2,5E-16 | 8,92E-17 | -6,4E-17 | 2,24E-16 | -1,6E-17 | 6,55E-17 | 1,22E-17 | -2,9E-17 | 1,79E-17 | -1E-17 | -1E-17 |
| -5,3E-16 | -2,37786 | 4,859713 | -2,37786 | -1E-15 | 4,76E-16 | -6,3E-16 | 2,67E-16 | -4E-16 | 3,98E-16 | -6,9E-16 | 3,4E-16 | -4,3E-16 | 1,84E-16 | -4,2E-17 | -2,1E-16 | 7,86E-17 | 3,29E-17 | 1,02E-17 | -2E-17 |
| 5,51E-16 | -1,1E-15 | -2,37786 | 4,859713 | -2,37786 | -1,5E-16 | 5,43E-16 | -2,1E-16 | 6,61E-16 | -7,3E-16 | 1,33E-15 | -1,3E-15 | 7,77E-16 | -5,5E-16 | 3,08E-16 | 7,8E-17 | -8,9E-17 | -5,5E-17 | -2E-17 | 8,13E-17 |
| 2,87E-16 | -1,6E-16 | -8E-16 | -2,37786 | 4,859713 | -2,37786 | -1E-15 | 8,2E-16 | -2,3E-17 | -4,7E-17 | -9,6E-16 | 1,6E-15 | -1,1E-15 | 7,37E-16 | -7E-16 | 5,62E-16 | -2,2E-16 | 5,9E-17 | -1E-16 | 2,03E-17 |
| -1,2E-16 | 1,26E-16 | 9,36E-16 | -1,1E-15 | -2,37786 | 4,859713 | -2,37786 | -2,2E-15 | 1,05E-15 | 9,06E-16 | -8,1E-16 | -2,2E-17 | 9,13E-16 | -8,5E-16 | 2,74E-16 | -6,9E-16 | 7,29E-16 | -3,4E-16 | 2,64E-16 | -1,6E-16 |
| 5,77E-18 | 1,38E-16 | -4E-16 | 6,6E-16 | -1,3E-15 | -2,37786 | 4,859713 | -2,37786 | -2,7E-15 | 1,4E-15 | 3,95E-16 | -5,3E-16 | -3,3E-16 | 5,06E-16 | 1,27E-16 | 2,51E-16 | -8,7E-16 | 5,86E-16 | -2,4E-16 | 8,13E-17 |
| -6,6E-17 | 2,76E-16 | -5,8E-18 | -4E-16 | 2,01E-15 | -4E-15 | -2,37786 | 4,859713 | -2,37786 | -3,4E-15 | 3,14E-15 | -1,3E-15 | -3,1E-16 | 3,21E-16 | 1,17E-16 | -3,2E-16 | 3E-16 | -1,6E-16 | 0 | 8,13E-17 |
| -1,3E-16 | 5,77E-18 | -2E-16 | 4,02E-16 | -6,7E-16 | 1,73E-15 | -4,3E-15 | -2,37786 | 4,859713 | -2,37786 | -6,6E-15 | 3,26E-15 | -1,9E-16 | 2,57E-16 | -1E-15 | 1,25E-15 | -1,6E-16 | -4,2E-16 | 2,44E-16 | -1,2E-16 |
| 3,3E-17 | -1,3E-16 | 5,31E-16 | -5,3E-16 | -2,8E-16 | 1,08E-15 | 1,56E-15 | -4,9E-15 | -2,37786 | 4,859713 | -2,37786 | -3,3E-15 | 6,07E-16 | -1,2E-15 | 2,6E-15 | -2,6E-15 | 1,04E-15 | -1,2E-16 | 3,25E-16 | -3,3E-16 |
| 3,44E-17 | -3E-32 | -6,1E-16 | 1,27E-15 | -1,1E-15 | -2,6E-16 | 2,52E-16 | 3,01E-15 | -6,2E-15 | -2,37786 | 4,859713 | -2,37786 | 5,07E-16 | -5,3E-16 | -6,5E-16 | 1,99E-15 | -1,9E-15 | 1,97E-15 | -2E-15 | 1,06E-15 |
| 3,44E-17 | -6,6E-17 | 3,3E-16 | -1,3E-15 | 1,94E-15 | -1,3E-15 | 6,48E-16 | -1,8E-15 | 3,16E-15 | -2,7E-15 | -2,37786 | 4,859713 | -2,37786 | 3,6E-15 | -3,9E-15 | 4,18E-16 | 2,45E-15 | -3,6E-15 | 3,5E-15 | -1,4E-15 |
| -3,3E-17 | 1,02E-16 | -1,7E-16 | 4,92E-16 | -1,2E-15 | 1,81E-15 | -1,6E-15 | 6,31E-16 | 2,41E-16 | -8,6E-16 | 1,33E-15 | -2,37786 | 4,859713 | -2,37786 | 5,34E-15 | -2,6E-15 | -6,7E-16 | 2,7E-15 | -2,9E-15 | 1,3E-15 |
| 7,87E-32 | -3,2E-17 | 1,99E-16 | -4E-16 | 6,33E-16 | -1E-15 | 9,87E-16 | -2,2E-16 | -7,1E-16 | 1,18E-15 | -1,6E-15 | 4,83E-15 | -2,37786 | 4,859713 | -2,37786 | 6,31E-15 | -3,3E-15 | 1,02E-15 | 1,46E-15 | -1,6E-15 |
| -1,6E-17 | 3,44E-17 | -1,3E-16 | 4,02E-16 | -4,7E-16 | 9,03E-17 | -3,4E-16 | 7,83E-16 | -7E-16 | 1,18E-15 | -1,4E-16 | -3,7E-15 | 6,52E-15 | -2,37786 | 4,859713 | -2,37786 | 4,11E-15 | -2,6E-15 | -3,3E-16 | 1,46E-15 |
| 7,22E-19 | 1,65E-17 | -1,8E-17 | -6,5E-17 | 2,64E-16 | -3E-16 | 5,94E-16 | -8,2E-16 | 9,84E-16 | -2,2E-15 | 1,6E-15 | 1,2E-16 | -2E-15 | 8,1E-15 | -2,37786 | 4,859713 | -2,37786 | 1,85E-15 | 1,63E-16 | -1,6E-16 |
| 3,44E-17 | -1E-16 | 5,02E-17 | 4,95E-17 | -2,7E-16 | 6,34E-16 | -7,1E-16 | 1,51E-16 | -8E-17 | 1,38E-15 | -2E-15 | 2,39E-15 | -1,1E-15 | -4,2E-15 | 6,5E-15 | -2,37786 | 4,859713 | -2,37786 | 0 | -6,5E-16 |
| -1,6E-17 | 6,6E-17 | -1,6E-17 | -1,4E-16 | 2,5E-16 | -4,7E-16 | 3,59E-16 | 1,25E-16 | -1,2E-16 | -8,8E-16 | 2E-15 | -3,8E-15 | 4,37E-15 | -7,5E-16 | -2,5E-15 | 2,45E-15 | -2,37786 | 4,859713 | -2,37786 | 6,51E-16 |
| 7,22E-19 | -1,7E-17 | 5,92E-17 | -1,8E-17 | -8,3E-17 | 2,65E-16 | -2,4E-16 | 1,26E-16 | -3,5E-16 | 1,06E-15 | -1,7E-15 | 3E-15 | -4,1E-15 | 3,67E-15 | -1,4E-15 | -1,2E-15 | 2,4E-15 | -2,37786 | 4,859713 | -2,37786 |
| 3,83E-18 | -3,8E-18 | -4,3E-17 | 7,37E-17 | -3,8E-18 | -1,1E-16 | 1,65E-16 | -1,5E-16 | 2,51E-16 | -6,4E-16 | 8,03E-16 | -9,7E-16 | 1,44E-15 | -2,2E-15 | 1,84E-15 | 6,72E-16 | -2,2E-15 | 9,76E-16 | -2,37786 | 2,929857 |

$$X^T S^{-1} X = \begin{pmatrix} 2{,}975991 & 272{,}2232 \\ 272{,}2232 & 32657{,}14 \end{pmatrix} \qquad X^T S^{-1} Y = \begin{pmatrix} 174{,}2274 \\ 20086{,}83 \end{pmatrix}$$

$$(X^T S^{-1} X)^{-1} = \begin{pmatrix} 1{,}414839 & -0{,}01179 \\ -0{,}01179 & 0{,}000129 \end{pmatrix} \qquad B = \begin{pmatrix} 9{,}603574 \\ 0{,}535029 \end{pmatrix}$$

Таким чином, параметри моделі $b_0 = 9{,}6036$, $b_1 = 0{,}5350$.

Вибіркова економетрична модель має вигляд:

$$\hat{y} = 9{,}6036 + 0{,}5350\, x_t.$$

Оцінки параметрів кореляційно-регресійної моделі, які визначені за методом перетворення вихідної інформації, не відрізняються від оцінок, що здобуті за методом Ейткіна. Це означає, що обидва методи є альтернативними, коли залишки описуються автокореляцією першого порядку.

Запишемо необхідні розрахунки для знаходження фактичного значення $DW$-критерію в табл. 6.5.

За формулою (6.8) фактичне значення $DW$-критерію

$$DW = \frac{\sum_{t=2}^{n} (v_t - v_{t-1})^2}{\sum_{t=1}^{n} v_t^2} = \frac{60{,}4042}{324{,}7868} \approx 0{,}186$$

(тут залишкі позначені $v_t$, для того, щоб їх відрізняти від залишків $e_t$, які обчислені у пункті 2 г)).

За табл. 5 Додатку при рівні значущості $\alpha = 0{,}05$, кількості параметрів моделі $m = 2$ та кількості спостережень $n = 20$ знаходимо критичні точки $d_L = 1{,}20$ і $d_U = 1{,}41$. Розраховане значення потрапляє в інтервал $[0; d_L)$, тому з ймовірністю 0,95 можна стверджувати, що у вибірковій сукупності наявна додатна автокореляція. Отже, від автокореляції залишків не звільнились. Це означає, що вихідна гіпотеза про те, що залишки описуються автокореляцією першого порядку, не виконується. Якщо залишки описуються автокореляцією вищого порядку, то доцільно виконати оцінювання параметрів моделі іншими методами, наприклад, методом Кочрена - Оркатта або Дарбіна [11].

Таблиця 6.5

| $t$ | $y_t$ | $\hat{y}$ | $v_t$ | $v_{t-1}$ | $v_t^2$ | $(v_t-v_{t-1})^2$ | $v_t\ v_{t-1}$ |
|---|---|---|---|---|---|---|---|
| 1 | 37,86 | 33,5729 | 4,2871 | | 18,3795 | | |
| 2 | 38,91 | 36,9971 | 1,9129 | 4,2871 | 3,6594 | 5,6368 | 8,2011 |
| 3 | 38,91 | 37,6391 | 1,2709 | 1,9129 | 1,6152 | 0,4122 | 2,4312 |
| 4 | 43,39 | 45,3435 | -1,9535 | 1,2709 | 3,8162 | 10,3969 | -2,4827 |
| 5 | 50,23 | 54,7600 | -4,5300 | -1,9535 | 20,5210 | 6,6384 | 8,8494 |
| 6 | 48,12 | 53,0479 | -4,9279 | -4,5300 | 24,2844 | 0,1583 | 22,323 |
| 7 | 51,01 | 55,6161 | -4,6061 | -4,9279 | 21,2158 | 0,1036 | 22,698 |
| 8 | 49,01 | 53,4759 | -4,4659 | -4,6061 | 19,9446 | 0,0196 | 20,570 |
| 9 | 45,34 | 48,7677 | -3,4277 | -4,4659 | 11,7490 | 1,0780 | 15,307 |
| 10 | 49,89 | 55,1880 | -5,2980 | -3,4277 | 28,0692 | 3,4982 | 18,160 |
| 11 | 50,79 | 55,1880 | -4,3980 | -5,2980 | 19,3427 | 0,8100 | 23,300 |
| 12 | 50,3 | 55,4020 | -5,1020 | -4,3980 | 26,0309 | 0,4956 | 22,439 |
| 13 | 49,06 | 53,0479 | -3,9879 | -5,1020 | 15,9035 | 1,2413 | 20,346 |
| 14 | 51,08 | 55,8301 | -4,7501 | -3,9879 | 22,5632 | 0,5809 | 18,942 |
| 15 | 61,2 | 65,4606 | -4,2606 | -4,7501 | 18,1526 | 0,2396 | 20,238 |
| 16 | 64,51 | 68,4568 | -3,9468 | -4,2606 | 15,5768 | 0,0985 | 16,815 |
| 17 | 72,23 | 74,6631 | -2,4331 | -3,9468 | 5,9199 | 2,2912 | 9,6028 |
| 18 | 79,45 | 79,1573 | 0,2927 | -2,4331 | 0,0857 | 7,4298 | -0,7121 |
| 19 | 91,52 | 86,8617 | 4,6583 | 0,2927 | 21,6993 | 19,0583 | 1,3633 |
| 20 | 92,2 | 87,0758 | 5,1242 | 4,6583 | 26,2579 | 0,2171 | 23,870 |
| **Сума** | | | **-40,542** | **-45,6657** | **324,786** | **60,4042** | **272,26** |

*Примітка***.** Знайдене рівняння регресії не може бути використане для прогнозу, оскільки у ньому не усунута автокореляція випадкових залишків, яка може мати різни причини. Автокореляція залишків може означати, що у рівняння не включений якийсь істотний фактор. Можливо також, що форма зв’язку не є точною.

## 6.10. Методи усунення автокореляції [11]

В силу ряду причин (помилок специфікації, інерційності розглянутих залежностей тощо) в регресійних моделях може мати місце кореляційна залежність між сусідніми випадковими відхиленнями. Це порушує одну з фундаментальних передумов МНК. Внаслідок цього оцінки, що отримані на основі МНК, перестають бути ефективними. Це робить ненадійними висновки про значущість параметрів регресії і за якістю самого рівняння. Тому важливим є вміння визначити наявність автокореляції та усунути її.

1. *Правильна специфікація моделі* (залучення значущих факторів або зміна форми залежності). Основною причиною наявності випадкової величини залишків $e_t$ в узагальненій кореляційно-регресійній моделі є неможливість урахувати всі значущі фактори і взаємозв'язки, що зумовлюють певне значення результуючої змінної. Тому властивості випадкових відхилень $e_t$, серед них і автокореляції, насамперед залежать від вибору форми залежності і переліку факторних ознак. Оскільки автокореляція здебільшого спричинена неправильною специфікацією моделі, то для її усунення потрібно, насамперед, спробувати скоректувати саму модель. Наприклад, автокореляція спричинена відсутністю деякої важливої факторної ознаки в кореляційно-регресійній моделі. Необхідно спробувати ідентифікувати цю факторну ознаку і врахувати її в кореляційно-регресійній моделі.

2. *Зміна форми залежності* (наприклад, лінійної на нелінійну).

3. *Використання $AR(1)$-моделі* (*авторегресійної моделі Маркова першого порядку*). Якщо всі доступні процедури зміни специфікації моделі вичерпані, а автокореляція наявна, можна припустити, що вона обумовлена внутрішніми властивостями певних значень відхилень $e_t$ і скористатися авторегресійним перетворенням. У лінійній моделі або в моделях, що зводяться до лінійної, найдоцільнішим перетворенням є $AR(1)$-модель, що розглянута у [11].

# РОЗДІЛ 7. ЕКОНОМЕТРИЧНІ МОДЕЛІ ІЗ ВЗАЄМОЗАЛЕЖНИМИ ЗМІННИМИ

## 7.1. Системи рівнянь при побудові економетричних моделей

Однією з причин корельованості регресорів з випадковими членами можуть служити фактори, що діють одночасно і на самі регресори, і на змінні, що пояснюються, при фіксованих значеннях регресорів. Іншими словами, в цій економічній ситуації значення залежних змінних і регресорів формуються одночасно під впливом деяких зовнішніх факторів. Отже, модель, що розглядається, не повна: її треба доповнити рівняннями, в яких залежними змінними виступали би самі регресори. Таким чином, необхідно розглядати системи одночасних рівнянь (СОР).

Прикладом [17] такого економічного явища може бути залежність ринкового або банківського курсу гривні до долара США від таких факторів: національного прибутку, середнього рівня заробітної плати, обсягів експорту, імпорту, ціни, за якою купуються енергоносії (нафта і газ), відсоткової ставки кредитів у банках, розміру емісії тощо. У свою чергу, імпорт може залежати від національного прибутку, експорту, інфляції гривні і в тому числі і від курсу гривні до долара США. Отже, спостерігається одночасний двосторонній зв'язок між курсом гривні до долара США й імпортом. Тому для опису економічного процесу потрібно використовувати систему взаємного відношення між характеристиками.

Позначимо через $Y_1$ позначити курс гривні до долара США, $Y_2$ - обсяг імпорту, $X_i$ $(i = \overline{1,n})$ – фактори, від яких залежать ці показники. У цьому випадку СОР можна записати у вигляді:

$$Y_1 = f_1\,(Y_2, X_1, X_2, \ldots, X_n, \varepsilon_1),$$
$$Y_2 = f_2\,(Y_1, X_1, X_2, \ldots, X_n, \varepsilon_2).$$

Записана система рівнянь є економетричною моделлю із взаємозалежними змінними. Тут $Y_1$, $Y_2$ – взаємопов’язані величини, що пояснюються прийнятою моделлю (ендогенні величини); фактори $X_i$ - екзогенні величини. Екзогенні величини визначають ендогенні величини, але самі від них не залежать, тобто між ними існує одностороння стохастична залежність.

Класичним прикладом СОР є одночасне формування попиту $Q_t^D$ і пропозиції $Q_t^S$ товару в залежності від його ціни $P_t$:

$$\begin{aligned} Q_t^D &= \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \varepsilon_{t1} \\ Q_t^S &= \beta_0 + \beta_1 P_t + \varepsilon_{t2}. \end{aligned} \tag{7.1}$$

Тут $I_t$ – доход. Якщо припустити, що ринок знаходиться у стані рівноваги, то в рівностях (7.1) покладаємо $Q_t^D = Q_t^S = Q_t$. В цьому випадку значення $P_t$, що спостерігається, є ціною рівноваги, що формується одночасно зі попитом та пропозицією. Отже, ми повинні вважати $P_t$ і $Q_t$ залежними змінними, а величину доходу $I_t$ – пояснювальною змінною.

Змінні $Q_t$ і $P_t$ формують свої значення згідно з рівняннями (7.1), тобто в середені моделі (ендогенні змінні); змінна $I_t$ вважається в рівняннях (7.1) заданою, її значення формуються зовні моделі (екзогенна змінна).

Наведемо ще декілька прикладів СОР.

*Модель «попит-пропозиція»* [16].

Однією з найпростіших СОР є модель попиту та пропозиції, яка має загальний вигляд:

$$\begin{array}{ll} \text{Функція попиту} & \\ \text{Функція пропозиції} & \\ \text{Функція рівноваги} & \end{array} \begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, \ \ \alpha_1 < 0, \\ q_t^s = \beta_0 + \beta_1 p_t + \varepsilon_{t2}, \ \ \beta_1 < 0, \\ q_t^D = q_t^s = q_t. \end{cases} \tag{7.2}$$

Має місце таке припущення: обсяг попиту $q_t^D$ та обсяг пропозиції $q_t^s$ певного товару в момент часу $t$ – лінійні регресійні функції від ціни цього товару $p_t$ у той самий момент часу. Наявність випадкових відхилень пов'язано у першу чергу з відсутністю низки важливих пояснювальних змінних (доходу, цін на супутні товари, очікувань, цін на ресурси, податків тощо). Зміна одного фактора може відбитися на моделі. Так, на рис. 7.1 показано, як зростання прибутку споживачів може зсунути лінію попиту вгору, що призведе до зміни рівноважної ціни та рівноважної кількості.

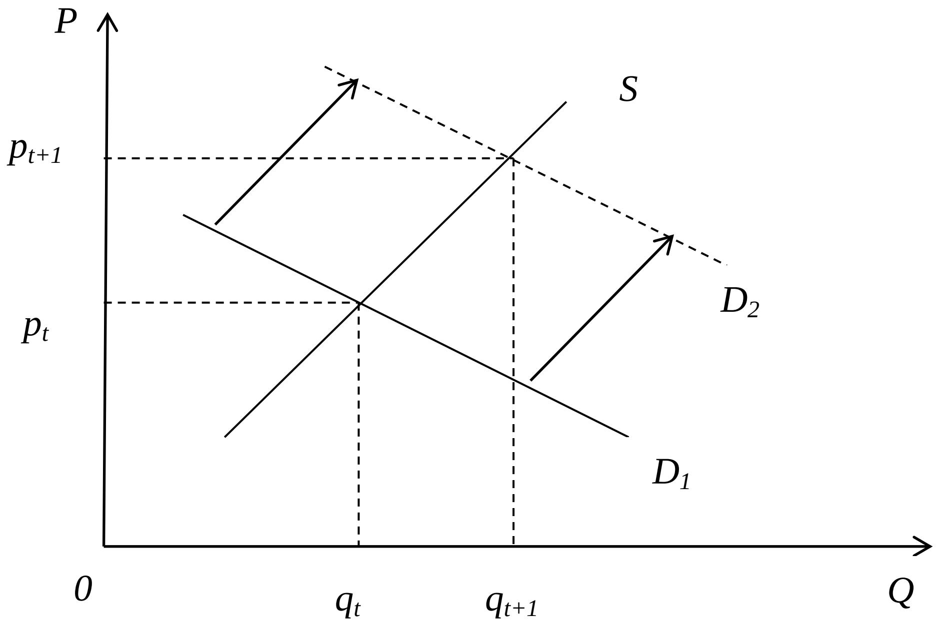



Рис. 7.1

Модель (7.2) може бути вдосконалена. Наприклад, якщо в функцію попиту додати доход споживачів *Y*, то отримаємо таку СОР:

Функція попиту $\quad q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{t1}, \ \ \alpha_1 < 0,$

Функція пропозиції $\quad q_t^s = \beta_0 + \beta_1 p_t + \varepsilon_{t2}, \ \ \beta_1 < 0,$

Функція рівноваги $\quad q_t^D = q_t^s = q_t.$

*Кейнсіанська модель формування доходів*

Опишемо найпростішу модель даного типу в припущенні, що розглядається закрита економіка без державних витрат:

Функція споживання $\quad c_t = \beta_0 + \beta_1 y_t + \varepsilon_{t1}$

Макроекономічна тотожність $\quad y_t = c_t + i_t + \varepsilon_{t2}.$

Тут *Y, C, I* - сукупний випуск, обсяги споживання та інвестицій відповідно ($y_t$, $c_t$, $i_t$ – значення вказаних змінних в момент часу $t$).

*Модель рівноваги ринку товарів* (*модель IS; «інвестиції, investment - заощадження, save»*).

Однією з можливих нестохастичних (що не містить випадкових відхилень) форм моделі *IS* (рівноваги на ринку товарів) є модель:

| | |
|---|---|
| Функція споживання | $c_t = \beta_0 + \beta_1 y_t,$ |
| Функція податків | $\tau_t = \alpha_0 + \alpha_1 y_t$ |
| Функція інвестицій | $i_t = \gamma_0 + \gamma_1 r_t$ |
| Визначення | $y(d)_t = y_t + \tau_t$ |
| Державні витрати | $g_t = \bar{g}$ |
| Макроекономічна тотожність | $y_t = c_t + i_t + g_t$ |

Тут використовуються в момент часу $t$ такі величини: $c_t$ - споживання, $y_t$ – національний доход, $\tau_t$ – обсяг податків, $i_t$ - бажаний обсяг чистих інвестицій, $r_t$ – процентна ставка, $y(d)_t$ - розміщення прибутку, $g_t$ – державні витрати ($g_t = \bar{g}$ = const).

Співвідношення між процентною ставкою й рівнем прибутку, при якому ринок товарів перебуває у стані рівноваги, має вигляд:

$$y_t = \pi_0 + \pi_1 r_{1,}$$

де

$$\pi_0 = \frac{\beta_0 + \alpha_0\beta_1 + \gamma_0 + \bar{g}}{1 - \beta_1(1 - \alpha_1)}; \quad \pi_1 = \frac{1}{1 - \beta_1(1 - \alpha_1)}.$$

*Модель рівноваги на ринку грошей* (*модель LM*; «*ліквідність, liquidity - гроші, money*»).

Рівновага на грошовому ринку задається таким співвідношенням між процентною ставкою та рівнем доходу, при якому попит на гроші дорівнює їх пропозиції. Одна із не стохастичних форм має вигляд:

Функція попиту на гроші
Функція пропозиції грошей
Умова рівноваги

$$\begin{cases} M_t^D = a + by_t - cr_t, \\ M_t^s = M \quad (\text{рівняння } LM) \\ M_t^D = M_t^s. \end{cases}$$

На рис. 7.2 зображена спільна модель *IS* – *LM*.

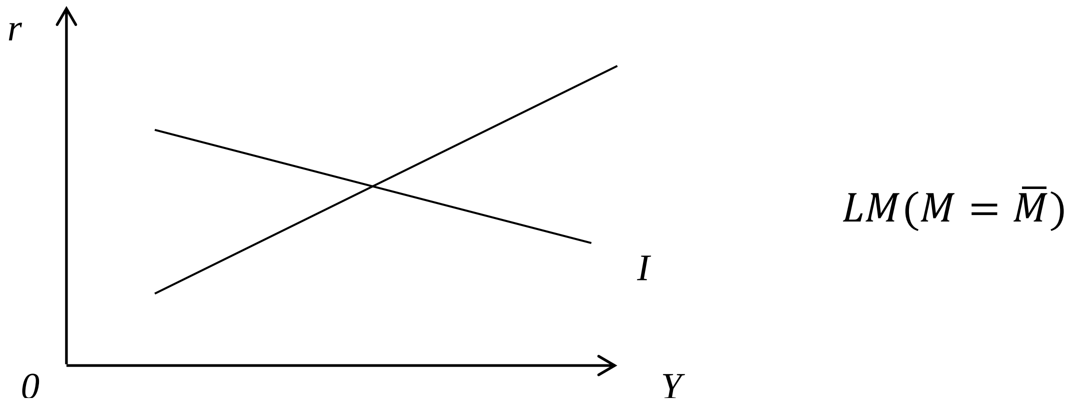



Рис. 7.2

## 7.2 Структурна та зведена форми системи рівнянь

Системи одночасних структурних рівнянь, як правило, включають лінійні рівняння. Нелінійність зв'язків здебільшого апроксимується лінійними співвідношеннями [17].

Запишемо економетричну модель на основі СОР:

$$\begin{aligned} y_1 &= a_{11}y_1 + \ldots + a_{1k}y_k + b_{10} + b_{11}x_1 + \ldots + b_{1m}x_m + \varepsilon_1 \\ y_2 &= a_{21}y_1 + \ldots + a_{2k}y_k + b_{20} + b_{21}x_1 + \ldots + b_{2m}x_m + \varepsilon_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_k &= a_{k1}y_1 + \ldots + a_{kk}y_k + b_{k0} + b_{k1}x_1 + \ldots + b_{km}x_m + \varepsilon_k \end{aligned} \tag{7.3}$$

Систему (7.3) можна записати в матричній формі

$$Y = AY + BX + \varepsilon, \tag{7.4}$$

де $Y$ - вектор ендогенних змінних розмірності ($k \times 1$);

$X$ - вектор екзогенних змінних розмірності ($m \times 1$);

$\varepsilon$ - вектор залишків розмірності ($k \times 1$);

$A$ - матриця коефіцієнтів при змінних $Y$ розмірності ($k \times k$);

$B$ - матриця коефіцієнтів при змінних $X$ розмірності ($k \times m$).

Економетрична модель у вигляді (7.3) безпосередньо відображає структуру зв'язків між змінними і тому називається *структурною формою* моделі. Якщо кожне рівняння системи розв’язати відносно $Y$, то одержимо *зведену форму* моделі, яка має вигляд:

$$Y = RX + u, \tag{7.5}$$

де $R = (E - A)^{-1}B$, $u = (E - A)^{-1}\varepsilon$ – залишки, які є лінійною комбінацією залишків $\varepsilon$.

Чисельна оцінка параметрів моделі на основі одночасних рівнянь пов'язана з проблемою *ідентифікації.*Якщо ніяка лінійна комбінація рівнянь структурної форми не може привести до рівняння, що має ті самі змінні, як і деякі рівняння в структурній формі,то модель буде ідентифікованою.

Необхідною умовою ідентифікації системи є справедливість виконання нерівності для кожного рівняння:

$$k_s - 1 \leq m - m_s, \tag{7.6}$$

де $k_s$ – кількість ендогенних змінних, які входять в $s$-те рівняння структурної форми; $m$ – загальна кількість екзогенних змінних моделі; $m_s$ – кількість екзогенних змінних, які не входять в $s$-те рівняння структурної форми моделі. Для ідентифікації моделей зведена форма (7.5) визначається однозначно.

Якщо співвідношення (7.6) виконується як рівність, то відповідне рівняння буде точно ідентифікованим (визначник $det(E\text{-}A) \neq 0$), а коли як нерівність, то відповідне рівняння буде надідентифікованим.

## 7.3. Методи оцінювання параметрів із взаємозалежними змінними

Якщо між змінними існує взаємозалежність, то одним із наслідків такої залежності буде порушення припущення про незалежність значень факторів і випадкових величин, а через це оцінки невідомих параметрів такої моделі (що отримані за допомогою МНК) будуть зміщеними (оцінками з відхиленнями).

Існують альтернативні методи оцінки параметрів моделі, основаних на СОР, які порівняно з МНК дозволяють уникнути зміщення. До таких методів відносяться:

1) метод зменшеної форми, або метод непрямих найменших квадратів;
2) метод інструментальних змінних;
3) двокроковий МНК;
4) метод найбільшої вірогідності обмеженої інформації;
5) метод змішаного оцінювання;
6) трикроковий МНК;
7) метод найбільшої вірогідності повної інформації.

Зазначимо, що двокроковий МНК [6] застосовується тоді, коли рівняння структурної форми моделі (7.4) надідентифіковані і непрямий МНК застосувати не можна, а користуватись МНК недоцільно. Двокроковий МНК призначений для оцінки параметрів окремих рівнянь системи моделі. Основна ідея методу полягає в тому, щоб очистити поточні ендогенні змінні від стохастичної складової, бо вони пов’язані із залишками.

Для практичного використання трикрокового МНК необхідно виконання вимог [18]:

- розпочинаючи оцінювати параметри моделі необхідно вилучити всі тотожності;
- виключити з системи кожне не ідентифіковане рівняння;
- за наявності серед рівнянь системи точно ідентифікованих та над ідентифікованих рівнянь доцільно застосовувати трикроковий метод найменших квадратів до кожної з груп рівнянь окремо;
- якщо група над ідентифікованих рівнянь має тільки одне рівняння, то трикроковий метод найменших квадратів перетворюється на двокроковий МНК;
- в разі блочно-діагональної матриці коваріацій залишків

використання трикрокового методу найменших квадратів може бути застосовано для кожної групи рівнянь, які відповідають одному блоку.

Перші п’ять методів називаються методами одного рівняння, оскільки вони застосовуються тільки до одного з рівнянь системи. Шостий та сьомий методи називають системними методами, оскільки, вони застосовуються одночасно до всіх рівнянь системи.

## 7.4. Рекурсивні системи

Особливим випадком моделей із взаємозалежними змінними є рекурсивні моделі. Розглянемо економетричну модель [19], що складається із системи рівнянь, де матриця $A$ має трикутний вид:

$$\begin{aligned} y_1 &= b_{10} + b_{11} x_1 + \ldots + b_{1m} x_m + \varepsilon_1, \\ y_2 &= a_{21} + b_{20} + b_{21} x_1 + \ldots + b_{2m} x_m + \varepsilon_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_k &= a_{k1} y_1 + \ldots + a_{k,k-1} y_{k-1} + b_{k0} + b_{k1} x_1 + \ldots + b_{km} x_m + \varepsilon_k. \end{aligned} \tag{7.7}$$

Для цієї системи рівнянь матриця $A$ має вигляд:

$$A = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ a_{21} & 0 & 0 & \ldots & 0 & 0 \\ a_{31} & a_{32} & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{k1} & a_{k2} & a_{k3} & \ldots & a_{k,k-1} & 0 \end{pmatrix}$$

Економетрична модель, в якій матриця $A$ має трикутний вид і залишки взаємонезалежні, називається рекурсивною моделлю. Рекурсивність системи дозволяє застосовувати МНК для оцінки невідомих параметрів.

Маючи значення екзогенних змінних, можемо застосувати МНК до першого рівняння і отримати оцінене значення $y_1$. Далі використовуємо обчислене $y_1$ як незалежну змінну в другому рівнянні. Знову можна застосувати МНК, оскільки всі змінні незалежні від випадкової величини $\varepsilon_2$ і т.д. Для того, щоб визначити, чи є модель рекурсивною, достатньо дослідити вид матриці *A*. Як свідчать дослідження, реальні економічні системи найчастіше описуються рекурсивними системами рівнянь.

### 7.5. Непрямий метод найменших квадратів

У випадку, коли кожне рівняння моделі точно ідентифіковане, для оцінки параметрів моделі можна застосувати непрямий МНК.

Алгоритм цього методу складається з трьох кроків [18].

*Крок 1*. Перехід від структурної форми моделі до зведеної. Для цього кожне рівняння структурної форми розв'язується відносно однієї з *k* залежних ендогенних змінних моделі.

*Крок 2.* Оцінка параметрів кожного рівняння зведеної форми моделі за допомогою МНК.

*Крок 3*. Розрахунок оцінок параметрів рівнянь структурної форми на основі співвідношення:

$$R = (E - A)^{-1}B,$$

де $A$ і $B$ – параметри структурних рівнянь (шукані величини), а $R$ – матриця оцінок параметрів зведеної форми моделі.

Отримані цим методом оцінки із збільшенням кількості спостережень наближаються до істинних значень.

### 7.6. Приклад динамічної моделі Клейна [ 9 ]

*Постановка задачі.* Розглядається динамічна модель Клейна, яка була розроблена в 1950 році для економіки США періоду 1920-1941 років. Структурна форма моделі складається з таких рівнянь і тотожностей:

$$C_t = \alpha_0 + \alpha_1 P_1 + \alpha_2 V_t + \alpha_3 P_{t-1} + \varepsilon_{t1} \quad \text{(функція споживання);}$$

$$I_t = \beta_0 + \beta_1 P_t + \beta_1 P_{t-1} + \beta_3 K_{t-1} + \varepsilon_{t2} \quad \text{(функція інвестицій);}$$

$$W_t = \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 t + \varepsilon_{t3} \quad \text{(функція заробітної плати);}$$

$$P_t = C_t + I_t + G_t - T_t - W_t \quad \text{(прибуток бізнес-сектора);}$$

$$X_t = T_t + P_t + W_t \quad \text{(доход бізнес-сектора);}$$

$$V_t = W_t + WG_t \quad \text{(загальна оплата праці);}$$

*Ендогенні змінні:*

$C_t$ – кінцеве споживання за період $t$;

$I_t$ - інвестиції за період $t$;

$W_t$ - заробітна плата у приватному секторі за період $t$;

$P_t$ - прибуток в бізнес-секторі за період $t$;

$X_t$ - доходи за період $t$ в приватному секторі;

$V_t$ - заробітна плата в цілому за період $t$.

Екзогенні змінні:

а) лагові змінні:

$P_{t-1}$ - прибуток за період $t - 1$;

$K_{t-1}$ - основний капітал за період $t - 1$;

$X_{t-1}$ - доходи за період $t - 1$;

б) екзогенні змінні:

$G_t$ - урядові витрати, виключаючи заробітну плату держслужбовців, за період $t$;

$T_t$ - сплачені податки за період $t$;

$WG_t$ - заробітна плата держслужбовців за період $t$; $t$ - фактор часу.

Вибіркові дані, які характеризують факторні ознаки моделі, наведені в табл. 7.1. Необхідно знайти оцінки параметрів системи одночасних рівнянь, використовуючи двокроковий МНК.

***Розв'язання***

На основі вибіркових даних таблиці 7.1 сформуємо у таблиці 7.2 значення ендогенних факторів та пояснювальних змінних системи.

*Крок* 1. Знайдемо залежність кожної ендогенної змінної від екзогенних змінних за допомогою МНК, тобто визначимо параметри зведеної форми СОР. Отримана система має вигляд:

$$C_t^* = R_{10}^* + R_{11}^* P_{t-1} + R_{12}^* X_{t-1} + R_{13}^* K_{t-1} +$$

$$+ R_{14}^* G_t + R_{15}^* T_t + R_{16}^* WG_t + R_{17}^* t;$$

$$I_t^* = R_{20}^* + R_{21}^* P_{t-1} + R_{22}^* X_{t-1} + R_{23}^* K_{t-1} +$$

$$+ R_{24}^* G_t + R_{25}^* T_t + R_{26}^* WG_t + R_{27}^* t;$$

$$W_t^* = R_{30}^* + R_{31}^* P_{t-1} + R_{32}^* X_{t-1} + R_{33}^* K_{t-1} +$$
$$+ R_{34}^* G_t + R_{35}^* T_t + R_{36}^* WG_t + R_{37}^* t;$$

$$P_t^* = R_{40}^* + R_{41}^* P_{t-1} + R_{42}^* X_{t-1} + R_{43}^* K_{t-1} +$$
$$+ R_{44}^* G_t + R_{45}^* T_t + R_{46}^* WG_t + R_{47}^* t;$$

$$X_t^* = R_{50}^* + R_{51}^* P_{t-1} + R_{52}^* X_{t-1} + R_{53}^* K_{t-1} +$$
$$+ R_{54}^* G_t + R_{55}^* T_t + R_{56}^* WG_t + R_{57}^* t;$$

$$V_t^* = R_{60}^* + R_{61}^* P_{t-1} + R_{62}^* X_{t-1} + R_{63}^* K_{t-1} +$$
$$+ R_{64}^* G_t + R_{65}^* T_t + R_{66}^* WG_t + R_{67}^* t.$$

Оцінки параметрів зведеної форми СОР обчислюємо за МНК для вибіркових даних, що записані у табл. 7.2.

$$C_t^* = 58{,}302 + 0{,}748\, P_{t-1} + 0{,}230\, X_{t-1} - 0{,}147\, K_{t-1} +$$
$$+ 0{,}205\, G_t + 0{,}366\, T_t + 0{,}193\, WG_t + 0{,}701\, t;$$

$$I_t^* = 35{,}518 + 0{,}926\, P_{t-1} - 0{,}113\, X_{t-1} - 0{,}193\, K_{t-1} +$$
$$+0{,}100\, G_t - 0{,}162\, T_t - 0{,}717\, WG_t + 0{,}332\, t;$$

$$W_t^* = 43{,}436 + 0{,}872\, P_{t-1} + 0{,}095\, X_{t-1} - 0{,}123\, K_{t-1} +$$
$$+0{,}866\, G_t - 0{,}604\, T_t - 0{,}444\, WG_t + 0{,}714\, t;$$

$$P_t^* = 50{,}384 + 0{,}803\, P_{t-1} + 0{,}022\, X_{t-1} - 0{,}216\, K_{t-1} +$$
$$+0{,}439\, G_t - 0{,}923\, T_t - 0{,}080\, WG_t + 0{,}319\, t;$$

$$X_t^* = 93{,}820 + 1{,}674\, P_{t-1} + 0{,}117\, X_{t-1} - 0{,}339\, K_{t-1} +$$
$$+1{,}305\, G_t - 0{,}527\, T_t - 0{,}523\, WG_t + 1{,}033\, t;$$

$$V_t^* = 43{,}436 + 0{,}872\, P_{t-1} + 0{,}095\, X_{t-1} - 0{,}123\, K_{t-1} +$$
$$+0{,}866\, G_t - 0{,}604\, T_t + 0{,}556\, WG_t + 0{,}714\, t.$$

Використовуючи знайдені рівняння залежностей кожної ендогенної змінної від пояснювальних змінних, визначаємо регресійні значення ендогенних змінних (табл. 7.3).

*Крок* 2. Знаходимо значення параметрів структурної форми. Для цього підставимо в кожне рівняння структурної форми моделі замість ендогенних змінних, що входять в якості пояснювальних змінних до правих частин рівнянь, їх регресійні значення, сформовані на попередньому кроці (табл. 4.3).

Структурна форма одночасних рівнянь приймає вигляд:

$$C_t = \alpha_0 + \alpha_1 P_t^* + \alpha_2 V_t^* + \alpha_3 P_{t\text{-}1} + v_{t1};$$
$$I_t = \beta_0 + \beta_1 P_t^* + \beta_2 P_{t\text{-}1} + \beta_3 K_{t\text{-}1} + v_{t2};$$
$$W_t = \gamma_0 + \gamma_1 X_t^* + \gamma_2 X_{t\text{-}1} + \gamma_0 t + v_{t3}.$$

Застосуємо МНК для знаходження оцінок параметрів цієї системи рівнянь:

$$C_t = 16{,}555 + 0{,}017\, P_t^* + 0{,}810\, V_t^* + 0{,}216\, P_{t\text{-}1} + v_{t1};$$
$$I_t = 20{,}278 + 0{,}150\, P_t^* + 0{,}616\, P_{t\text{-}1} - 0{,}158\, K_{t\text{-}1} + v_{t2};$$
$$W_t = 1{,}497 + 0{,}439\, X_t^* + 0{,}146\, X_{t\text{-}1} + 0{,}130\, t + v_{t3}.$$

Не приймаючи до уваги взаємозв’язки між економічними показниками, які виражені структурною формою моделі, оцінимо кожне рівняння звичайним МНК окремо від інших (не розглядаючи всі рівняння як систему). Отримаємо такі результати для кожного окремого рівняння:

$$C_t = 16{,}237 + 0{,}193 P_t + 0{,}796\, V_t + 0{,}090\, P_{t\text{-}1} + e_{t1};$$
$$I_t = 10{,}126 + 0{,}4807\, P_t + 0{,}333\, P_{t\text{-}1} - 0{,}112\, K_{t\text{-}1} + e_{t2};$$
$$W_t = 1{,}497 + 0{,}439\, X_t + 0{,}146\, X_{t\text{-}1} + 0{,}130\, t + e_{t3}.$$

Таким чином, нехтування структурними зв’язками СОР приводить до розбіжностей в оцінюванні параметрів деяких рівнянь.

Таблиця 7.1

| Період | $C_t$ | $I_t$ | $W_t$ | $P_t$ | $X_t$ | $V_t$ | $K_t$ | $G_t$ | $T_t$ | $WG_t$ | $t$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1920 | 39,8 | 2,7 | 28,8 | 12,7 | 44,9 | 31,0 | 182,8 | 2,4 | 3,4 | 2,2 | -11 |
| 1921 | 41,9 | -0,2 | 25,5 | 12,4 | 45,6 | 28,2 | 182,6 | 3,9 | 7,7 | 2,7 | -10 |
| 1922 | 45,0 | 1,9 | 29,3 | 16,9 | 50,1 | 32,2 | 184,5 | 3,2 | 3,9 | 2,9 | -9 |
| 1923 | 49,2 | 5,2 | 34,1 | 18,4 | 57,2 | 37,0 | 189,7 | 2,8 | 4,7 | 2,9 | -8 |
| 1924 | 50,6 | 3,0 | 33,9 | 19,4 | 57,1 | 37,0 | 192,7 | 3,5 | 3,8 | 3,1 | -7 |
| 1925 | 52,6 | 5,1 | 35,4 | 20,1 | 61,0 | 38,6 | 197,8 | 3,3 | 5,5 | 3,2 | -6 |
| 1926 | 55,1 | 5,6 | 37,4 | 19,6 | 64,0 | 40,7 | 203,4 | 3,3 | 7,0 | 3,3 | -5 |
| 1927 | 56,2 | 4,2 | 37,9 | 19,8 | 64,4 | 41,5 | 207,6 | 4,0 | 6,7 | 3,6 | -4 |
| 1928 | 57,3 | 3,0 | 39,2 | 21,1 | 64,5 | 42,9 | 210,6 | 4,2 | 4,2 | 3,7 | -3 |
| 1929 | 57,8 | 5,1 | 41,3 | 21,7 | 67,0 | 45,3 | 215,7 | 4,1 | 4,0 | 4,0 | -2 |
| 1930 | 55,0 | 1,0 | 37,9 | 15,6 | 61,2 | 42,1 | 216,7 | 5,2 | 7,7 | 4,2 | -1 |
| 1931 | 50,9 | -3,4 | 34,5 | 11,4 | 53,4 | 39,3 | 213,3 | 5,9 | 7,5 | 4,8 | 0 |
| 1932 | 45,6 | -6,2 | 29,0 | 7.0 | 44,3 | 34,3 | 207,1 | 4,9 | 8,3 | 5,3 | 1 |
| 1933 | 46,5 | -5,1 | 28,5 | 11,2 | 45,1 | 34,1 | 202,0 | 3,7 | 5,4 | 5,6 | 2 |
| 1934 | 48,7 | -3,0 | 30,6 | 12,3 | 49,7 | 36,6 | 199,0 | 4,0 | 6,8 | 6,0 | 3 |
| *1935* | *51,3* | *-1,3* | *33,2* | *14,0* | *54,4* | *39,3* | *197,7* | *4,4* | *7,2* | *6,1* | *4* |
| 1936 | 57,7 | 2,1 | 36,8 | 17,6 | 62,7 | 44,2 | 199,8 | 2,9 | 8,3 | 7,4 | 5 |
| 1937 | 58,7 | 2,0 | 41,0 | 17,3 | 65,0 | 47,7 | 201,8 | 4,3 | 6,7 | 6,7 | 6 |
| 1938 | 57,5 | -1,9 | 38,2 | 15,3 | 60,9 | 45,9 | 199,9J | 5,3 | 7,4 | 7,7 | 7 |
| 1939 | 61,6 | 1,3 | 41,6 | 19,0 | 69,5 | 49,4 | 201,2 | 6,6 | 8,9 | 7,8 | 8 |
| 1940 | 65,0 | 3,3 | 45,0 | 21,1 | 75,7 | 53,0 | 204,5 | 7,4 | 9,6 | 8,0 | 9 |
| 1941 | 69,7 | 4,9 | 53,3 | 23,5 | 88,4 | 61,8 | 209,4 | 13,8 | 11,6 | 8,5 | 10 |

Таблиця 7.2

| Період | $C_t$ | $I_t$ | $W_t$ | $P_t$ | $X_t$ | $V_t$ | $P_{t-1}$ | $X_{t-1}$ | $K_{t-1}$ | $G_t$ | $T_t$ | $WG_t$ | $t$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1921 | 41,9 | -0,2 | 25,5 | 12,4 | 45,6 | 28,2 | 12,7 | 44,9 | 182,8 | 3,9 | 7,7 | 2,7 | -10 |
| 1922 | 45,0 | 1,9 | 29,3 | 16,9 | 50,1 | 32,2 | 12,4 | 45,6 | 182,6 | 3,2 | 3,9 | 2,9 | -9 |
| 1923 | 49,2 | 5,2 | 34,1 | 18,4 | 57,2 | 37,0 | 16,9 | 50,1 | 184,5 | 2,8 | 4,7 | 2,9 | -8 |
| 1924 | 50,6 | 3,0 | 33,9 | 19,4 | 57,1 | 37,0 | 18,4 | 57,2 | 189,7 | 3,5 | 3,8 | 3,1 | -7 |
| 1925 | 52,6 | 5,1 | 35,4 | 20,1 | 61,0 | 38,6 | 19,4 | 57,1 | 192,7 | 3.3 | 5,5 | 3,2 | -6 |
| 1926 | 55,1 | 5,6 | 37,4 | 19,6 | 64,0 | 40,7 | 20,1 | 61,0 | 197,8 | 3,3 | 7,0 | 3,3 | -5 |
| 1927 | 56,2 | 4,2 | 37,9 | 19,8 | 64,4 | 41,5 | 19,6 | 64,0 | 203.4 | 4,0 | 6,7 | 3,6 | -4 |
| 1928 | 57,3 | 3,0 | 39,2 | 21,1 | 64,5 | 42,9 | 19,8 | 64,4 | 207,6 | 4,2 | 4,2 | 3,7 | -3 |
| 1929 | 57,8 | 5,1 | 41,3 | 21,7 | 67,0 | 45,3 | 21,1 | 64,5 | 210,6 | 4,1 | 4,0 | 4,0 | -2 |
| 1930 | 55,0 | 1,0 | 37,9 | 15,6 | 61,2 | 42,1 | 21,7 | 67,0 | 215,7 | 5,2 | 7,7 | 4,2 | -1 |
| 1931 | 50,9 | -3,4 | 34,5 | 11,4 | 53,4 | 39,3 | 15,6 | 61,2 | 216,7 | 5,9 | 7,5 | 4,8 | 0 |
| 1932 | 45,6 | -6,2 | 29,0 | 7,0 | 44,3 | 34,3 | 11,4 | 53,4 | 213,3 | 4,9 | 8,3 | 5,3 | 1 |
| 1933 | 46,5 | -5,1 | 28,5 | 11,2 | 45,1 | 34,1 | 7,0 | 44,3 | 207,1 | 3,7 | 5,4 | 5,6 | 2 |
| 1934 | 48,7 | -3,0 | 30,6 | 12,3 | 49,7 | 36,6 | 11,2 | 45,1 | 202,0 | 4,0 | 6,8 | 6,0 | 3 |
| 1935 | 51,3 | -1,3 | 33,2 | 14,0 | 54,4 | 39,3 | 12,3 | 49,7 | 199,0 | 4,4 | 7,2 | 6,1 | 4 |
| 1936 | 57,7 | 2,1 | 36,8 | 17,6 | 62,7 | 44,2 | 14,0 | 54,4 | 197,7 | 2,9 | 8,3 | 7,4 | 5 |
| 1937 | 58,7 | 2,0 | 41,0 | 17,3 | 65,0 | 47,7 | 17,6 | 62,7 | 199,8 | 4,3 | 6,7 | 6,7 | 6 |
| 1938 | 57,5 | -1,9 | 38,2 | 15,3 | 60,9 | 45,9 | 17,3 | 65,0 | 201,8 | 5,3 | 7,4 | 7,7 | 7 |
| 1939 | 61,6 | 1,3 | 41,6 | 19,0 | 69,5 | 49,4 | 15,3 | 60,9 | 199,9 | 6,6 | 8,9 | 7,8 | 8 |
| 1940 | 65,0 | 3,3 | 45,0 | 21,1 | 75,7 | 53,0 | 19,0 | 69.5 | 201,2 | 7,4 | 9,6 | 8,0 | 9 |
| 1941 | 69,7 | 4,9 | 53,3 | 23,5 | 88,4 | 61,8 | 21,1 | 75,7 | 204,5 | 13,81 | 11,6 | 8,5 | 10 |

Таблиця 7.3

| Період | $C_t^*$ | $I_t^*$ | $W_t^*$ | $P_t^*$ | $X_t^*$ | $V_t^*$ |
|---|---|---|---|---|---|---|
| 1921 | 42,838 | 0,923 | 26,706 | 13,256 | 47,662 | 29,406 |
| 1922 | 44,790 | 1,337 | 28,850 | 16,577 | 49,327 | 31,750 |
| 1923 | 49,240 | 4,795 | 32,853 | 19,282 | 56,835 | 35,753 |
| 1924 | 52,446 | 4,787 | 35,973 | 20,960 | 60,733 | 39,073 |
| 1925 | 52,789 | 5,113 | 35,936 | 19,767 | 61,202 | 39,136 |
| 1926 | 53,634 | 4,358 | 36,054 | 18,239 | 61,292 | 39,354 |
| 1927 | 54,142 | 2,714 | 36,583 | 17,573 | 60,856 | 40,183 |
| 1928 | 55,444 | 2,730 | 38,632 | 19,542 | 62,374 | 42,332 |
| 1929 | 56,812 | 3,484 | 40,021 | 20,375 | 64,396 | 44,021 |
| 1930 | 56,700 | 2,478 | 39,498 | 17,180 | 64,378 | 43,698 |
| 1931 | 51,690 | -2,707 | 34,678 | 12,705 | 54,883 | 39,478 |
| 1932 | 47,552 | -5,320 | 29,832 | 9,000 | 47,132 | 35,132 |
| 1933 | 44,649 | -6,712 | 27,184 | 9,054 | 41,638 | 32,784 |
| 1934 | 49,050 | -2,080 | 31,499 | 12,671 | 50,970 | 37,499 |
| 1935 | 52,027 | -0,766 | 34,040 | 14,421 | 55,661 | 40,140 |
| 1936 | 54,813 | -0,399 | 34,303 | 14,712 | 57,315 | 41,703 |
| 1937 | 60,546 | 2,829 | 41,178 | 19,796 | 67,675 | 47,878 |
| 1938 | 61,401 | 1,509 | 41,603 | 19,207 | 68,210 | 49,303 |
| 1939 | 59,678 | 0,632 | 40,591 | 17,420 | 66,911 | 48,391 |
| 1940 | 64,882 | 2,996 | 45,372 | 20,306 | 75,278 | 53,372 |
| 1941 | 68,774 | 3,899 | 52,216 | 22,657 | 86,473 | 60,716 |

# ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

## Тести

I. *Вказати правильну відповідь* (*відповіді*) *на запитання*

1. При виконанні передумов МНК залишки рівняння регресії характеризуються:

а) гомоскедастичністю;

б) нульовою середньою величиною;

в) випадковим характером;

г) високим ступенем автокореляції.

2. До методів виявлення гетероскедастичності залишків відносяться:

а) критерій Дарбіна - Уотсона ;

б) тест Голдфельда – Квандта;

в) графічний аналіз залишків;

г) метод найменших квадратів.

3. На практиці гетероскедастичність наявна, якщо маємо підставу вважати, що:

а) ймовірнісні розподіли випадкових похибок при різних спостереженнях будуть різними;

б) ймовірнісні розподіли випадкових похибок при різних спостереженнях будуть однаковими;

в) дисперсії випадкових похибок однакові.

4. Мультиколінеарність означає, що в моделі регресії має місце:

а) матриця парних коефіцієнтів кореляції додатна;

б) визначник матриці коефіцієнтів кореляції не дорівнює 0;

в) визначник матриці коефіцієнтів кореляції близький до 0.

5. Узагальненою лінійною множинною називають регресію:

а) зі стохастичними регресійними пояснювальними змінними;

б) з довільною коваріаційною матрицею вектора похибок;

в) з ненульовим математичним сподіванням.

6. Критерій Дарбіна - Уотсона використовується для:

а) перевірки моделі на кореляцію залишків;

б) визначення економічної значущості моделі у цілому;

в) визначення статистичної значущості моделі у цілому;

г) порівняння двох альтернативних варіантів моделі.

7. Для рівняння регресії $y = b_0 + b_1 x + e$ МНК використовується при оцінюванні:

а) $b_1$; б) $x$; в) $y$; г) $e$.

8. Тест Гольдфельда - Квандта використовується для перевірки залишків на

а) гетероскедастичність;

б) гомоскедастичність;

в) автокорельованість;

г) мультиколінеарність.

9. Корельованість двох або декількох факторів у рівнянні – це:

а) автокореляція;

б) мультиколінеарність;

в) гетероскедастичність;

г) гомоскедастичність.

10. Якщо розрахункове значення $F$-статистики значущості коефіцієнта детермінації більше критичного, то нульова гіпотеза $H_0$:

а) відкидається, тобто коефіцієнт детермінації є значущим;

б) приймається, тобто коефіцієнт детермінації є незначущим;

в) відкидається, тобто коефіцієнт детермінації є незначущим;

г) приймається, тобто коефіцієнт детермінації є значущим.

11. Ідентифікація моделі – це:

а) єдність відповідності між зведеною та структурною формами моделі;

б) перевага ендогенних змінних над екзогенними;

в) перевага екзогенних змінних над ендогенними.

12. В системі одночасних рівнянь

$$y_1 = b_{12} y_2 + a_{11} x_1 + \varepsilon_1$$
$$y_2 = b_{21} y_1 + a_{22} x_2 + \varepsilon_2$$

екзогенними змінними є

а) $x_1$ **;** б) $x_2$ ; в) $y_1$ ; г) $y_2$ ; д) $\varepsilon_1$ ; е) $\varepsilon_2$ .

13. Методи оцінювання коефіцієнтів структурної моделі:

а) непрямий МНК;

б) двокроковий та трьохкроковий МНК;

в) метод максимальної правдоподібності;

г**)** непрямий, двокроковий та трьохкроковий МНК.

14. Величина коефіцієнта регресії показує:

а) щільність зв’язку між факторами, що досліджуються;

б) щільність зв’язку між фактором та результатом;

в) характер зв’язку між фактором та результатом;

г) середню зміну результату при зміні фактора на одиницю.

15. Узагальнений МНК використовується, якщо випадкові члени рівняння регресії:

а) гетероскедастичні;

б) мультиколінеарні;

в) гомоскедастичні;

г) незалежні від спостережень.

16. Передумовами МНК є:

а) дисперсії випадкових відхилень сталі для всіх спостережень;

б) випадкові відхилення є взаємозалежними;

в) випадкові відхилення корелюють між собою.

17. Узагальнений МНК використовується у випадку порушення передумов МНК про:

а) гомоскедастичність залишків;

б) мінімізації залишків;

в) нормальному розподілі залишків;

г) мультиколінеарність.

18. Число ступенів вільності пов’язане із:

а) обсягом сукупності та типом рівняння регресії;

б) характером змінних, що досліджуються;

в) тільки з обсягом сукупності;

г) тільки з типом рівняння регресії.

19. Критичні значення критерію Ст'юдента визначаються за:

а) рівнем значущості та числом ступенів вільності;

б) рівнем значущості;

в) числом ступенів вільності;

г) двом ступеням вільності.

20. Синонімом системи взаємозалежних рівнянь є система:

а) одночасних рівнянь;

б) спільних систем;

в) мультиколінеарних систем;

г) структурних рівнянь.

21. Змінні, що задаються «ззовні», автономні від моделі, називаються:

а) екзогенними;

б) структурними;

в) ендогенними;

г) лаговими.

22. Оцінки параметрів ідентифікованої системи економетричних рівнянь можуть бути знайдені за допомогою:

а) непрямого МНК;

б) узагальненого МНК;

В) звичайного МНК.

23. Під частинною кореляцією розуміють:

а) залежність між результативним і одним факторним ознаками при фіксованих значеннях інших факторних ознак;

б) залежність результативної ознаки і двох або більше факторних ознак, які включені у дослідження;

в) зв'язок між двома ознаками (результативною і факторною чи двома факторними);

г) залежність між якісними ознаками.

24. Для оцінювання значущості коефіцієнта кореляції використовують критерій:

а) Ст'юдента; б) Фішера; в) Пірсона; г) Дарбіна - Уотсона.

25. Для оцінювання значущості рівняння регресії використовують критерій:

а) Ст'юдента; б) Фішера; в) Пірсона; г) Дарбіна - Уотсона.

26. Системами економетричних рівнянь є системи:

а) одночасних рівнянь;

б) рекурсивних рівнянь;

г) незалежних рівнянь.

27. Система одночасних рівнянь відрізняється від інших типів економетричних систем тим, що в ній:

а) одні й ті ж самі ендогенні змінні в одних рівняннях знаходяться у лівій частині, а в інших рівняннях – у правій частині;

б) ендогенна змінна одного рівняння знаходиться у другому рівнянні як фактор;

в) кожна ендогенна змінна є функцією однієї й тієї сукупності екзогенних змінних.

28. Екзогенні змінні моделі характеризуються тим, що вони:

а) є незалежними та визначаються поза системою;

б) датуються попередніми моментами часу;

в) є залежними та визначаються усередині системи.

29. Виберіть аналог поняття «ендогенна змінна»:

а) залежна змінна, що визначається усередині системи;

б) результат;

в) фактор.

30. Рівняння множинної регресії має вигляд: $y=20+0{,}8x_1+0{,}3x_2+\varepsilon$.

На результативну ознаку більш впливає:

а) $x_1$; б) $x_1$ і $x_2$; в) $x_2$; г) неможливо зробити висновок.

31. Для усунення мультиколінеарності використовується метод:

а) включення факторів;

б) виключення факторів;

в) метод найменших квадратів;

г) ідентифікації.

32. Для отримання якісних оцінок параметрів системи одночасних рівнянь використовують:

а) звичайний МНК;

б) двокроковий МНК;

в) метод ідентифікації рівнянь;

г) метод перетворення моделі до зведеної форми.

33. Якщо автокореляція відсутня, то статистика $DW$ наближено дорівнює:

а) 1; б) -1; в) 2; г) 0; д) -2.

34. Значення статистики $DW$ знаходиться між значеннями:

а) 0 і 6; б) -3 і 3; в) 0 і 4; г) -2 і 2; д) 0 і 2.

35. При до додаванні ще однієї змінної у рівняння регресії коефіцієнт детермінації:

а) залишається незмінним;

б) зменшується;

в) не зменшується;

г) збільшується.

36. Чим більше число спостережень, тим зона невизначеності для $DW$-критерію:

а) розташована лівіше;

б) більш вузька;

в) більш ширша;

г) розташована правіше;

д) незмінна.

37. Зона невизначеності тесту Дарбіна - Уотсона:

а) $DW > d_2$; б) $DW < d_1$; в) $d_1 < DW < d_2$; г) $DW = 0$; д) $DW \neq 0$, де $d_1$, $d_2$ – нижня та верхня границі.

38. Гетероскедастичність призводить до:

а) зміщення оцінок параметрів регресії;

б) зменшення дисперсії оцінок параметрів регресії;

в) неефективності оцінок параметрів регресії;

г) збільшення дисперсії оцінок параметрів регресії.

39. Параметри множинної регресії $\beta_1, \beta_2, \ldots, \beta_k$ характеризують:

а) ступінь впливу;

б) випадковість;

в) рівень незалежності;

г) несталість.

відповідних економічних факторів.

40. Гетероскедастичність полягає у тому, що дисперсія випадкового члена регресії:

а) залежить від номера спостережень;

б) залежить від числа спостережень;

в) залежить від часу проведення спостережень;

г) однакова для всіх спостережень.

41. Які з тверджень є вірними:

а) у множинну лінійну модель доцільно включати мультиколінеарні фактори;

б) мультиколінеарність факторів призводить до зменшення надійності оцінок параметрів рівняння регресії;

в) мультиколінеарність факторів проявляється у наявності парних коефіцієнтів між факторної кореляції із значеннями, більшими 0,7;

г) мультиколінеарність факторів проявляється у наявності парних коефіцієнтів між факторної кореляції із значеннями, меншими 0,3.

42. В лінійній множинній регресії $y = b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k$ параметри при факторах характеризують:

а) долю дисперсії результативної змінної, яка пояснюється регресією у загальній дисперсії;

б) щільність зв'язку між результативною змінною та відповідним фактором при усуненні впливу інших факторів, що включені у модель;

в) середню зміну результативної змінної із зміною відповідного фактора на одиницю при незмінному значенні інших факторів, що закріплені на середньому рівні;

г) на скільки приблизно процентів у середньому змінюється результативна змінна при зміні відповідного фактора на 1%.

43. Щільність сумісного впливу факторів на результат у рівнянні лінійної множинної регресії оцінює:

а) коефіцієнт парної кореляції;

б) коефіцієнт частинної кореляції;

в) коефіцієнт множинної кореляції;

г) коефіцієнт множинної детермінації.

44. МНК – оцінка вектора параметрів множинної лінійної регресійної моделі має вигляд:

а) $(X\cdot X^T)^{-1}\cdot X^T\cdot Y$;

б) $(X^T\cdot X)^{-1}\cdot X^T\cdot Y$;

в) $(X^T\cdot X)^{-1}\cdot Y\cdot X^T$;

г) $(X^T\cdot X)\cdot X^T\cdot Y$.

45. Залежність дисперсії випадкових залишків від номера спостереження називається:

а) гомоскедастичністю;

б) гетероскедастичністю;

в) автокореляцією;

г) мультиколінеарністю.

46. Причини гетероскедастичності:
    а) дослідження неоднорідних об’єктів;
    б) помилки вимірювань;
    в) масштабність вимірювань;
    г) розкид спостережень.

47. Причини автокореляції :
    а) дослідження неоднорідних об’єктів;
    б) помилки вимірювань;
    в) масштабність вимірювань;
    г) нелінійність моделі;
    д) характер даних.

48. Гетероскедастичність дає:
    а) оцінки параметрів з відхиленням;
    б) найкращі лінійні оцінки;
    в) ефективні оцінки параметрів;
    г) проблеми із статистичними висновками;
    д) високий ступень кореляції між залишками та залежною змінною.

49. Мультиколінеарність дає:
    а) оцінки параметрів з відхиленням;
    б) найкращі лінійні оцінки;
    в) неефективні оцінки параметрів;
    г) проблеми із статистичними висновками.

50. Автокореляція дає:
    а) оцінки параметрів з відхиленням;
    б) найкращі лінійні оцінки;
    в) неефективні оцінки параметрів;
    г) проблеми із статистичними висновками.

II. *Дати відповіді на запитання з поясненнями* [19 ]

| № | Запитання | Варіанти відповіді | |
|---|---|---|---|
| 1 | Гетероскедастичність може бути визначена за допомогою тесту Дарбіна - Уотсона. | так | ні |
| 2 | Коли наявна досконала мультиколінеарність, неможливо застосовувати МНК. | так | ні |
| 3 | Мультиколінеарність наявна тоді, коли незалежна змінна сильно корелює із залежною змінною. | так | ні |
| 4 | Невключення в модель доречної незалежної змінної призводить до помилки специфікації. | так | ні |
| 5 | Включення недоречної пояснювальної змінної в регресію призводить до помилки специфікації. | так | ні |
| 6 | Використання незалежних змінних, що виміряні з помилкою, веде до помилки специфікації. | так | ні |
| 7 | Включення недоречної пояснювальної змінної в регресію гірше, ніж невключення доречної пояснювальної змінної. | так | ні |
| 8 | Тест Дарбіна — Уотсона завжди засвідчить наявність автокореляції. | так | ні |
| 9 | Тест Дарбіна — Уотсона засвідчить наявність мультиколінеарності. | так | ні |
| 10 | Одним із методів для тестування гетероскедастичності є графічний аналіз випадкових відхилень | так | ні |

## Запитання для самоконтролю

1. Який загальний вигляд має узагальнена лінійна множинна кореляційно-регресійної модель (для генеральної сукупності) у скалярній та матричній формах?

2. Який загальний вигляд має вибіркова лінійна множинна кореляційно-регресійна модель у скалярній та матричній формах?

3. За якою формулою обчислюються оцінки параметрів лінійної множинної кореляційно-регресійної моделі?

4. Запишіть властивості оцінок параметрів лінійної множинної кореляційно-регресійної моделі.

5. Назвіть основні припущення багатофакторного кореляційно-регресійного аналізу.

6. У чому сутність теореми Гаусса - Маркова?

7. Які наслідки невиконання кожного з припущень теореми Гаусса - Маркова.

8. Що означає мультиколінеарність змінних?

9. Які наслідки явища мультиколінеарності?

10. Назвіть ознаки мультиколінеарності.

11. Які критерії використовуються для виявлення явища мультиколінеарності?

12. Дайте характеристику алгоритму Феррара - Глобера.

13. Назвіть способи усунення мультиколінеарності.

14. Для чого призначений метод головних компонент та у чому полягає його ідея?

15. Дайте означення гомоскедастичності.

16. Дайте означення гетероскедастичності.

17. Як на оцінку параметрів моделі впливає явище гетероскедастичності?

18. Назвіть методи визначення гетероскедастичності.

19. Порівняйте сутність явищ гетероскедастичності та автокореляції.

20. У яких випадках застосовується параметричний тест Гольдфельда - Квандта?

21. Сутність тесту Глейзера.

22. У яких випадках використовується узагальнений МНК (метод Ейткена)?

23. Сутність методу Ейткена.

24. Як використовується матриця кореляції залишків $S$ у методі Ейткена?

25. Властивості матриці кореляції залишків $S$.

26. Виконання прогнозу за методом Ейткена.

27. Дайте означення автокореляції.

28. При порушені якої умови застосування МНК виникає автокореляція залишків?

29. Назвіть причини виникнення автокореляції.

30. До яких наслідків може призвести автокореляція залишків?

31. За якими методами перевіряється наявність автокореляції?

32. Використання критерію Дарбіна - Уотсона та його структура.

33. Використання критерію Неймана та його структура.

34. Циклічний коефіцієнт автокореляції та його використання.

35. Які методи використовують для оцінювання параметрів моделі з автокорельованими залишками?

36. Коли використовується метод перетворення вхідної інформації?

37. Коли використовується ітераційний метод Корчена-Оркатта? Дайте характеристику цьому методу.

38. Коли використовується ітераційний метод Дарбіна? Дайте характеристику методу.

39. Запишіть формулу прогнозу залежної змінної при автокореляції залишків.

40. У чому сутність одночасних рівнянь?

41. Наведіть приклади систем одночасних рівнянь.

42. Дайте означення ендогенних та екзогенних величин.

43. Що розуміють під структурною формою економетричної моделі?

44. Записати загальний вигляд зведеної моделі.

45. Сутність проблеми ідентифікації моделі.

46. Методи оцінювання параметрів із взаємозалежними змінними.

47. Яку модель називають рекурсивною?

48. Сутність непрямого МНК.

49. Сутність двокрокового МНК.

50. Чи є зв'язок між ідентифікованістю рівняння та оцінками параметрів непрямим та двокроковим МНК?

## ПІСЛЯМОВА

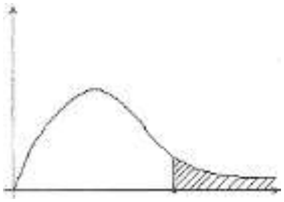
Неможливо стати висококваліфікованим економістом, якщо тільки сидіти вдома та вивчати книжки з економіки. Так і студент, який вивчить матеріал цього посібника, не стане відразу добрим економетристом. Прикладні економічні дослідження містять набагато більше проблем і задач, ніж автори розглянули у цьому посібнику. Годі думати, що практика економетрики складається тільки у механічному використанні відомих процедур. Прикладна економетрика – це вельми тонке балансування між економічною теорією, доступністю даних, попередніми ідеями і, звичайно, економічною теорією. На думку давньогрецького мислителя Платона, нікого нічого не можна навчити – можна лише навчитися.

Автори бажають успіхів у вивченні наведеного матеріалу і сподіваються, що посібник в якійсь мірі допоможе стати висококваліфікованими спеціалістами.

*Through hardship to the stars*!

# ДОДАТКИ

## Таблиця 1
## Графік і таблиця *F*-розподілу Фішера

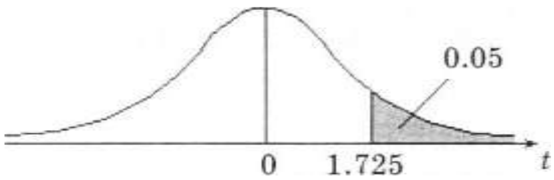


$F(V_1\ V_2, \alpha)$

| $k_2$ \ $k_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 241 | 242 |
| **2** | 18.5 | 19.0 | 19.2 | 19.2 | 19.3 | 19.3 | 19.4 | 19.4 | 19.4 | 19.4 |
| **3** | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 |
| **4** | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 |
| **5** | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 |
| **6** | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 |
| **7** | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 |
| **8** | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 |
| **9** | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 |
| **10** | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 |
| **11** | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 |
| **12** | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 |
| **13** | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 |
| **14** | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 |
| **15** | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 |
| **16** | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 |
| **17** | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 |
| **18** | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 |
| **19** | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 |
| **20** | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 |
| **21** | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 |
| **22** | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 |
| **23** | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 |
| **24** | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 |
| **25** | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 |
| **30** | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 |
| **40** | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 |
| **60** | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 |
| **120** | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.18 | 2.09 | 2.02 | 1.96 | 1.91 |
| **∞** | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 |

| $k_2$ \ $k_1$ | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 244 | 246 | 248 | 249 | 250 | 251 | 252 | 253 | 254 |
| **2** | 19.4 | 19.4 | 19.4 | 19.4 | 19.5 | 19.5 | 19.5 | 19.5 | 19.5 |
| **3** | 8.74 | 8.70 | 8.66 | 8.64 | 8.62 | 8.59 | 8.57 | 8.55 | 8.53 |
| **4** | 5.91 | 5.86 | 5.80 | 5.77 | 5.75 | 5.72 | 5.69 | 5.66 | 5.63 |
| **5** | 4.68 | 4.62 | 4.56 | 4.53 | 4.50 | 4.46 | 4.43 | 4.40 | 4.37 |
| **6** | 4.00 | 3.94 | 3.87 | 3.84 | 3.81 | 3.77 | 3.74 | 3.70 | 3.67 |
| **7** | 3.57 | 3.51 | 3.44 | 3.41 | 3.38 | 3.34 | 3.30 | 3.27 | 3.23 |
| **8** | 3.28 | 3.22 | 3.15 | 3.12 | 3.08 | 3.04 | 3.01 | 2.97 | 2.93 |
| **9** | 3.07 | 3.01 | 2.94 | 2.90 | 2.86 | 2.83 | 2.79 | 2.75 | 2.71 |
| **10** | 2.91 | 2.85 | 2.77 | 2.74 | 2.70 | 2.66 | 2.62 | 2.58 | 2.54 |
| **11** | 2.79 | 2.72 | 2.65 | 2.61 | 2.57 | 2.53 | 2.49 | 2.45 | 2.40 |
| **12** | 2.69 | 2.62 | 2.54 | 2.51 | 2.47 | 2.43 | 2.38 | 2.34 | 2.30 |
| **13** | 2.60 | 2.53 | 2.46 | 2.42 | 2.38 | 2.34 | 2.30 | 2.25 | 2.21 |
| **14** | 2.53 | 2.46 | 2.39 | 2.35 | 2.31 | 2.27 | 2.22 | 2.18 | 2.13 |
| **15** | 2.48 | 2.40 | 2.33 | 2.29 | 2.25 | 2.20 | 2.16 | 2.11 | 2.07 |
| **16** | 2.42 | 2.35 | 2.28 | 2.24 | 2.19 | 2.15 | 2.11 | 2.06 | 2.01 |
| **17** | 2.38 | 2.31 | 2.23 | 2.19 | 2.15 | 2.10 | 2.06 | 2.01 | 1.96 |
| **18** | 2.34 | 2.27 | 2.19 | 2.15 | 2.11 | 2.06 | 2.02 | 1.97 | 1.92 |
| **19** | 2.31 | 2.23 | 2.16 | 2.11 | 2.07 | 2.03 | 1.98 | 1.93 | 1.88 |
| **20** | 2.28 | 2.20 | 2.12 | 2.08 | 2.04 | 1.99 | 1.95 | 1.90 | 1.84 |
| **21** | 2.25 | 2.18 | 2.10 | 2.05 | 2.01 | 1.96 | 1.92 | 1.87 | 1.81 |
| **22** | 2.23 | 2.15 | 2.07 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.78 |
| **23** | 2.20 | 2.13 | 2.05 | 2.01 | 1.96 | 1.91 | 1.86 | 1.81 | 1.76 |
| **24** | 2.18 | 2.11 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.79 | 1.73 |
| **25** | 2.16 | 2.09 | 2.01 | 1.96 | 1.92 | 1.87 | 1.82 | 1.77 | 1.71 |
| **30** | 2.09 | 2.01 | 1.93 | 1.89 | 1.84 | 1.79 | 1.74 | 1.68 | 1.62 |
| **40** | 2.00 | 1.92 | 1.84 | 1.79 | 1.74 | 1.69 | 1.64 | 1.58 | 1.51 |
| **60** | 1.92 | 1.84 | 1.75 | 1.70 | 1.65 | 1.59 | 1.53 | 1.47 | 1.39 |
| **120** | 1.83 | 1.75 | 1.66 | 1.61 | 1.55 | 1.50 | 1.43 | 1.35 | 1.25 |
| **∞** | 1.75 | 1.67 | 1.57 | 1.52 | 1.46 | 1.39 | 1.32 | 1.22 | 1.00 |

## Таблиця 2.
## Графік і таблиця *t*-розподілу Ст'юдента




| *Df/Pr* | 0,25<br>0,50 | 0,10<br>0,20 | 0,05<br>0,10 | 0,025<br>0,05 | 0,01<br>0,02 | 0,005<br>0,010 | 0,001<br>0,002 |
|---|---|---|---|---|---|---|---|
| 1 | 1,000 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 | 318,31 |
| 2 | 0,861 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 22,327 |
| 3 | 0,765 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 10,214 |
| 4 | 0,741 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 7,173 |
| 5 | 0,727 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 | 5,893 |
| 6 | 0,718 | 1,440 | 1,943 | 2,227 | 3,143 | 3,707 | 3,208 |
| 7 | 0,711 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 4,785 |
| 8 | 0,706 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 4,501 |
| 9 | 0,703 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 4,297 |
| 10 | 0,700 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 | 4,144 |
| 11 | 0,697 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 | 4,025 |
| 12 | 0,695 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 | 3,930 |
| 13 | 0,694 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 | 3,852 |
| 14 | 0,692 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 | 3,787 |
| 15 | 0,691 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 | 3,733 |
| 16 | 0,690 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 | 3,686 |
| 17 | 0,689 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 | 3,646 |
| 18 | 0,688 | 1,330 | 1,434 | 2,101 | 2,552 | 2,878 | 3,610 |
| 19 | 0,688 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 | 3,579 |
| 20 | 0,687 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 | 3,552 |
| 21 | 0,686 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 | 3,527 |
| 22 | 0,686 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 | 3,505 |
| 23 | 0,685 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 | 3,485 |
| 24 | 0,685 | 1,318 | 1,711 | 2,064 | 2,492 | 2,797 | 3,467 |
| 25 | 0,684 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 | 3,450 |
| 26 | 0,684 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 | 3,435 |
| 27 | 0,684 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 | 3,421 |
| 28 | 0,683 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 | 3,408 |
| 29 | 0,683 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 | 3,396 |
| 30 | 0,683 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 | 3,385 |
| 40 | 0,081 | 1,303 | 1,684 | 2,021 | 2,423 | 2,704 | 3,307 |
| 60 | 0,679 | 1,296 | 1,671 | 2,000 | 2,390 | 2,660 | 3,232 |
| 120 | 0,677 | 1,289 | 1,658 | 1,980 | 2,358 | 2,167 | 3,160 |
| ∞ | 0,674 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 3,090 |

Ст'юдент [псевдонім Вільяма Сілі Госсета (1876-1937)] - англійський математик і статистик.

## Таблиця 3.

## Критичні значення розподілу $\chi^2(N)$

| *N* | Довірча ймовірність *P* | | | | | |
|---|---|---|---|---|---|---|
| | **0,99** | **0,95** | **0,9** | **0,1** | **0,05** | **0,01** |
| **1** | 6,63 | 3,84 | 2,71 | 0,0158 | 0,0039 | 0,0002 |
| **2** | 9,21 | 5,99 | 4,61 | 0,2107 | 0,1026 | 0,0201 |
| **3** | 11,34 | 7,81 | 6,25 | 0,584 | 0,352 | 0,115 |
| **4** | 13,28 | 9,49 | 7,78 | 1,064 | 0,711 | 0,297 |
| **5** | 15,09 | 11,07 | 9,24 | 1,61 | 1,15 | 0,554 |
| **6** | 16,81 | 12,59 | 10,64 | 2,20 | 1,64 | 0,872 |
| **7** | 18,48 | 14,07 | 12,02 | 2,83 | 2,17 | 1,24 |
| **8** | 20,09 | 15,51 | 13,36 | 3,49 | 2,73 | 1,65 |
| **9** | 21,67 | 16,92 | 14,68 | 4,17 | 3,33 | 2,09 |
| **10** | 23,21 | 18,31 | 15,99 | 4,87 | 3,94 | 2,56 |
| **11** | 24,73 | 19,68 | 17,28 | 5,58 | 4,57 | 3,05 |
| **12** | 26,22 | 21,03 | 18,55 | 6,30 | 5,23 | 3,57 |
| **13** | 27,69 | 22,36 | 19,81 | 7,04 | 5,89 | 4,11 |
| **14** | 29,14 | 23,68 | 21,06 | 7,79 | 6,57 | 4,66 |
| **15** | 30,58 | 25,00 | 22,31 | 8,55 | 7,26 | 5,23 |
| **16** | 32,00 | 26,30 | 23,54 | 9,31 | 7,96 | 5,81 |
| **18** | 34,81 | 28,87 | 25,99 | 10,86 | 9,39 | 7,01 |
| **20** | 37,57 | 31,41 | 28,41 | 12,44 | 10,85 | 8,26 |
| **24** | 42,98 | 36,42 | 33,20 | 15,66 | 13,85 | 10,86 |
| **30** | 50,89 | 43,77 | 40,26 | 20,60 | 18,49 | 14,95 |
| **40** | 63,69 | 55,76 | 51,81 | 29,05 | 26,51 | 22,16 |
| **60** | 88,38 | 79,08 | 74,40 | 46,46 | 43,19 | 37,48 |
| **80** | 112,33 | 101,88 | 96,58 | 64,28 | 60,39 | 53,54 |
| **100** | 135,81 | 124,34 | 118,50 | 82,36 | 77,93 | 70,06 |
| **120** | 158,95 | 146,57 | 140,23 | 100,62 | 95,70 | 86,92 |

## Таблиця 4.
## Критичні значення кількості рядів для визначення наявності автокореляції за допомогою методу рядів

Нижня межа $k_1$

| $n_1$ | $n_2$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 2 | | | | | | | | | | | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | | | | | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | | | | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 |
| 5 | | | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 |
| 6 | | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 |
| 7 | | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 6 |
| 8 | | 2 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 7 |
| 9 | | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 8 |
| 10 | | 2 | 3 | 3 | 4 | 5 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 9 |
| 11 | | 2 | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 9 |
| 12 | 2 | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 10 | 10 |
| 13 | 2 | 2 | 3 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 10 | 10 |
| 14 | 2 | 2 | 3 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 9 | 9 | 9 | 10 | 10 | 11 | 11 |
| 15 | 2 | 3 | 3 | 4 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 12 |
| 16 | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 12 | 12 |
| 17 | 2 | 3 | 4 | 4 | 5 | 6 | 7 | 7 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 11 | 12 | 12 | 13 |
| 18 | 2 | 3 | 4 | 5 | 5 | 6 | 7 | 8 | 8 | 9 | 9 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 |
| 19 | 2 | 3 | 4 | 5 | 6 | 6 | 7 | 8 | 8 | 9 | 10 | 10 | 11 | 11 | 12 | 12 | 13 | 13 | 13 |
| 20 | 2 | 3 | 4 | 5 | 6 | 6 | 7 | 8 | 9 | 9 | 10 | 10 | 11 | 12 | 12 | 13 | 13 | 13 | 14 |

| $n_1$ | $n_2$ | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** |
| **4** | | | | 9 | 9 | | | | | | | | | | | | | | |
| **5** | | | 9 | 10 | 10 | 1 | 1 | | | | | | | | | | | | |
| **6** | | | 9 | 10 | 11 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | |
| **7** | | | | 11 | 12 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | |
| **8** | | | | 11 | 12 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **9** | | | | | 13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **10** | | | | | 13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 |
| **11** | | | | | 13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| **12** | | | | | 13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **13** | | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **14** | | | | | | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **15** | | | | | | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **16** | | | | | | | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **17** | | | | | | | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **18** | | | | | | | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **19** | | | | | | | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **20** | | | | | | | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

Верхня межа

## Таблиця 5.

## Критичні значення $d_1$ і $d_2$ для статистики Дарбіна-Уотсона при $P$ =0,95

| ***n*** | **Розмерність моделі** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *m* = 2 | | *m* = 3 | | *m* =4 | | *m* =5 | | *m* = 6 | |
| | $d_1$ | $d_u$ | $d_1$ | $d_u$ | $d_1$ | $d_u$ | $d_1$ | $d_u$ | $d_1$ | $d_u$ |
| 6 | 0,61 | 1,40 | | | | | | | | |
| 7 | 0,70 | 1,36 | 0,47 | 1,90 | | | | | | |
| 8 | 0,76 | 1,33 | 0,56 | 1,78 | 0,37 | 2,29 | | | | |
| 9 | 0,82 | 1,32 | 0,63 | 1,70 | 0,46 | 2,13 | 0,30 | 2,59 | | |
| 10 | 0,88 | 1,32 | 0,70 | 1,64 | 0,53 | 2,02 | 0,38 | 2,41 | 0,24 | 2,82 |
| 11 | 0,93 | 1,32 | 0,76 | 1,60 | 0,60 | 1,93 | 0,44 | 2,28 | 0,32 | 2,65 |
| 12 | 0,97 | 1,33 | 0,81 | 1,58 | 0,66 | 1,86 | 0,51 | 2,18 | 0,38 | 2,51 |
| 13 | 1,01 | 1,34 | 0,86 | 1,58 | 0,72 | 1,82 | 0,57 | 2,09 | 0,45 | 2,39 |
| 14 | 1,05 | 1,35 | 0,91 | 1,56 | 0,77 | 1,78 | 0,63 | 2,03 | 0,51 | 2,30 |
| 15 | 1,08 | 1,36 | 0,95 | 1,54 | 0,81 | 1,75 | 0,69 | 1,98 | 0,56 | 2,22 |
| 16 | 1,11 | 1,37 | 0,98 | 1,54 | 0,86 | 1,73 | 0,73 | 1,94 | 0,62 | 2,16 |
| 17 | 1,13 | 1,38 | 1,02 | 1,54 | 0,90 | 1,71 | 0,78 | 1,90 | 0,66 | 2,10 |
| 18 | 1,16 | 1,39 | 1,05 | 1,54 | 0,93 | 1,70 | 0,82 | 1,87 | 0,71 | 2,06 |
| 19 | 1,18 | 1,40 | 1,07 | 1,54 | 0,97 | 1,69 | 0,86 | 1,85 | 0,75 | 2,02 |
| 20 | 1,20 | 1,41 | 1,10 | 1,54 | 1,00 | 1,68 | 0,89 | 1,83 | 0,79 | 1,99 |
| 21 | 1,22 | 1,42 | 1,13 | 1,54 | 1,03 | 1,67 | 0,93 | 1,81 | 0,83 | 1,98 |
| 22 | 1,24 | 1,43 | 1,16 | 1.54 | 1,05 | 1,66 | 0,96 | 1,80 | 0,86 | 1,94 |
| 23 | 1,26 | 1,44 | 1.17 | 1,54 | 1.08 | 1,66 | 0,99 | 1,79 | 0,90 | 1,92 |
| 24 | 1,27 | 1,45 | 1,19 | 1,55 | 1,10 | 1,66 | 1,01 | 1,78 | 0,93 | 1,90 |
| 25 | 1,29 | 1,45 | 1,21 | 1,55 | 1,12 | 1,65 | 1.04 | 1,77 | 0,95 | 1,89 |
| 26 | 1,30 | 1,46 | 1,23 | 1,55 | 1,14 | 1,65 | 1,06 | 1,76 | 0,98 | 1,87 |
| 27 | 1,32 | 1,47 | 1,24 | 1.56 | 1,16 | 1,65 | 1,08 | 1,75 | 1,00 | 1,86 |
| 28 | l,33 | 1,48 | 1,26 | 1,56 | 1,18 | 1,65 | 1,10 | 1,75 | 1,03 | 1,85 |
| 29 | 1,34 | 1,48 | 1,27 | 1,56 | 1,20 | 1,65 | 1,12 | 1,74 | 1,05 | 1,84 |
| 30 | 1,35 | 1,49 | 1,28 | 1,57 | 1,21 | 1,65 | 1,14 | 1,74 | 1,07 | 1,83 |
| 35 | 1,40 | 1,52 | 1,34 | 1,58 | 1,28 | 1,65 | 1,22 | 1,73 | 1,16 | 1,80 |
| 40 | 1,44 | 1,54 | 1,39 | 1,60 | 1,34 | 1,66 | 1,29 | 1,72 | 1,23 | 1,79 |
| 50 | 1,50 | 1,59 | 1,46 | 1,63 | 1,42 | 1,67 | 1,38 | 1,72 | 1,34 | 1,77 |
| 60 | 1,55 | 1,62 | 1,51 | 1,65 | 1,48 | 1,69 | 1,44 | 1,73 | 1,41 | 1,77 |
| 70 | 1,58 | 1,64 | 1,55 | 1,67 | 1,53 | 1,70 | 1,49 | 1,74 | 1,46 | 1,77 |
| 80 | 1,61 | 1,66 | 1,59 | 1.69 | 1,56 | 1,72 | 1,53 | 1,74 | 1,51 | 1,77 |
| 100 | 1,65 | 1,69 | 1,63 | 1,72 | 1,61 | 1,74 | 1,59 | 1,76 | 1,57 | 1,78 |
| 150 | 1,72 | 1,75 | 1,71 | 1,76 | 1,69 | 1,77 | 1,68 | 1,79 | 1,67 | 1,80 |
| 200 | 1,76 | 1,78 | 1,75 | 1,79 | 1,74 | 1,80 | 1,73 | 1,81 | 1,72 | 1,82 |

## Таблиця 6

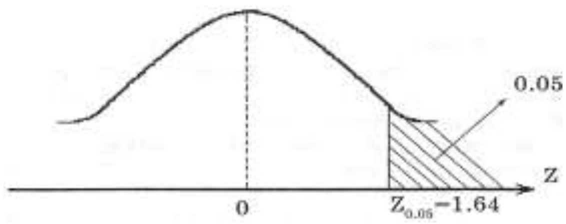
## Критичні значення для відношення фон Неймана

| *k* | *n* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *4* | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| **0,25** | | | | **0,00001** | **0,00001** | **0,00001** | **0,00001** | | **0,00001** |
| **0,30** | | | | **0,00007** | **0.00007** | **0,00005** | **0,00004** | **0,00002** | **0,00003** |
| **0,35** | | | **0,00006** | **0,00027** | **0,00021** | **0,00014** | **0,00009** | **0,00005** | **0,00007** |
| **0,40** | | | **0,00047** | **0,00065** | **0,00047** | **0,00031** | **0,00019** | **0,00012** | **0,00016** |
| **0,45** | | | **0,00126** | **0,00126** | **0,00088** | **0,00059** | **0,00038** | **0,00025** | **0,00031** |
| **0,50** | | **0,00038** | **0,00246** | **0,00214** | **0,00150** | **0,00103** | **0,00069** | **0,00046** | **0,00055** |
| **0,55** | | **0,00223** | **0,00409** | **0,00333** | **0,00237** | **0,00168** | **0,00116** | **0,00080** | **0,00094** |
| **0,60** | | **0,00493** | **0,00615** | **0,00486** | **0,00355** | **0,00259** | **0,00185** | **0,00132** | **0,00152** |
| **0,65** | | **0,00830** | **0,00865** | **0,00678** | **0,00511** | **0,00382** | **0,00282** | **0,00208** | **0,00235** |
| **0,70** | | **0,01225** | **0,01161** | **0,00913** | **0,00710** | **0,00544** | **0,00414** | **0,00313** | **0,00351** |
| **0,75** | | **0,01673** | **0,01505** | **0,01197** | **0,00958** | **0,00753** | **0,00587** | **0,00455** | **0,00508** |
| **0,80** | **0,00356** | **0,02171** | **0,01900** | **0,01534** | **0,01263** | **0,01015** | **0,00809** | **0,00642** | **0,00714** |
| **0,85** | **0,01301** | **0,02717** | **0,02348** | **0,01932** | **0,01631** | **0,01338** | **0,01089** | **0,00883** | **0,00980** |
| **0,90** | **0,02257** | **0,03310** | **0,02851** | **0,02403** | **0,02068** | **0,01729** | **0,01436** | **0,01188** | **0,01316** |
| **0,95** | **0,03221** | **0,03949** | **0,03412** | **0,02957** | **0,02579** | **0,02196** | **0,01858** | **0,01565** | **0,01733** |
| **1,00** | **0,04191** | **0,04634** | **0,04035** | **0,03598** | **0,03171** | **0,02745** | **0,02363** | **0,02025** | **0,02241** |
| **1,05** | **0,05186** | **0,05364** | **0,04728** | **0,04325** | **0,03849** | **0,03384** | **0,02959** | **0,02578** | **0,02852** |
| **1,10** | **0,0618^** | **0,06140** | **0,05500** | **0,05137** | **0,04618** | **0,04120** | **0,03655** | **0,03232** | **0,03577** |
| **1,15** | **0,07194** | **0,06963** | **0,06361** | **0,06036** | **0,05482** | **0,04957** | **0,04458** | **0,03997** | **0,04425** |
| **1,20** | | | **0,07323** | **0,07020** | **0,06445** | **0,05901** | **0,05375** | **0,04882** | **0,05407** |
| **1,25** | | | | | | **0,06956** | **0,06412** | **0,05894** | **0,06531** |
| **1,30** | | | | | | | | **0,07040** | |

| *k* | *n* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 15 | 20 | 25 | 30 | 40 | 50 | 60 |
| 0,35 | 0.00001 | | | | | | |
| 0,40 | 0,00002 | | | | | | |
| 0,45 | 0,00004 | | | | | | |
| 0,50 | 0,00009 | 0,00001 | | | | | |
| 0,55 | 0,00018 | 0,00002 | | | | | |
| 0,60 | 0,00033 | 0,00005 | 0,00001 | | | | |
| 0,65 | 0,00059 | 0,00012 | 0,00002 | | | | |
| 0,70 | 0,00100 | 0,00024 | 0,00005 | 0,00001 | | | |
| 0,75 | 0,00161 | 0,00044 | 0,00011 | 0,00003 | | | |
| 0,80 | 0.00250 | 0,00076 | 0,00023 | 0,00007 | 0,00001 | | |
| 0.85 | 0,00375 | 0,00127 | 0,00044 | 0,00015 | 0,00002 | | |
| 0,90 | 0,00547 | 0,00206 | 0,00079 | 0,00030 | 0,00004 | 0,00001 | |
| 0,95 | 0,00778 | 0,00323 | 0,00135 | 0,00057 | 0,00010 | 0,00002 | |
| 1.00 | 0.01079 | 0,00489 | 0,00222 | 0,00102 | 0,00022 | 0,00005 | 0,00001 |
| 1,05 | 0,014 65 | 0.00720 | 0,00355 | 0,00176 | 0,00044 | 0,00012 | 0,00003 |
| 1,10 | 0,01950 | 0,01033 | 0,00550 | 0,00294 | 0,00085 | 0,00026 | 0,00008 |
| 1,15 | 0,02550 | 0,01448 | 0,00826 | 0,00474 | 0,00158 | 0,00054 | 0,00019 |
| 1,20 | 0,03280 | 0,01986 | 0,01208 | 0,00738 | 0,00280 | 0,00108 | 0,00043 |
| 1,25 | 0,04155 | 0,02670 | 0,01723 | 0,01117 | 0,00476 | 0,00206 | 0,00092 |
| 1,30 | 0,05189 | 0,03524 | 0,02402 | 0,01644 | 0,00780 | 0,00376 | 0,00185 |
| 1,35 | 0,06396 | 0,04571 | 0,03276 | 0,02357 | 0,01235 | 0,00656 | 0,00355 |
| 1,40 | 0,07787 | 0,05834 | 0,04379 | 0,03298 | 0,01892 | 0,01098 | 0,00649 |
| 1.45 | | 0,07333 | 0,05743 | 0,04511 | 0,02810 | 0,01769 | 0,01133 |
| 1,50 | | | 0,07398 | 0,06038 | 0,04055 | 0,02750 | 0,01893 |
| 1,55 | | | | 0,07920 | 0,05696 | 0,04131 | 0,03034 |
| 1,60 | | | | | 0,07797 | 0,06006 | 0,04675 |
| 1,70 | | | | | | | 0,09949 |

# Таблиця 7

## Графік і таблиця нормального закону розподілу




| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| 0.1 | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0753 |
| 0.2 | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| 0.3 | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| 0.4 | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| 0.5 | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| 0.6 | .2257 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2517 | .2549 |
| 0.7 | .2580 | .2611 | .2642 | .2673 | .2704 | .2734 | .2764 | .2794 | .2823 | .2852 |
| 0.8 | .2881 | .2910 | .2939 | .2967 | .2995 | .3023 | .3051 | .3078 | .3106 | .3133 |
| 0.9 | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| 1.0 | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| 1.1 | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| 1.2 | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| 1.3 | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| 1.4 | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| 1.5 | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| 1.6 | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| 1.7 | .4554 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| 1.8 | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| 1.9 | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| 2.0 | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| 2.1 | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| 2.2 | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| 2.3 | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| 2.4 | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| 2.5 | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| 2.6 | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2.7 | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| 2.8 | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| 2.9 | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| 3.0 | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |
| 3.1 | .4990 | .4991 | 4991. | .4992 | .4992 | .4992 | .4992 | .4992 | .4993 | .4993 |
| 3.2 | .4993 | .4993 | .4994 | .4994 | .4994 | .4994 | .4994 | .4995 | .4995 | .4995 |
| 3.3 | .4995 | .4995 | .4996 | .4996 | .4996 | .4996 | .4996 | .4996 | .4996 | .4997 |
| 3.4 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4998 | .4998 |
| 3.5 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 |
| 3.6 | .4998 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| 3.7 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| 3.8 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .49995 | .49995 | .49995 |
| 3.9 | .49995 | .49995 | .49996 | .49996 | .49996 | .49996 | .49996 | .49996 | .49997 | .49997 |
| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |

## СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ


1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы економетрики. – М.:ЮНИТИ, 1998. – 1022 с.
2. Берк К., Кєйри П. Анализ данных с помощью Microsoft Excel.: Пер. с англ. - М.: Изд. дом «Вильямс», 2005. – 560 с.
3. Берндт Э.Р. Практика эконометрики:классика и современность: Учебник: Пер. с англ. – М.: ЮНИТИ-ДАНА, 2005. – 863 с.
4. Вітлинський В.В. Моделювання економіки: Навч. посібник. – К.: КНЕУ, 2007. – 407 с.
5. Денисенко В.Ю., Ковальова І.Л., Молчанюк І.В., Окара Д.В., Чернишев В.Г. Методичні вказівки з курсу «Економетрія. Розділ «Лінійні моделі регресії» для студентів спеціальностей 6.030507 «Маркетинг» і 6.030601 «Менеджмент» - Одеса: ОДАБА, 2014. – 127 с.
6. Грубер Й. Економетрія: Вступ до множинної регресії та економетрії. – В 2-х т. – К.: Нічлава, 1998. Т. 1. – 384 с.; 1999. Т.2. – 308 с.
7. Доугерти К. Введение в економетрику. – М.: ИНФРА – М,1997. – 402 с.
8. Економетрія. За редакцією Кабака А.Ф., Проценко О.В. – Одеса: ТОВ «Автограф», 2003. – 561 с.
9. Жлуктенко В.І., Водянова Н.К., Савіна С.С., Колодінська О.В. Економетрія. Навчальний посібник. – К.: Вид-во Європ. ун-ту 2005. – 552 с.
10. Захарченко Н.Н. Бизнес-статистика и прогнозирование в Excel - М.: Изд.дом «Вильямс», 2004. – 208 с.
11. Здрок В.В., Лагоцький Т.Я. Економетрія: Підручник. – К.: Знання, 2010. – 541 с.
12. Корольов О.А., Рязанцева В.В. Практикум з економетрії: завдання з практичними рекомендаціями, алгоритмами та

прикладом їх наскрізного виконання. Ч.1. Регресійний аналіз: Навч.посібник. – К.: Вид-во Європ.ун-ту, 2002. – 250 с.
13. Кремер Н.Ш., Путко Б.А. Эконометрика. – М.: ЮНИТИ-ДАНА, 2002. – 311 с.
14. Кузьмичов А.І., Медведєв М.Г. Економетрія. Моделювання засобами MS Excel: Навчальний посібник. - К.: Вид-во «Ліра - К», 2011. – 214 с.
15. Левин Д. и др. Статистика для менеджеров с использованием Microsoft Excel, 4-е изд. Пер. с англ. – М.: Изд. дом «Вильямс», 2005. – 1312 с.
16. Лещинський О.Л. Економетрія: Навч. посібник / О.Л. Лещинський, В.В. Рязанцева, О.О.Юнькова. – К.: МАУП, 2003. – 208 с.
17. Лондар С.Л., Юринець Р.В. Економетрія засобами MS Excel: Навчальний посібник. - К.: Вид-во Європ.ун-ту, 2004. – 242 с.
18. Лугінін О.Є., Білоусова С.В., Білоусов О.М. Економетрія. : Навчальний посібник. - К.: Центр навчальної літератури, 2005. – 252 с.
19. Лук'яненко І.Г., Краснікова Л.І. Економетрика: Підручник. – К.: Товариство «Знання», КОО, 1998. – 494 с.
20. Лук'яненко І.Г., Краснікова Л.І. Економетрика: Практикум з використанням комп'ютера. – К.: Товариство «Знання», КОО, 1998. – 220 с.
21. Каплан А.В. Решение экономических задач на компьютере. – М.: ДМК Пресс; СПб.: Питер, 2004. – 600 с.
22. Медведєв М.Г. Економетричні методи моделювання: Навч. посібник. – К.: Вид-во Європ. ун-ту, 2003. – 140 с.
23. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник. – К.: КНЕУ, 2004. – 520 с.
24. Магнус Я.Р.,Катышев П.К., Пересецкий А.А. Эконометрика: Учеб. пособие.- М.: Дело, 1998. – 248 с.

25. Назаренко О.М. Основи економетрії. Підручник: К.: Центр навчальної літератури, 2004. – 392 с.
26. Практикум по эконометрике: Учеб. пособие / И.И.Елисеева, С.В.Курышева, Н.М.Гордиенко и др.; Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2001. – 192 с.
27. Сигел Э. Практическая бизнес-статистика.: Пер. с англ. – М.: Изд. Дом «Вильямс», 2002. -1056 с.
28. Толбатов Ю.А. Економетрика.– К: Четверта хвиля, 1997.–320 с.
29. Фишер Ф. Проблема идентификации в эконометрии. – М.: Статистика, 1978. - 224 с.
30. Чернишев В.Г., Денисенко В.Ю., Ковальова І.Л., Окара Д.В., Молчанюк І.В. Методичні вказівки до лекційного курсу «Економетрія. Частина 2» для студентів напряму 6.030504 «Економіка підприємства» - Одеса: ОДАБА, 2012. – 56 с.

**Навчальне електронне видання**

**Окара Діана Василівна**

# ЕКОНОМЕТРІЯ

**Навчальний посібник**

Підписано до публікації 16.02.2018 р.
Гарнітура Times. Зам. №18-20
Видавець і виготовлювач:
**Одеська державна академія будівництва та архітектури**
**Свідоцтво ДК № 4515 від 01.04.2013 р.**
**Україна, 65029, м. Одеса, вул. Дідріхсона, 4.**
**тел.: (048) 729-85-34, e-mail: rio@ogasa.org.ua**