ЛЬВІВСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ПРАВА, ПСИХОЛОГІЇ ТА ЕКОНОМІКИ

Олександра ВОЛОШИН
Наталія ГАЛАЙКО

ЕКОНОМЕТРІЯ

ЧАСТИНА 1

Навчальний посібник

Львів
2012

УДК 519.22
ББК 22.17
Е40

*Рекомендовано Вченою радою
Львівського державного університету внутрішніх справ
(протокол № 8 від 25 квітня 2012 р.)*

Р е ц е н з е н т и:
***І.В. Огірко**, доктор фізико-математичних наук, професор, професор кафедри інформаційно-мультимедійних технологій Української академії друкарства;*
***С.І. Фединяк**, кандидат фізико-математичних наук, доцент, доцент кафедри теорії функцій та теорії ймовірностей Львівського національного університету імені Івана Франка*

Волошин О.Р., Галайко Н.В.
Е40 Економетрія. Ч. 1: навч. посібник / О. Волошин, Н. Галайко. – Львів: Львівський державний університет внутрішніх справ, 2012. – 192 с.

ISBN 978-617-511-112-3

У навчальному посібнику розглянуто основні математико-статистичні методи побудови та аналізу економетричних моделей. Наведено приклади застосування даних моделей для дослідження різноманітних економічних процесів як на макро-, так і на мікрорівнях.

Посібник призначається студентам економічних спеціальностей вищих навчальних закладів та фахівцям у сфері економіки.

УДК 519.22
ББК 22.17

ISBN 978-617-511-112-3

ЗМІСТ

ПЕРЕДМОВА………………………………………………………. 7

Розділ I. **ЕКОНОМЕТРІЯ ЯК НАУКА**………………………… 9

Тема 1. ВСТУП ДО ЕКОНОМЕТРІЇ……………………………... 9

1.1. Поняття економетрії……………………………………….. 9

1.2. Історія виникнення і розвитку економетрії……………… 11

Запитання та завдання для самоконтролю……………………….. 13

Тема 2. ОСНОВИ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ……………………………………………….… 13

2.1. Суть і методологічні основи економетричного моделювання…… 13

2.2. Статистична база економетричних моделей……..…… 16

2.3. Етапи економетричного моделювання…………………… 18

2.4. Специфікація моделі………………………...……………… 19

Запитання та завдання для самоконтролю……………………….. 20

Розділ II. **ОСНОВИ ІНФОРМАЦІЙНОГО ТА МАТЕМАТИКО-СТАТИСТИЧНОГО ЗАБЕЗПЕЧЕННЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ**………… 22

Тема 3. ОСНОВИ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ……………………...……… 22

3.1. Завдання та об’єкти математичної статистики…………... 22

3.2. Поняття динамічних і варіаційних рядів…………………. 24

3.3. Табличне зображення статистичних даних……………… 25

3.4. Графічне зображення статистичних даних……………….. 28

3.5. Середні величини, міри розсіяння та інші числові характеристики…………………………………………... 30

3.6. Характеристики динамічного ряду……………………….. 37

Запитання та завдання для самоконтролю……………………….. 39

Тема 4. РОЗПОДІЛ ВИПАДКОВИХ ВЕЛИЧИН………………….. 43
4.1. Випадкові величини. Загальні поняття…………………… 43
4.2. Числові характеристики випадкової величини…………. 48
4.3. Деякі найпоширеніші закони розподілу………………….. 58
4.3.1. Рівномірний розподіл……………………………….. 58
4.3.2. Біноміальний розподіл……………………………….. 59
4.3.3. Розподіл Пуассона ………………………………... 59
4.3.4. Показниковий (експоненціальний) розподіл……………………………………………….. 60
4.3.5. Нормальний розподіл………………………………... 60
4.3.6. Логнормальний розподіл…………………………….. 63
Запитання та завдання для самоконтролю……………………….. 63

Тема 5. ТОЧКОВІ ТА ІНТЕРВАЛЬНІ ОЦІНКИ ПАРАМЕТРІВ……………………………………………………… 67
5.1. Точкові оцінки…………………………………………… 67
5.2. Інтервальні оцінки……………………………………… 74
Запитання та завдання для самоконтролю……………………….. 79

Тема 6. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ……………... 86
6.1. Поняття статистичної гіпотези…………………………… 86
6.2. Перевірка гіпотези про рівність центрів розподілу двох нормальних генеральних сукупностей при відомій дисперсії………….…………… 88
6.3. Перевірка гіпотези про рівність центрів розподілу двох нормально розподілених генеральних сукупностей при невідомій дисперсії……………….… 89
6.4. F-розподіл і перевірка гіпотези про рівність дисперсій двох нормальних генеральних сукупностей…………………………………………… 91
Запитання та завдання для самоконтролю……………………….. 92

Тема 7. ОСНОВИ КОРЕЛЯЦІЙНОГО АНАЛІЗУ……………….. 100
7.1. Функціональні, стохастичні (статистичні) та кореляційні залежності…………………………… 100

7.2. Коефіцієнт кореляції…………….……………………… 102
7.3. Оцінка значущості коефіцієнта кореляції………………... 104
7.4. Розкладання дисперсій. Кореляційне відношення. Коефіцієнт детермінації. Стандартна похибка оцінки……………………………………………….…… 106
7.5. Багатовимірний кореляційний аналіз…..………………… 113
Запитання та завдання для самоконтролю………………………….. 114

***Розділ ІІІ.* КЛАСИЧНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ. ОСНОВИ РЕГРЕСІЙНОГО АНАЛІЗУ**……………... 126
Тема 8. ПАРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ………………………... 126
8.1. Модель парної лінійної регресії…………………...… 126
8.2. Метод найменших квадратів…………………….….. 129
8.3 Передумови застосування регресійного аналізу………… 130
8.4. Незміщеність і точність оцінок коефіцієнтів регресії……………………………………………………… 133
8.5. Перевірка значущості оцінок коефіцієнтів регресії……………………………………………………… 135
8.6. Довірчі інтервали для коефіцієнтів регресії b_0 і b_1……… 136
8.7. Перевірка значущості рівняння регресії в цілому. Прогнозування………………………………………… 138
Запитання та завдання для самоконтролю………………………….. 144

Тема 9. МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ…………………. 148
9.1. Класична нормальна модель множинної регресії……………………………………………………… 148
9.2. Оцінювання параметрів багатофакторної регресії методом найменших квадратів……………………………. 153
9.3. Коваріаційна матриця оцінок параметрів регресійної моделі…………………………………………………….. 157
9.4. Значущість оцінок параметрів множинної регресії……………………………………………………… 159

9.5. Коефіцієнт множинної кореляції та детермінації. Скоригований коефіцієнт детермінації. Оцінка значущості рівняння множинної регресії………………………………………………………. 161
9.6. Прогнозування за множинною регресійною моделлю….. 166
9.7. Методи побудови множинної регресійної моделі…………………………………………………….. 169
Запитання та завдання для самоконтролю………………………….. 172

ЛІТЕРАТУРА………………………………………………………………. 178

ДОДАТКИ……………………………………………………………….. 180

ПЕРЕДМОВА

Поряд з макроекономікою і мікроекономікою економетрія є однією з найважливіших дисциплін фундаментальної підготовки бакалаврів з економіки підприємства та менеджменту.

Економетрія – це галузь економічної науки, яка вивчає методи кількісного вимірювання взаємозв'язків між економічними показниками [9, с. 3].

Предметом вивчення дисципліни «Економетрія» є економіко-математичні методи та засоби для дослідження економічних явищ і процесів, що відбуваються на макро- і мікрорівнях.

«Економетрія» викладається після вивчення студентами блоку дисциплін «Вища математика», «Дослідження операцій», «Теорія ймовірностей та математична статистика», пов'язує дисципліни математичного циклу з економічними науками, передує вивченню професійно орієнтованих дисциплін, становить основу для проведення економічних досліджень.

Основною *метою* вивчення економетрії є формування у майбутніх економістів та менеджерів сучасного економічного мислення та спеціальних знань з використанням системного аналізу, різних методів економетричного аналізу як складової підтримки прийняття рішень щодо економічних об'єктів різної складності, ієрархії та організації.

Основними завданнями, що мають бути вирішені у процесі вивчення дисципліни, є формування у студентів певних умінь, а саме:

- правильно задати специфікацію економетричної моделі;
- обчислити оцінки її параметрів;
- оцінити якість самої моделі;
- будувати багатофакторні лінійні економетричні моделі;

– визначати мультиколінеарність, гетероскедастичність і автокореляцію в економетричних моделях, застосовувати способи їх усунення;

– використовувати узагальнений метод найменших квадратів та інші спеціальні економетричні методи;

– використовувати прикладні програми під час проведення розрахунків на ПЕОМ та розробки практичних рекомендацій щодо прийняття рішень.

Вивчення економетрії сприяє розвитку математичної культури, економічного мислення, формуванню аналітичних навичок, кращому оволодінню сучасними інформаційними технологіями, що дозволяє якісно проводити економічні дослідження.

Розділ I

ЕКОНОМЕТРІЯ ЯК НАУКА

Тема 1. ВСТУП ДО ЕКОНОМЕТРІЇ

1. *Суть економетрії як наукової дисципліни.*
2. *Альтернативні підходи до визначення економетрії як науки.*
3. *Історія виникнення і розвитку економетрії.*

1.1. Поняття економетрії

Економетрія як наукова дисципліна виникла у 30-х роках XX століття. Так, у 1930 р. у США було засновано «економетричне товариство», яке визначило себе як «Міжнародне товариство для розвитку економічної теорії у зв'язку зі статистикою і математикою».

Сам термін «економетрія» (econometrics) складається з двох грецьких слів: «економія», тобто управління господарством, господарювання, і «метрон» – вимірювання. Отже, у буквальному перекладі термін «економетрія» означає «вимірювання в економіці».

Існують дві версії щодо походження терміна «економетрія». Згідно з першою термін «економетрія» був запропонований представником Львівсько-Варшавської логіко-філософської наукової школи П. Чомпою в 1910 р. Згідно з другою версією назву «економетрія» було введено у 1926 р. відомим норвезьким вченим, Нобелівським лауреатом Рагнаром Фрішем (R. Frisch).

Нині відсутнє єдине загальновизнане визначення економетрії. Більше того, серед українських та російських учених немає єдності думок щодо самого терміна. Так, С.А. Айвазян і В.С. Мхітарян [1], Н.М. Кремер і Б.А. Путко [5], І.Г. Лук'яненко та Л.І. Краснікова [6], А.Н. Мардас [8], Ю.А. Толбатов [13] та ряд інших дослідників застосовують термін «економетрика». Водночас В.І. Єлейко [3], С.І. Наконечний, Т.О. Терещенко і Т.П. Романюк [9], С.В. Ржевський [12] та інші вчені вживають термін «Економетрія». У галузевих стандартах Міністерства освіти і науки України з підготовки бакалаврів за напрямами «Економіка і підприємництво» та «Менеджмент» назву дисципліни визначено як «Економетрія». Тому надалі саме цей термін ми і будемо застосовувати.

У західній літературі подано таке найбільш узагальнене визначення *економетрії,* зокрема Л. Клейном, Г. Тінтнером, Г. Хансеном, С. Лізером, А. Класом (Klein L.R., Tintner G., Hansen G., Leser C., Klas A.): *економетрія* – це дисципліна, основне завдання якої полягає у встановленні й вимірюванні кількісних взаємозв'язків між економічними показниками за допомогою статистичних методів.

Вітчизняні підручники та стандарти освіти по-різному визначають предмет економетрії. Більшістю авторів у вітчизняній літературі *економетрія* визначається як наука, яка вивчає конкретні кількісні закономірності та взаємозв'язки економічних об'єктів і процесів за допомогою математико-статистичних методів та моделей.

Отже, *економетрія* – це синтез математики, економіки і статистики.

Економетричні моделі є важливим інструментом прогнозування як на макроекономічному рівні для урядових установ щодо зміни податків, цін тощо, так і на мікроекономічному рівні для прийняття рішень конкретними фірмами, підприємствами.

Вивчення економетрії вимагає добрих знань з вищої математики, зокрема, знання диференціального числення, матричної алгебри, теорії визначників, а також найважливіших понять теорії ймовірностей. Для оволодіння економетрією необхідно також мати уявлення про основні поняття математичної статистики, такі як: випадкова змінна, стандартна помилка, рівень значущості, теорія статистичного оцінювання і перевірки гіпотез, кореляційний і регресійний аналіз.

Необхідні також добрі знання основних показників і категорій макроекономіки, мікроекономіки, економічної статистики й економіки галузі.

Останнім часом мова економіки стає все більше мовою математики, а економіка перетворюється в одну з найбільш математизованих наук. Досягнення сучасної економічної науки висувають нові вимоги до вищої професійної економічної освіти. Якщо в період централізованої планової економіки основна увага приділялася побудові оптимізаційних моделей народного господарства, галузей і підприємств, то в умовах ринкової економіки суттєво підвищується роль економетричних методів. Саме за допомогою цих методів, які враховують імовірнісний характер ринкових економічних процесів, можливі проведення економічного аналізу та розробка достовірних прогнозів у банківській справі, фінансах, різноманітних сферах бізнесу.

1.2. Історія виникнення і розвитку економетрії

На початку XX століття були зроблені перші спроби складання так званих «барометрів розвитку», найвідоміший з яких – «гарвардський барометр», за допомогою якого намагались передбачити поведінку товарного і грошового ринків. Однак цей барометр не зумів передбачити світову кризу 1929 р. Тим не менше, гарвардська школа вважалась на той час центром економічних досліджень.

Саме тут уперше почали вивчати ряди економічних показників у сукупності, досліджувати їх тенденції і циклічний рух. Криза 1929–1933 років спонукала до критичного перегляду методів аналізу. У дослідженнях почали враховуватися випадкові аспекти економічних явищ. Засновниками економетрії вважаються Р. Фріш і Я. Тінберген (Tinbergen J.).

У 1928 р. були опубліковані роботи Ч. Кобба і П. Дугласа (Cobb C. і Douglas P.) про виробничі функції. У 30-ті роки розглядалися тільки деякі моделі попиту і пропозиції. Лише після Другої світової війни з’явились комплексні економетричні моделі на макрорівні. Тоді ж відбуваються і перші спроби побудови мікроекономічних моделей.

Подальша розробка економетричних моделей здійснювалась у двох напрямках. *Перший* пов’язаний з динамізацією моделей. Такий підхід є характерним для *голландської школи,* засновником якої був професор Я. Тінберген. *Для другого напрямку* характерною є часова і галузева деталізація. Цей напрямок розробляється *американською школою* під керівництвом професорів П. Клейна і А. Гольдбергера.

На жаль, впродовж тривалого часу в СРСР економетрія вважалася «буржуазною псевдонаукою», що завдало значної шкоди розвитку вітчизняної економічної науки. Лише на початку 1960-х років роботи у галузі економетрії розпочались в СРСР, Польщі і Чехословаччині.

У 70-х і 80-х роках минулого століття певний внесок у розвиток економетрії зробили радянські, в т.ч. українські вчені. Зокрема, А. Ємельянов і Ф. Кушнірський розробили економетричну модель розвитку народного господарства УРСР (УКР-2). Ця модель складається зі 101 регресійного рівняння, які об’єднані у 7 взаємопов’язаних блоків.

Сьогодні економетричні моделі є одним із найпоширеніших і водночас найдієвіших наукових інструментів, які використовуються в економічних дослідженнях.

Упродовж двох останніх десятиліть економетрія як наукова дисципліна бурхливо розвивається, що особливо спостерігається у США, Німеччині, Англії. Зростає кількість наукових публікацій і досліджень із застосуванням економетричних методів. Видаються і перевидаються десятки підручників, спеціалізовані економетричні журнали, проводяться наукові конференції.

Економетрія стала одним з основних наукових методів аналізу і прогнозування цінних паперів на фондових ринках, що ще більше сприяло її поширенню серед науковців і фінансових аналітиків.

Широкому впровадженню і розвитку економетричних методів сприяла поява персональних комп’ютерів та сучасних інформаційних технологій. Сучасні економетричні комп’ютерні пакети дозволяють автоматизувати найбільш трудомістку роботу з розрахунку таблиць, графіків, статистик, характеристик, оцінок. Спеціалісту ж з економетрії залишається найбільш важлива і творча робота – економічна постановка задачі, вибір змінних до відповідної моделі, оцінювання та інтерпретація результатів розрахунків.

Свідченням всесвітнього визнання економетрії є присудження Нобелівських премій з економіки таким видатним вченим: Р. Фрішу і Я. Тінбергену (1969 р.), Л. Клейну (1980 р.), Т. Хаавельмо (1989 р.), Дж. Хекману і Д. Макфаддену (2000 р.).

Запитання та завдання для самоконтролю

1. Дайте визначення предмета курсу «економетрія», основної мети та завдань вивчення дисципліни.

2. Що означає термін «економетрія», кого вважають його автором?

3. Дайте визначення економетрії як наукової дисципліни у західній і вітчизняній літературі.

4. Охарактеризуйте роль і місце курсу в підготовці бакалаврів економічних спеціальностей.

5. Коротко опишіть історію виникнення і розвитку економетрії як наукової дисципліни.

6. Охарактеризуйте основні сфери застосування економетрії в бізнесі та економіці.

7. Назвіть видатних учених у галузі економетрії – Нобелівських лауреатів з економіки.

Тема 2. ОСНОВИ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ

1. *Суть і методологічні основи економетричного моделювання.*
2. *Статистична база економетричних моделей.*
3. *Етапи економетричного моделювання.*
4. *Специфікація економетричної моделі.*

2.1. Суть і методологічні основи економетричного моделювання

Суть економетричного моделювання можна продемонструвати на такому прикладі. Нехай ми хочемо продати автомобіль і розмістити відповідне оголошення про продаж у газеті «Ваш магазин» та журналі «Автосалон». При цьому виникає запитання: яку ціну вказати в оголошенні?

Очевидно, що при визначенні ціни необхідно керуватись інформацією про ціни на аналогічні автомобілі на вторинному ринку.

В даному випадку аналогічні автомобілі – це автомобілі тієї ж самої марки і моделі (наприклад, автомобілі марки Honda, моделі Civic) з близькими значеннями таких параметрів, як рік випуску і пробіг у тис. км. Вивчивши оголошення у згаданих виданнях за декілька останніх тижнів (місяців), ми можемо сформувати власну думку про стан вторинного ринку, а отже, визначити ціну.

Наше завдання полягає у визначенні ціни продажу автомобіля, тобто величини, яка формується під впливом деяких факторів (у першу чергу – року випуску і пробігу). Перша величина називається *залежною,* або *пояснюваною*, а фактори, від яких вона залежить, – *незалежними,* або *пояснювальними* змінними. Вивчаючи стан вторинного автомобільного ринку, ми можемо одержати очікуване значення залежної змінної при заданих значеннях пояснювальних змінних.

Зауважимо також, що *дані спостережень*, тобто ціни на аналогічні автомобілі на ринку, залежать і від інших *випадкових факторів:* сезону (у весняний період попит зростає, торгівля пожвавлюється, ціни можуть зростати, у зимовий – навпаки), терміну продажу автомобіля і потреби продавця в грошах, перебування автомобіля у ДТП тощо.

Загальним моментом для будь-якої економетричної моделі є *розбиття залежної змінної на дві частини – пояснену і випадкову.* Тоді задача економетричного моделювання може бути сформульована так: *на підставі експериментальних даних визначити пояснену частину і, розглядаючи випадкову складову як випадкову величину, одержати оцінки параметрів її розподілу.*

Отже, економетрична модель має такий загальний вигляд:

$$\mathrm{Y} = f\,(X) + \varepsilon, \qquad (2.1)$$

де Y – значення залежної змінної, яке спостерігається; X – пояснена частина, яка залежить від значень пояснювальних змінних; ε – випадкова (стохастична) складова.

Припустимо, що в результаті дослідження одержано таку залежність для поясненої частини змінної Y – ціни автомобіля (при цьому змінні будемо позначати великими буквами, а їх значення – малими):

$$Y = 17000 - 800x_1 - 0{,}4x_2, \qquad (2.2)$$

де Y – очікувана ціна автомобіля в доларах США; x_1 – термін експлуатації автомобіля (років); x_2 – пробіг (тис. км); 17000 – ціна в доларах США нового автомобіля аналогічної моделі і марки.

Розглянемо на даному прикладі *цілі* економетричного моделювання.

Одержаний результат, по-перше, дозволяє зрозуміти, як саме формується дана економічна змінна, тобто ціна автомобіля. По-друге, він дозволяє виявити вплив кожної з пояснювальних змінних на ціну автомобіля. Так, при збільшенні терміну експлуатації на один рік ціна у середньому зменшується на *$800,* а при збільшенні пробігу на одну тис. км – на *$0,4.* По-третє, одержаний результат дозволяє *прогнозувати* ціну автомобіля, якщо відомі його основні параметри.

Розглянемо тепер суть економетричного моделювання у загальному вигляді. Нехай є p пояснювальних змінних $X_1, X_2,\ldots, X_p$ і залежна змінна Y.

Змінна величина Y є *випадковою* величиною, яка має при заданих значеннях незалежних змінних деякий розподіл. Якщо випадкова величина Y неперервна, то можна вважати, що її розподіл при кожному допустимому наборі значень змінних $(x_1, x_2,\ldots, x_p)$ має умовну щільність $f\, x_1, x_2, \ldots, x_p\, (y)$.

Як правило, робиться припущення щодо характеру розподілу Y.

Найчастіше припускають, що умовні розподіли Y при кожному допустимому значенні незалежних змінних є *нормальними.* Саме таке припущення дозволяє одержати найкращі результати.

Пояснювальні змінні X_J $(J = 1,2\ \ldots,\ P)$ можуть бути як *випадковими,* так *детермінованими,* тобто такими, які приймають певні значення.

Класична економетрична модель розглядає пояснювальні змінні X_J як *детерміновані,* натомість основні результати статистичного дослідження моделі залишаються значною мірою такими ж самими, як і у випадках, якщо X_J вважаються *випадковими* змінними.

Пояснена частина економетричної моделі (Y_e) – це функція від значень незалежних факторів, тобто пояснювальних змінних:

$$Y_e = f\,(x_1, x_2, \ldots, x_p) + \varepsilon. \qquad (2.3)$$

Найбільш природним вибором поясненої частини випадкової величини Y є її середнє значення – умовне математичне сподівання $M\, x_1, x_2, \ldots, x_p$ (Y), одержане за даного набору пояснювальних змінних ($x_1, x_2, \ldots, x_p$). Надалі математичне сподівання будемо позначати $M_x(Y)$. Рівняння $M_x(Y) = =f(x_1, x_2, \ldots, x_p)$ називається *рівнянням регресії.*

За такого вибору поясненої частини економетрична модель має вигляд:

$$Y = M_x(Y) + \varepsilon, \qquad (2.4)$$

де ε – випадкова складова, яка називається *похибкою* або *збуренням.*

Необхідно зауважити, що економетрична модель не обов’язково є регресійною, тобто пояснена частина не завжди є умовним математичним сподіванням залежної змінної. Розглянемо деякі властивості регресійної моделі. Візьмемо від обидвох частин рівняння (2.4) математичне сподівання при заданому наборі значень пояснювальних змінних X. У даному випадку M_x(Y) є числовою величиною, яка дорівнює своєму математичному сподіванню, і ми одержимо рівність

$$M_x(\varepsilon) = 0, \qquad (2.5)$$

а отже, і $M(\varepsilon) = 0$, тобто, у регресійній моделі очікуване значення випадкової помилки дорівнює нулю. Доведено, що звідси випливає некорельованість випадкових похибок і пояснювальних змінних X (якщо останні розглядаються як випадкові величини). Ця обставина є найбільш істотною умовою *обґрунтованості* кількісних результатів аналізу економетричної моделі.

2.2. Статистична база економетричних моделей

Щоб одержати достовірні та інформативні дані про розподіл будь-якої випадкової величини, необхідно мати вибірку її спостережень досить великого обсягу. Отже, вибірки спостережень залежної змінної Y і пояснювальних змінних X_j $(j = 1, 2, ..., p)$ є основою економетричного дослідження. Такі вибірки – це набори значень $(x_{j1}, x_{j2},\ldots, x_{jp},y_i)$, де $i = 1, 2, \ldots, n;$ n – кількість спостережень.

В економетрії розглядається два типи вибіркових даних.

Перший тип – це *просторова вибірка*, або *просторові дані.* В економіці під просторовою вибіркою розуміють набір економічних змінних, одержаних у певний момент часу. Для економетрії таке визначення є незручним у зв'язку із невизначеністю поняття «момент часу». Тому просторовою вибіркою вважають спостереження, які одержані у приблизно незмінних умовах, тобто вони є набором незалежних вибіркових даних з відповідної генеральної сукупності.

Отже, *просторовою* вибіркою називається серія з *n незалежних* спостережень ($p+1$) – мірної випадкової величини $(x_1,\ldots, x_p; Y)$. У цьому разі різні випадкові величини Y_i є незалежними між собою. Звідси випливає некорельованість їх збурень, тобто

$$r(\varepsilon_i,\varepsilon_j) = 0 \quad при\ i = j, \qquad (2.6)$$

де $r(\varepsilon_i,\varepsilon_j)$ – коефіцієнт кореляції між збуреннями ε_i та ε_j.

Економетрична модель, побудована на основі просторової вибірки експериментальних даних (x_i, y_i), має вигляд:

$$y_i = f(x_i) + \varepsilon_i,\ i = 1, \ldots, n, \qquad (2.7)$$

де похибки регресії задовольняють умовам

$$M(\varepsilon_i) = 0, \qquad (2.8)$$
$$r(\varepsilon_i,\varepsilon_j) = 0, \qquad (2.9)$$
$$D(\varepsilon_i) = \sigma_i^2, \qquad (2.10)$$

де $D(\varepsilon_i)$ – дисперсія збурення. Умова (2.10) означає, що дисперсія збурення є постійною для будь-якого *i.*

Другий тип вибіркових даних – *це часовий ряд.* Часовим рядом називається вибірка спостережень, у якій важливими є не лише значення випадкових величин, що спостерігаються, але й порядок їх слідування один за одним. Найчастіше впорядкованість обумовлена тим, що експериментальні дані є серією спостережень однієї й тієї ж випадкової величини у послідовні моменти часу. При цьому припускається, що тип розподілу випадкової величини, що спостерігається, залишається тим самим (наприклад, нормальним), але параметри його змінюються залежно від часу.

Моделі часових рядів, як правило, є складнішими від моделей просторових вибірок.

Це пояснюється тим, що спостереження у випадку часового ряду *є залежними,* тому похибки регресії можуть корелювати одна з одною, тобто умова (2.6) не виконується. Це значно ускладнює статистичний аналіз моделі.

2.3. Етапи економетричного моделювання

Можна виокремити *п'ять основних етапів (кроків)* економетричного моделювання: постановка задачі; параметризація моделі; формування інформаційної бази; ідентифікація моделі; верифікація моделі.

Крок 1. Постановка задачі. Формується мета дослідження, відбираються економічні змінні моделі.

Цілями економетричного моделювання можуть бути: аналіз економічного об'єкта або процесу; прогнозування економічних показників; імітація розвитку об'єкта; підготовка управлінського рішення.

При виборі змінних необхідне економічне обґрунтування кожної змінної. Пояснювальні змінні не повинні бути пов'язані між собою функціональною або тісною кореляційною залежністю, тому що це може призвести до неможливості оцінки параметрів моделі.

Для вибору змінних моделі можуть бути використані спеціальні методи, наприклад, покроковий регресійний аналіз та деякі інші. Натомість визначальним при включенні у модель тих чи інших змінних є аналіз економічної суті об'єкта (процесу), який досліджується.

Крок 2. Параметризація моделі. Здійснюється безпосередньо моделювання, тобто вибір функції f(x) економетричної моделі (2.1).

Важливою проблемою на даному і попередньому етапах є *специфікація* моделі, тобто запис у математичній формі економічних зв'язків і співвідношень; виявлення складу *екзогенних* (тих, що задаються ззовні) і *ендогенних* (тих, що формуються всередині об'єкта) змінних, а також *лагових* змінних (тих, що взяті у попередні моменти часу); формування вихідних передумов та обмежень моделі.

Крок 3. Формування інформаційної бази. Здійснюється збір відповідної статистичної інформації, тобто економічних змінних:

$$(x_{i1}, x_{i2}, \ldots x_{ip}, y_i)\ i = 1, 2, \ldots, n.$$

Крок 4. Ідентифікація моделі. Здійснюється *статистичний аналіз моделі* та *оцінка її параметрів*. Реалізації цього кроку присвячена значна частина курсу «Економетрії».

Крок 5. Верифікація моделі. Проводиться *перевірка істинності, адекватності моделі.* З'ясовується, як розв'язані проблеми специфікації та ідентифікації моделі, точність розрахунків за даною моделлю, якою мірою побудована модель відповідає реальному економічному об'єкту або процесу.

Необхідно зауважити, що розглянутий поділ економетричного моделювання на етапи (кроки) є умовним. Розглянуті етапи можуть перетинатись та взаємно доповнювати один одного.

2.4. Специфікація моделі

Специфікація моделі – це аналітична форма економетричної моделі. На підставі досвіду економетричних досліджень і теоретичного аналізу взаємозв'язків між економічними показниками, основними функціями, які описують такі взаємозв'язки, є [9, с. 86]:

1) лінійна функція

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m; \qquad (2.11)$$

2) степенева функція

$$y = b_0 x_1^{\ b1} x_2^{\ b_2} \ldots x_m^{\ bm}; \qquad (2.12)$$

3) гіперболічна функція

$$y = b_0 + \frac{b_1}{x_1} + \frac{b_2}{x_2} + \frac{b_m}{x_m}; \qquad (2.13)$$

4) квадратична функція

$$y = b_0 + b_1 x_1^{\ 2} + b_2 x_2^{\ 2} + \ldots + b_m x_m^{\ 2}. \qquad (2.14)$$

За допомогою нескладних перетворень нелінійні функції можна звести до лінійного виду:

1) для степеневої функції

$$\ln y = \ln b_0 + b_1 \ln x_1 + b_2 \ln x_2 + ... + b_m \ln x_m; \qquad (2.15)$$

2) для гіперболічної функції

$$y = b_0 + b_1 z_1 + b_2 z_2 + ... + b_m z_m, \text{zj} = \frac{1}{x_j}; \qquad (2.16)$$

3) для квадратичної функції

$$y = b_0 + b_1 t_1 + b_2 t_2 + ... + b_m t_m, \; t_j = x_j^{\,2}. \qquad (2.17)$$

Лінійні моделі є найбільш простими і надійними, тому вивчення можливості лінеаризації є однією з основних задач економетричного моделювання.

Вибір аналітичної форми або специфікації економетричної моделі передбачає відбір конкретних незалежних змінних. Можлива ситуація, коли вид функції та її складові не відповідають (або недостатньо повно відповідають) реальним економічним процесам. Тоді можуть мати місце *помилки специфікації моделі.*

Можливі *три види* помилок специфікації моделі:

1) ігнорування при побудові моделі істотної пояснювальної змінної;

2) введення в модель незалежної змінної, яка не є істотною для взаємозв'язків, що досліджуються;

3) використання невідповідних математичних функцій.

Правила перевірки побудованої моделі будуть розглянуті у наступних темах.

Запитання та завдання для самоконтролю

1. *У чому полягає задача економетричного моделювання?*
2. *Що таке пояснювальні та пояснювані змінні?*
3. *З яких частин складається економетрична модель?*
4. *Що таке випадкова складова економетричної моделі?*

5. Якими можуть бути пояснювальні змінні?

6. Задайте загальне формулювання економетричної моделі.

7. Як сформулювати загальну економетричну модель за допомогою рівняння регресії?

8. Яка основна умова обґрунтованості результатів аналізу економетричної моделі?

9. Які типи вибіркових даних використовуються при економетричному моделюванні?

10. Що таке просторова вибірка, які її властивості?

11. Сформулюйте економетричну модель, побудовану на основі просторової вибірки.

12. Що таке часовий ряд? У чому полягає складність моделей часових рядів?

13. Назвіть етапи економетричного моделювання та поясніть їх зміст.

14. Що таке специфікація економетричної моделі?

15. Які основні функції описують взаємозв'язки між економічними показниками?

16. Що таке помилка специфікації моделі?

Розділ II

ОСНОВИ ІНФОРМАЦІЙНОГО ТА МАТЕМАТИКО-СТАТИСТИЧНОГО ЗАБЕЗПЕЧЕННЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ

У цьому розділі подано короткий огляд основних понять математичної статистики і теорії ймовірностей, які складають основу математично-статистичного інструментарію економетрії. Головна мета розділу – нагадати найважливіші поняття. Відповідним чином підготовлений читач може відразу перейти до вивчення тем наступного розділу.

Тема 3. ОСНОВИ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ

1. *Завдання та об'єкти математичної статистики.*
2. *Сутність динамічних і варіаційних рядів.*
3. *Основи табличного і графічного зображення статистичних даних.*
4. *Основні числові характеристики вибірки (середні величини, міри розсіяння та ін.).*
5. *Основні характеристики динамічного ряду.*

3.1. Завдання та об'єкти математичної статистики

Математична статистика – це система методів збору, обробки, аналізу та інтерпретації результатів експериментів.

Предмет математичної статистики – визначення загальних характеристик, закономірностей явищ і процесів на основі збору й аналізу даних.

Завдання математичної статистики можна поділити на дві великі *групи,* кожна з яких виконує свої функції.

Перша функція – систематизація і опис результатів спостережень (*описова функція* статистики).

Друга – оцінка отриманих спостережень або зібраних даних. Це функція здійснення *статистичних висновків,* або *пояснювальна функція.*

Для реалізації цієї функції слід розв'язати два класи задач:

– з'ясувати питання про точність результатів на основі методів *теорії статистичного оцінювання;*

– оцінити відповідність одержаних даних зробленим припущенням (гіпотезам).

Відповідь на ці питання дає *теорія перевірки статистичних гіпотез.*

Об'єктом дослідження математичної статистики є *масові випадкові явища.* Кількісно такі явища характеризуються *випадковими величинами.*

Сукупність усіх можливих значень випадкової величини називається *генеральною сукупністю.*

На практиці дослідник економічних явищ і процесів майже ніколи не має справу з генеральною сукупністю (винятки становлять, наприклад, загальнодержавні переписи населення, які проводяться не частіше, ніж раз на 10 років), оскільки її надзвичайно складно або неможливо дослідити й опрацювати.

Вибірка, випадкова вибірка, статистична вибірка – це сукупність значень випадкової величини, отриманих у результаті збору даних, вимірювань, спостережень, експериментів.

Одна з найважливіших задач математичної статистики полягає в тому, щоб на підставі вивчення вибірки одержати можливість оцінити (з певною довірчою ймовірністю або допустимим рівнем похибки) характеристики генеральної сукупності.

Для розв'язання даної задачі розроблені спеціальні методи і прийоми, які називаються *статистичною процедурою,* що передбачає ряд послідовних кроків:

– збір статистичних даних, тобто одержання вибірки;

– попередній аналіз вибірки (описова статистика);

– визначення статистичних оцінок сукупності (точкових та інтервальних);

– перевірку статистичних гіпотез;
– кореляційний аналіз;
– регресійний аналіз.

Під час збору інформації необхідно забезпечити *однорідність, достовірність* та *незміщеність* вибіркових даних.

3.2. Поняття динамічних і варіаційних рядів

Спостереження за економічними процесами (явищами) з метою їх математичного моделювання передбачає деяку кількість вимірювання значень окремих показників цих процесів. При цьому можуть бути отримані два типи даних: динамічні та варіаційні ряди. Ці ряди складають основу інформаційної бази економетричних моделей.

Динамічний ряд – це послідовність спостережень за економічним явищем або процесом у рівновіддалені проміжки часу, які обчислюються за однією і тією ж методикою.

Нехай x_i – значення деякого показника економічного процесу в *і-й* проміжок часу. Тоді, обчислюючи значення цього показника через рівновіддалені проміжки часу за однаковою методикою, отримаємо динамічний ряд:

$$x_1, x_2, \ldots, x_i, \ldots x_n . \qquad (3.1)$$

Значення x_i називається *і-м рівнем* динамічного ряду. Для того, щоб динамічні ряди можна було використовувати як інформаційну базу економетричних моделей, необхідно, щоб усі її рівні можна було *порівняти* між собою. Так, усі рівні динамічного ряду повинні мати однакові одиниці виміру, враховувати зміни в часі, фактор інфляції тощо. Крім того, дані про рівні динамічного ряду повинні бути *достовірними.*

Лише за дотримання таких умов динамічні ряди можна використати як інформаційну базу економетричних моделей.

Другий тип вхідної інформації економетричних моделей – це варіаційні ряди.

Варіаційний ряд у найбільш загальному визначенні – це послідовність значень спостережуваної величини, впорядкованих у порядку зростання або спадання.

Прикладами варіаційних рядів можуть бути:

– дані про тарифні розряди робітників дільниці (цеху, підприємства);

– дані про середню заробітну платню працівників різних галузей економіки;

– дані про середню заробітну платню у різних регіонах країни тощо.

Нехай маємо варіаційний ряд (3.1), де x_i – числа, які показують варіацію ознаки, що вивчається, *i – номер варіанти.* Отже, число x_i, що знаходиться на *i*-му місці послідовності (3.1), називається *i-ю варіантою.*

Якщо всі варіанти $\{x_i\}$ впорядкувати у порядку зростання або спадання, то ми одержимо *ранжований варіаційний ряд.*

Дані статистичних спостережень зображають у вигляді *таблиць* і *графіків.*

3.3. Табличне зображення статистичних даних

Для забезпечення зручності та наочності статистичні дані зображають у вигляді таблиць.

Приклад. Дослідник отримав такі дані про тарифні розряди робітників дільниці складання меблів: *5,1,4, 5, 4, 3, 5, 5, 2, 5, 5, 6, 4, 3, 1, 5, 2, 5, 6, 5.*

Для вивчення результатів ці дані необхідно згрупувати. Проранжуємо значення у порядку зростання. У результаті отримаємо *ранжований статистичний ряд:*

$$\underbrace{1,1}_{2\,рази}\ \underbrace{2,2}_{2\,рази}\ \underbrace{3,3}_{2\,рази}\ \underbrace{4,4,4}_{3\,рази}\ \underbrace{5,5,5,5,5,5,5,5,5}_{9\,разів}\ \underbrace{6,6}_{2\,рази}.$$

Величини, які показують, скільки разів зустрічаються ті чи інші значення варіант, називаються *частотами* варіант m_x. Якщо n – обсяг вибірки, то

$$\sum_x m_X = n. \qquad (3.2)$$

Відношення частоти до обсягу вибірки називається *відносною частотою* w_x

$$w_x = \frac{m_x}{n} = \frac{m_x}{\sum_x m_x}. \qquad (3.3)$$

Крім частоти і відносної частоти, у статистиці розглядають поняття нагромадженої частоти $m_x^н$ і нагромадженої відносної частоти $w_x^н$, які показують кількість спостережень, що менше або дорівнюють заданому:

$$m_x^н = \sum_{x \le x} m_x; \qquad (3.4)$$

$$w_x^н = \frac{m_x^н}{n}. \qquad (3.5)$$

Значення величин m_x, w_x, $m_x^н$ та $w_x^н$ для варіаційного ряду, який характеризує тарифні розряди, зображені у табл. 1.

Таблиця, яка показує розподіл частот або відносних частот між варіантами, називається *дискретним варіаційним рядом.*

У розглянутому прикладі ознака приймає лише цілі значення. Саме тому даний ранжований варіаційний ряд є *дискретним.* У багатьох інших випадках економічні показники змінюються неперервно або вважаються неперервними.

У таких випадках будується *інтервальний* варіаційний ряд. При цьому весь діапазон зміни ознаки поділяється на інтервали і визначаються *інтервальні частоти* m_h – кількість випадків ознаки, які потрапляють у даний інтервал, та *інтервальні відносні частоти.*

Відповідно визначаються *нагромаджені інтервальні частоти* і *відносні частоти.*

Таблиця 1

Значення величин дискретного варіаційного ряду

Значення ознаки, x	Частота, m_x	Відносна частота, w_x	Нагромаджена частота, $m_x^н$	Нагромаджена відносна частота, $w_x^н$
1	2	0,10	2	0,10
2	2	0,10	4	0,20
3	2	0,10	6	0,30
4	3	0,15	9	0,45
5	9	0,45	18	0,90
6	2	0,10	20	1,00
$\sum$	20	1,00	-	-

Інтервальний варіаційний ряд будується двома способами:

1) на основі спостережень за ознакою, що неперервно змінюється, тобто дослідним шляхом;

2) шляхом визначення оптимальної величини інтервала h за формулою Стерджеса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n}, \qquad (3.6)$$

де $x_{\max}$ та $x_{\min}$ – відповідно найбільше та найменше значення ознаки у вибірці.

Початком першого інтервалу вважається і приймається величина

$$a_1 = x_{\min} - \frac{1}{2} h, \qquad (3.7)$$

наступні інтервали

$$a_2 = a_1 + h\ , a_3 = a_2 + h, \ldots, a_k = a_{k-1} + h. \qquad (3.8)$$

До кожного інтервалу включаються дані, що є більшими від нижньої межі інтервалу або меншими чи рівними верхній межі інтервалу, або навпаки.

3.4. Графічне зображення статистичних даних

На підставі результатів ранжування варіаційні ряди зображаються графічно у вигляді *полігона* частот або відносних частот для дискретних рядів, *гістограми* частот або відносних частот для інтервальних рядів, графіка нагромаджених частот або відносних частот *(кумулятивних кривих).*

Полігон використовується, як правило, для зображення *дискретного* варіаційного ряду. Для його побудови у системі координат $(x;\ m_x)$ або $(x;\ w_x)$ наносимо точки з відповідними координатами. Точки з'єднуються лініями (рис. 1) (дані з табл. 1).

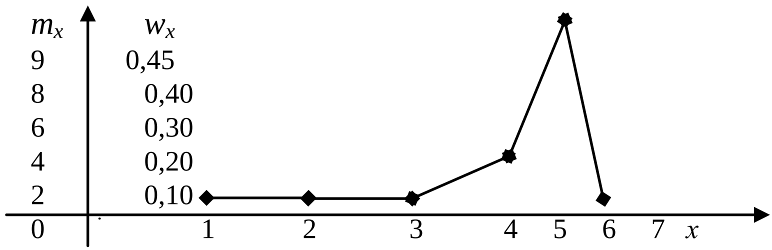


Рис. 1. Полігон частот і відносних частот дискретного варіаційного ряду

Гістограма – це діаграма, яка зображає *інтервальний* варіаційний ряд. При цьому частота і відносна частота зображаються не точкою, а прямою, яка паралельна до осі абсцис (рис. 2).

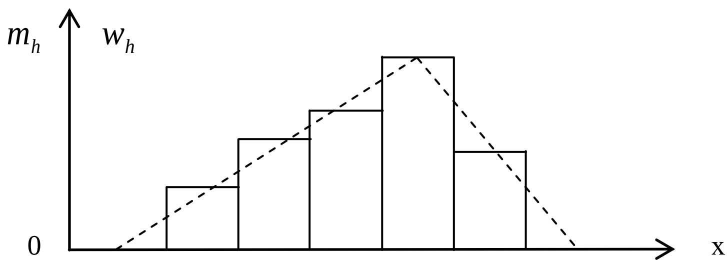


Рис. 2. Гістограма і полігон частот варіаційного ряду

Іноді інтервальний ряд зображається у вигляді полігону. Тоді значення частоти або відносної частоти відносять до середини інтервалу (пунктирна лінія, рис. 2).

Кумулятивна крива – це крива нагромаджених частот (відносних частот).

Вона будується так: по осі абсцис відкладається значення *x* для дискретного ряду та інтервалу – для інтервального ряду; по осі ординат відкладаються нагромаджені частоти або відносні частоти.

Верхнім межам інтервалів відповідають нагромаджені частоти (відносні частоти) (див. рис. 3, нижня крива).

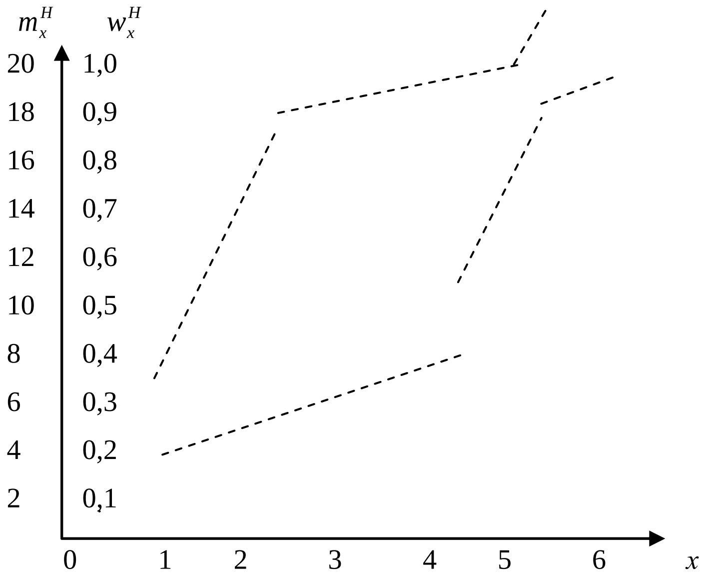


Рис. 3. Кумулятивна крива і огива

При зміні осей кумулятивна крива називається *огивою* (рис. 3, верхня крива), дані з табл. 1.

Варіаційний ряд у табличному і графічному зображеннях є значно інформативнішим, ніж неорганізована вибірка, і дає достатнє первинне уявлення про характеристики явища або процесу, який досліджується.

3.5. Середні величини, міри розсіяння та інші числові характеристики

Числові характеристики отримують на основі вибіркових даних. При цьому припускається, що такі характеристики по суті дають уявлення про значення генеральної сукупності. Оскільки дослідник майже ніколи не має інформації про генеральну сукупність, то і значення характеристик, що отримуються з вибірки, можуть певною мірою відрізнятися від істинних значень характеристик. Зважаючи на цю обставину, в статистиці здебільшого говорять не про самі характеристики, а про їх вибіркові оцінки.

У практиці статистичних досліджень найчастіше використовуються такі числові характеристики, як середні величини, математичне сподівання, міри розсіяння та моменти більш високих порядків.

Середні величини – це характеристики, які узагальнено зображають вибірку одним числом. Найчастіше використовуються середнє арифметичне, середнє геометричне, мода і медіана.

Середнє арифметичне. Найбільш загальна формула:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad (3.9)$$

де x_i – значення випадкової величини, одержане в i-спостереженні; n – обсяг вибірки.

Для вибірок невеликих обсягів формула (3.9) є досить зручною при розрахунку середнього арифметичного. Для вибірок великих обсягів зручніше користуватися характеристиками варіаційного ряду:

для дискретних рядів

$$\bar{x} = \frac{\sum_{x} x m_x}{n} = \sum_{x} x w_x; \qquad (3.10)$$

для інтервальних рядів

$$\bar{x} = \frac{\sum_h x_h m_h}{n} = \sum_h x_h w_h . \qquad (3.11)$$

Середнє геометричне:

$$x_{геом} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} . \qquad (3.12)$$

Мода (M_0 *або* $\hat{x}$ *)* – це значення ознаки, яке спостерігалось у вибірці найбільшу кількість разів (йому відповідає максимум частот або відносних частот варіаційного ряду).

Для дискретного ряду мода дорівнює варіанті з найбільшою частотою. Значення $\hat{x}$ визначається безпосередньо з таблиці або за полігоном частот (відносних частот). Так, для табл. 1 $M_0 = \hat{x} = 5$.

Для інтервального ряду спочатку визначається модальний інтервал a_0 (інтервал, який відповідає найбільшій частоті ознаки), потім частота (відносна частота) модального *(*m_0 *або* w_0), попереднього *(*m_0' *або* w_0') і наступного *(*m_0'' *або* w_0'') інтервалу. Тоді

$$M_0 = \hat{x} = a_0 + h\frac{m_0 - m_0'}{2m_0 - m_0' - m_0''} , \qquad (3.13)$$

або

$$M_0 = \hat{x} = a_0 + h\frac{w_0 - w_0'}{2w_0 - w_0' - w_0''} . \qquad (3.14)$$

Медіана (M_e або $\tilde{x}$ *)* – це значення ознаки, для якої половина всіх спостережень менше від цієї ознаки, тобто це серединне значення ознаки. Значення медіани припадає на середину варіаційного ряду.

Найпростіше медіану можна знайти за графіком кумулятивної кривої, визначаючи x, що відповідає $w_x^n = 0{,}5$, або найближчого до нього цілого числа для дискретного ряду. Так, для рис. 3 значенням $w_x^n = 0{,}5$ і $w_x^n = 12$ відповідає $x = 4{,}8$, тоді $M_e = \tilde{x} = 5$.

Крім того, при знаходженні медіани можливі два випадки:

– кількість членів ряду непарна *(n = 2k+1);*

– кількість членів ряду парна *(n = 2k).*

У першому випадку

$$M_e = x_{k+1}, \tag{3.15}$$

де x_{k+1} – значення *(k+1)-го* члена варіаційного ряду.

У другому випадку

$$M_e = \frac{x_k + x_{k+1}}{2}, \tag{3.16}$$

де x_k, x_{k+1} – значення відповідно *k-ї* та (*k*+1)*-ї* варіанти.

Математичним сподіванням, або *середнім значенням M* (*X*) дискретної випадкової величини *X* називається сума добутків усіх значень та ймовірності, що їм відповідає:

$$M(X) = \sum_{i=1}^{n} x_i p_i , \tag{3.17}$$

де x_i – значення випадкової величини;

p_i – ймовірність того, що випадкова величина прийме дане значення.

Властивості математичного сподівання:

1) $M(C) = C$, де C – константа;

2) $M(kX) = kM(X)$;

3) $M(X \pm Y) = M(X) \pm M(Y)$;

4) $M(X\ Y) = M(X) \cdot M(Y)$, де X, Y – незалежні випадкові величини;

5) $M(X \pm C) = M(X) \pm C$;

6) $M(X - \mu) = 0$, де $\mu = M(X)$.

Середні величини, відображаючи варіаційний ряд єдиним числом, не характеризують *мінливості* значень ознаки, яка спостерігається, тобто варіацію.

Тому необхідні інші міри, які б характеризували розсіяння даних навколо центра.

Найпростішим показником варіації є *варіаційний розмах:*

$$R_B = x_{\max} - x_{\min}. \qquad (3.18)$$

Найбільшого поширення отримали міри розсіяння спостережень навколо середніх величин, зокрема *центральні моменти.*

Центральний момент *другого* порядку – це *дисперсія.*

Дисперсією D(X) випадкової величини *X*, або її *теоретичною дисперсією,* називається математичне сподівання квадрата її відхилення від математичного сподівання:

$$D(X) = M\left[X - M(X)\right]^2. \qquad (3.19)$$

Теоретична дисперсія випадкової величини *X* позначається також як *pop.var(x)* або σ_x^2, і якщо зрозуміло, яка змінна розглядається, то нижній індекс опускається (σ^2).

Дисперсія характеризує *відхилення (розсіяння, варіацію)* значень випадкової величини навколо середнього значення.

Якщо випадкова величина *X* є дискретною із скінченною кількістю значень, то

$$D(X) = \sigma_x^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p_i. \qquad (3.20)$$

Дисперсія *D(X)* має розмірність квадрата випадкової величини, що не завжди зручно. Зокрема, дисперсія і середнє арифметичне мають різні розмірності, що ускладнює аналіз.

Тому як міра розсіяння використовується також величина $\sqrt{\sigma_x^2}$.

Середнім квадратичним відхиленням (або *стандартним відхиленням*) σ^2 випадкової величини називається арифметичне значення кореня квадратного з її дисперсії:

$$\sigma_x = \sqrt{D(X)}. \qquad (3.21)$$

Зауважимо, що величина σ_x є *теоретичним стандартним відхиленням.*

Як зазначалось вище, дослідник майже ніколи не має справу з генеральною сукупністю випадкової величини X, а може досліджувати на основі випадкової обмеженої вибірки $(x_1, x_2, ..., x_n)$, *де* n – обсяг вибірки. Тому на практиці може бути розрахована *емпірична (вибіркова) дисперсія,* або *оцінка дисперсії,* яку будемо позначати S^2.

Для неранжованих рядів

$$S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}, \tag{3.22}$$

для інтервальних ранжованих рядів

$$S_x^2 = \frac{\sum_h (x_h - \bar{x})^2 m_h}{n} = \sum_h (x_h - \bar{x})^2 w_h. \tag{3.23}$$

Відповідно *оцінка стандартного відхилення (емпіричне* або в*ибіркове стандартне відхилення)* визначається за формулами:

для неранжованих рядів

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}, \tag{3.24}$$

для інтервальних ранжованих рядів

$$S = \sqrt{\sum_h (x_h - \bar{x})^2 w_h}. \tag{3.25}$$

Третій центральний момент (або *асиметрія*) A_s визначається за формулами:

$$A_s = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n} \tag{3.26}$$

або

$$A_S = \sum_h (x_h - \bar{x})^3 w_h. \qquad (3.27)$$

Коефіцієнт асиметрії γ_1 – це відношення центрального моменту 3-го порядку до куба стандартного відхилення*:*

$$\gamma_1 = \frac{A_S}{S^3}. \qquad (3.28)$$

Величини A_s і γ_1 характеризують симетричність полігону і гістограми відносно середнього арифметичного. Так, при $A_S < 0$ і $\gamma_1 < 0$ середнє арифметичне зсунуте ліворуч відносно моди $\hat{x}$ (лівостороння асиметрія).

При $A_S > 0$ і $\gamma_1 > 0$ – навпаки, асиметрія правостороння (див. рис. 4).

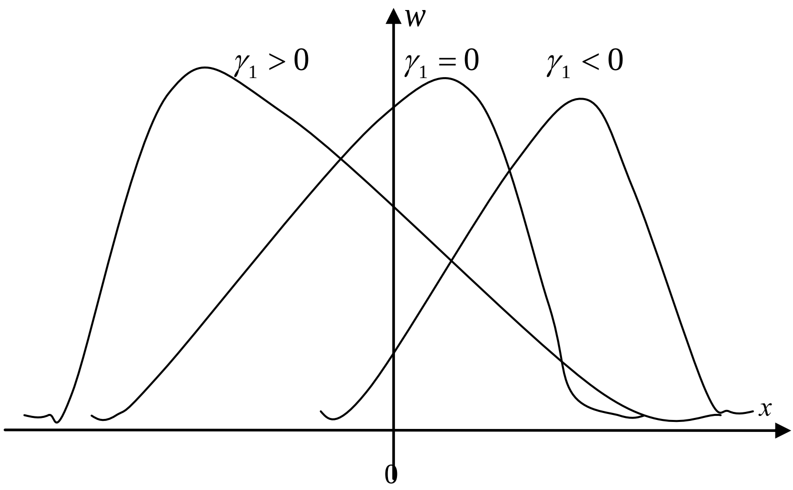


Рис. 4. Характер «геометрії» розподілу залежно від A_s і γ_1

Четвертий центральний момент (або ексцес) ε_x визначається за формулами:

$$\varepsilon_x = \sum_{i=1}^{n} (x_1 - \bar{x})^4 / n \tag{3.29}$$

або

$$\varepsilon_x = \sum_{h} (x_h - \bar{x})^4 w_h. \tag{3.30}$$

Коефіцієнт ексцесу

$$\gamma_2 = \frac{\varepsilon_x}{S^4} - 3. \tag{3.31}$$

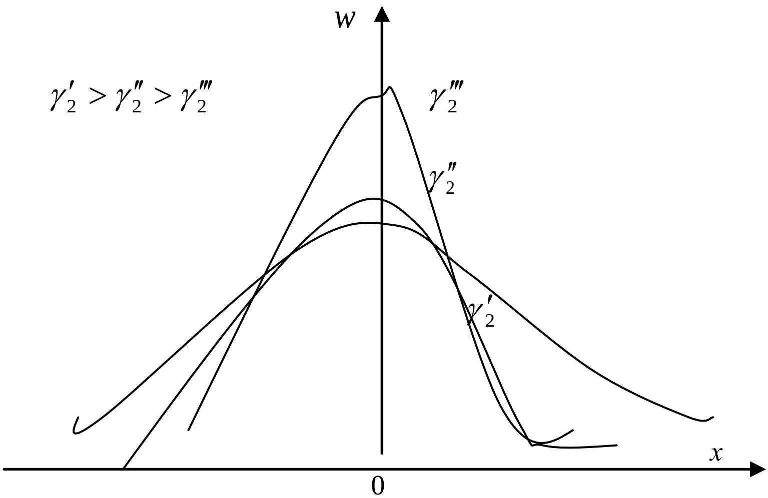


Рис. 5. Характер «геометрії» вершини розподілу залежно від ε_x і γ_2

Величини ε_x і γ_2 і характеризують гостроту вершини полігону гістограми. Чим більше значення цих величин, тим гострішою є вершина графіка розподілу (рис. 5).

Для *нормального* закону розподілу характерні такі співвідношення між числовими характеристиками:

$$\bar{x} = \hat{x} = \tilde{x}, \ \gamma_1 = 0, \ \varepsilon_x = 3S^4, \ \gamma_2 = 0.$$

Мірою взаємозв'язку між *двома* випадковими величинами є к*оваріація.* Якщо X і Y – випадкові величини, то *теоретична коваріація* σ_{xy} визначається як математичне сподівання добутку відхилень цих величин від їх середніх значень:

$$\sigma_{xy} = M\left[(X - \mu_x)\cdot(Y - \mu_y)\right], \qquad (3.32)$$

де μ_x та μ_y – теоретичні середні значення (математичні сподівання) випадкових величин X та Y відповідно.

Теоретична коваріація позначається також як *pop.cov(x.y).*

Якщо теоретична коваріація невідома, то для її оцінки може бути використана *вибіркова коваріація cov(x,y),* яка обчислюється за рядом спостережень.

Якщо є n спостережень двох змінних X та Y, то вибіркова коваріація між X та Y визначається за формулою

$$\mathrm{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}). \qquad (3.33)$$

3.6. Характеристики динамічного ряду

Основними характеристиками динамічного ряду є середня хронологічна, середній абсолютний приріст, середній коефіцієнт (темп) зростання, середній коефіцієнт (темп) приросту, дисперсія, середнє квадратичне відхилення та коефіцієнт варіації.

Середня хронологічна $\bar{x}_{xp}$ характеризує середнє значення рівнів динамічного ряду (3.1) і обчислюється за формулою

$$\bar{x}_{xp} = \sum_{i=1}^{n} \frac{x_i}{n}, \qquad (3.34)$$

де x_i – i-й рівень динамічного ряду; n – кількість рівнів.

Середні хронологічні використовуються для порівняльного аналізу двох або кількох динамічних рядів (порівняння середнього рівня заробітної платні працівників різних галузей економіки, різних регіонів країни; порівняння середніх рівнів урожайності зернових культур у різних областях тощо).

Середній абсолютний приріст $\overline{\Delta x}$ характеризує величину зміни кінцевого рівня динамічного ряду x_n відносно його початкового рівня x_1 і визначається за формулою

$$\overline{\Delta x} = \frac{x_n - x_1}{n-1}. \qquad (3.35)$$

Середній коефіцієнт зростання $\overline{k}_{зр}$ характеризує середню швидкість зміни динамічного ряду і визначається за формулою

$$\overline{\kappa}_{зр} = \sqrt[n-1]{\frac{x_n}{x_1}}. \qquad (3.36)$$

Середній коефіцієнт приросту $k_{пр}$ визначається за формулою

$$\overline{\kappa}_{пр} = \sqrt[n-1]{\frac{x_n}{x_1}} - 1. \qquad (3.37)$$

Середній коефіцієнт зростання та середній коефіцієнт приросту, виражені у відсотках, називаються, відповідно, *темпом зростання* $\overline{T}_{зр}$ та *середнім темпом приросту* $\overline{T}_{пр}$ і визначаються за формулою

$$\overline{T}_{зр} = \overline{\kappa}_{зр} \cdot 100\%. \qquad (3.38)$$

Відхилення від середнього рівня динамічного ряду характеризує його *дисперсія,* яка визначається за формулою (3.22).

Для порівняння дисперсії з середніми характеристиками динамічного ряду використовується його *середнє квадратичне відхилення* (формула (3.24)).

Ступінь коливання рівнів динамічного ряду у відсотках характеризується *коефіцієнтом варіації* V_x, який визначається за формулою

$$V_x = \frac{S}{\bar{\bar{x}}_{xp}} \cdot 100\%, \qquad (3.39)$$

де S – середнє квадратичне відхилення.

Запитання та завдання для самоконтролю

1. Дайте визначення предмета й основних завдань математичної статистики. У чому полягає статистична процедура?
2. Дайте визначення динамічного та варіаційного рядів.
3. Охарактеризуйте показники дискретного варіаційного ряду: частота, відносна частота, нагромаджена частота і відносна частота.
4. Як будується інтервальний варіаційний ряд?
5. Що таке полігон частот, гістограма частот, кумулятивна крива, огива? Як вони будуються?
6. Охарактеризуйте основні середні величини варіаційного ряду (середнє арифметичне, середнє геометричне, мода, медіана).
7. Що таке математичне сподівання, які його властивості?
8. Що таке теоретична дисперсія та теоретичне стандартне відхилення випадкової величини?
9. Чим оцінки дисперсії та стандартного відхилення випадкової величини відрізняються від їх теоретичних значень?
10. Що характеризують показники асиметрії?
11. Що характеризують показники ексцесу?
12. Дайте визначення теоретичної та вибіркової коваріації.
13. Охарактеризуйте основні характеристики динамічного ряду.
14. Запишіть вибірку 4, 2, 10, 3, 5, 4, 10, 7, 3, 2, 4, 3, 5, 2 у вигляді: варіаційного ряду; статистичного ряду частот; статистичного ряду відносних частот.
15. Запишіть емпіричну функцію розподілу та побудуйте її графік для вибірки, поданої у вигляді частотної таблиці:

x_i	2	5	7	8
n_i	1	3	2	8

16. Знайдіть емпіричну функцію розподілу та побудуйте її графік; побудуйте гістограму та полігон частот і відносних частот для вибірки, заданої у вигляді таблиці:

Інтервал	[-3;-2)	[-2;-1)	[-1;0)	[0;1)	[1;2)	[2;3)	[3;4)	[4;5)
n_i	3	10	15	24	25	13	7	3

17. Складіть частотну таблицю розподілу за частотами букв українського алфавіту у висловлюванні американського математика А. Нівена: «Математику не можна вивчати, спостерігаючи, як це робить сусід».

18. Нехай маємо таку частотну таблицю (статистичний розподіл частот):

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Запишіть емпіричну функцію розподілу та побудуйте її графік.

19. Запишіть вибірку 5, 3, 7, 10, 5, 5, 2, 10, 7, 2, 7, 7, 4, 2, 4 у вигляді варіаційного ряду, статистичного ряду частот та відносних частот. Запишіть емпіричну функцію розподілу та побудуйте її графік.

20. Запишіть вибірку 17, 15, 17, 20, 16, 17, 20, 14, 15, 20, 17, 15 у вигляді варіаційного ряду; складіть статистичний ряд частот і статистичний ряд відносних частот; знайдіть емпіричну функцію розподілу та побудуйте її графік.

21. Сім вимірювань відношення маси Землі до маси Місяця, одержані сімома різними космічними кораблями, є такими:

81,3001; 81,3015; 81,3006; 81,3011; 81,2997; 81,3005; 81,3021.

Для даного статистичного матеріалу запишіть варіаційний ряд. Знайдіть медіану, середнє арифметичне і розмах.

22. За даними таблиці про врожайність пшениці на різних ділянках посівної площі запишіть функцію розподілу та побудуйте частотну криву, полігон та гістограму частот:

Врожайність, ц/га	20-25	25-30	30-35	35-40	40-45	45-50
Частка ділянки від загальної площі, %	5	10	33	21	25	6

23. Обчисліть вибіркове середнє, моду, медіану, варіанту, стандарт, розмах для вибірки, заданої частотною таблицею:

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

24. На олімпіаді учні одержали такі бали:

Номер учня за списком	1	2	3	4	5	6	7	8	9	10	11	12
Кількість балів	1	2	1	3	2	3	3	12	11	7	12	10

Складіть частотну таблицю, побудуйте полігон частот, визначте міри центральних тенденцій.

25. Дано вибірку: -40; -40; 0; 0; 1; 1; 2; 8; 9; 9. Знайдіть моду, медіану, середнє значення та середньоквадратичне відхилення.

26. Знайдіть числові характеристики (вибіркове середнє, вибіркову дисперсію, вибіркове середньоквадратичне відхилення, стандарт, моду, медіану, розмах) для вибірки:

$[x_i, x_{i+1})$	[4, 6)	[6, 8)	[8, 10)	[10, 14)	[14, 16)
n_i	5	3	2	8	2

27. Знайдіть числові характеристики (вибіркове середнє, вибіркову дисперсію, вибіркове середньоквадратичне відхилення, стандарт, моду, медіану, розмах) для вибірки:

$[x_i, x_{i+1})$	[3, 5)	[5, 7)	[7, 11)	[11, 13)	[13, 15)
n_i	4	3	6	3	4

28. Знайдіть числові характеристики (вибіркове середнє, вибіркову дисперсію, вибіркове середньоквадратичне відхилення, стандарт, моду, медіану, розмах) для вибірки:

$[x_i, x_{i+1})$	[4, 8)	[8, 12)	[12, 16)	[16, 20)	[20, 24)
n_i	2	5	6	8	4

29. У таблиці наведено вибірку середньомісячної платні ста співробітників фірми. Впорядкуйте вибірку. Запишіть розподіл частоти та відносної частоти середньомісячної платні співробітників фірми; згрупований розподіл частоти та відносної частоти середньомісячної платні співробітників фірми; згрупований розподіл накопиченої частоти та накопиченої відносної частоти середньомісячної платні співробітників фірми; згрупований розподіл щільності частоти та щільності відносної частоти. Побудуйте гістограму та полігон частот та відносних частот, полігон накопиченої частоти та емпіричну функцію розподілу середньомісячної платні співробітників фірми.

Вибірка середньомісячної платні 100 співробітників фірми									
338	348	304	314	326	314	324	304	342	308
336	304	302	338	314	304	320	321	322	321
312	323	336	324	312	312	364	356	362	302
322	310	334	292	362	381	304	366	298	304
381	368	304	298	368	290	340	328	316	322
302	314	292	342	321	322	290	332	298	296
296	298	324	338	352	326	318	304	332	322
360	312	331	331	304	316	332	282	342	338
342	322	324	325	302	328	354	330	316	324
334	350	334	324	332	340	324	314	326	323

Тема 4. РОЗПОДІЛ ВИПАДКОВИХ ВЕЛИЧИН

1. Закони розподілу випадкової величини.
2. Функція розподілу та щільність розподілу випадкової величини.
3. Деякі закони розподілу випадкових величин, які найчастіше використовуються в економетрії.

4.1. Випадкові величини. Загальні поняття

Поняття випадкової величини є одним із найважливіших у теорії ймовірностей.

Випадкова величина – це змінна, яка в результаті випробувань приймає одне з можливих своїх значень, наперед не відоме.

Строго *випадкова величина X* визначається як функція, що задана на множині елементарних наслідків або в просторі елементарних подій, тобто

$$X = f(\omega), \qquad (4.1)$$

де ω – елементарний наслідок (результат) або елементарна подія, яка належить до Ω,тобто $\omega \in \Omega$.

Для *дискретної* випадкової величини множина Θ можливих значень випадкової величини, тобто функції $f(\omega)$, є *скінченною* і *зчисленною,* для *неперервної – нескінченною* і *незчисленною.* Наприклад, кількість дорожньо-транспортних пригод в обласному центрі протягом тижня – це дискретна випадкова величина; віддаль польоту куль при стрільбі з автоматичної зброї – неперервна випадкова величина.

Розподілом випадкової величини називають будь-яке співвідношення між можливими значеннями випадкової величини і ймовірностями, що відповідають їм.

Форма запису таких співвідношень називається *функцією* або *законом розподілу.*

Для дискретної випадкової величини закон розподілу задається у вигляді таблиці, формули (аналітично) або графіка.

Закон розподілу випадкової величини X у вигляді таблиці називається *рядом розподілу* дискретної випадкової величини.

x_1	x_2	…	x_1	…	x_n
p_1	p_2	…	p_1	…	p_n

Якщо по осі абсцис відкладати значення випадкової величини, а по осі ординат – імовірності, що їм відповідають, то ламана, яку отримуємо шляхом з'єднання точок, називається *многокутником* або *полігоном розподілу ймовірностей.*

Для будь-якої дискретної випадкової величини

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1. \quad (4.2)$$

Дві випадкові величини називаються *незалежними,* якщо закон розподілу однієї з них залежить від того, яких можливих значень набула друга величина.

Закони розподілу задаються у *двох формах: інтегральній та диференціальній.*

Інтегральним законом або *функцією розподілу* випадкової величини X називається функція $F(x)$ ймовірності того, що випадкова величина X набуде значення, яке є менше від деякої невипадкової величини x:

$$F(x) = P(X < x). \quad (4.3)$$

Властивості функції розподілу випадкової величини (ФРВВ)

1. ФРВВ є невід'ємною функцією, значення якої знаходяться між нулем і одиницею:

$$0 \le F(x) \le 1. \quad (4.4)$$

2. ФРВВ є неспадною функцією на всій числовій осі, тобто якщо $x_2 > x_l$, то

$$F(x_2)>F(x_l). \quad (4.5)$$

3. $$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, \qquad (4.6)$$

$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1. \qquad (4.7)$$

4. Імовірність попадання випадкової величини X в інтервал $[x_1, x_2)$ дорівнює приросту її функції розподілу на даному інтервалі, тобто

$$P(x_1 \le X < x_2) = F(x_2) - F(x_1). \qquad (4.8)$$

Випадкова величина називається *неперервною*, якщо її функція розподілу є неперервною в будь-якій точці і диференційована скрізь, крім, можливо, окремих точок.

Для дискретних випадкових величин *інтегральний* закон – це ступінчаста лінія з розривами в точках, де випадкова величина приймає конкретні значення (рис. 6).

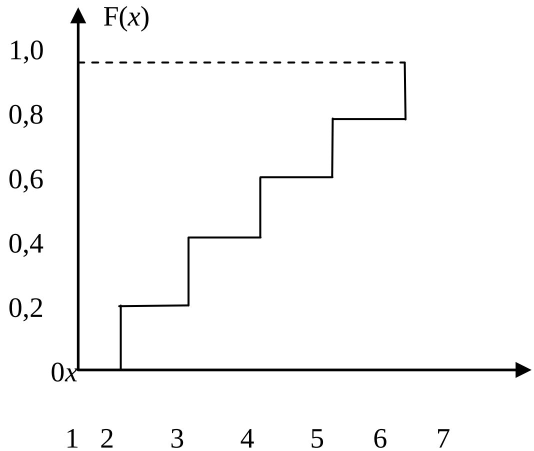


Рис. 6. Інтегральний закон розподілу дискретної випадкової величини

Для *неперервних* випадкових величин інтегральний закон – це «плавна» крива (рис. 7).

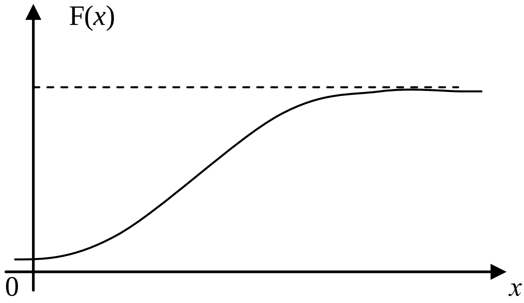


Рис. 7. Інтегральний закон розподілу неперервної випадкової величини

Диференціальний закон розподілу $f(x)$ (*щільність* імовірності) графічно зображається у вигляді *кривої розподілу* (рис. 8).

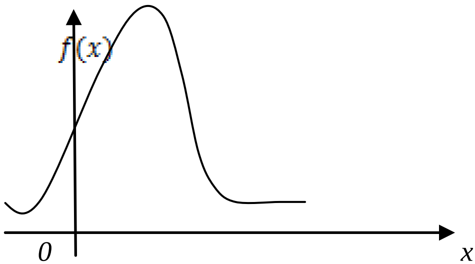


Рис. 8. Диференціальний закон розподілу (крива розподілу)

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

$$f(x) = F'(x) \tag{4.9}$$

Щільність імовірності $f(x)$, як і функція розподілу $F(x)$, є однією з двох форм закону розподілу, але на відміну від функції розподілу вона існує лише для *неперервних* випадкових величин.

Властивості щільності розподілу ймовірностей неперервної випадкової величини:

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

2. Ймовірність попадання неперервної випадкової величини в інтервал $[a_1, a_2]$ дорівнює визначеному інтегралу від її щільності ймовірності в межах від a_1 до a_2 (рис. 9), тобто

$$P(a_1 \leq X \leq a_2) = \int_{a_1}^{a_2} f(x)dx. \qquad (4.10)$$

3. Функція розподілу неперервної випадкової величини (рис. 10) може бути виражена через щільність імовірності:

$$F(x) = \int_{-\infty}^{x} f(x)dx. \qquad (4.11)$$

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

$$\int_{-\infty}^{+\infty} f(x)dx = 1. \qquad (4.12)$$

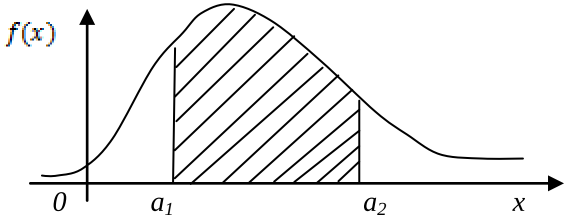


Рис. 9. Ймовірність попадання неперервної випадкової величини в інтервал $[a_1, a_2]$

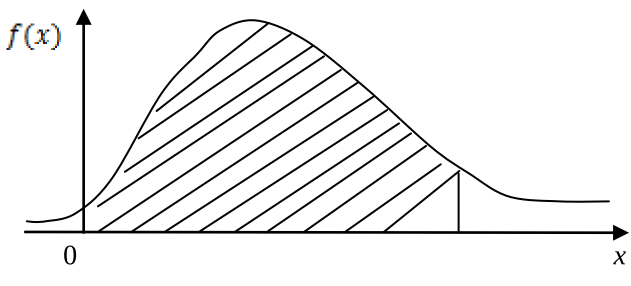


Рис. 10. Вираження функції розподілу неперервної випадкової величини через щільність імовірності

Властивості щільності ймовірності означають, що крива розподілу лежить вище від осі абсцис, а повна площа фігури, обмеженої кривою розподілу і віссю абсцис, дорівнює одиниці.

4.2. Числові характеристики випадкової величини

Закон розподілу ймовірностей як для дискретних, так і для неперервних випадкових величин дає повну інформацію про них. Проте на практиці немає потреби так докладно описувати ці величини, а достатньо знати лише певні параметри, що характеризують їх істотні ознаки. Ці параметри і називають *числовими характеристиками випадкових величин*.

Однією з найчастіше застосовуваних на практиці характеристик є *математичне сподівання*. Термін «математичне сподівання» випадкової величини X є синонімом поняття «середнє значення» випадкової величини X.

Математичним сподіванням випадкової величини X, визначеної на дискретному просторі Ω, є величина

$$M(X)=\sum_{i=1}^{\infty} x_i p_i \,. \tag{4.13}$$

Якщо Ω – обмежена множина, то

$$M(X)=\sum_{s=1}^{n} x_i p_i \,. \tag{4.14}$$

Якщо простір Ω є неперервним, то математичним сподіванням неперервної випадкової величини X є величина

$$M(X)=\int_{\Omega} x\ f(x)dx\,. \tag{4.15}$$

Якщо $\Omega = (-\infty; \infty)$, то

$$M(X)=\int_{-\infty}^{\infty} x\ f(x)dx\,. \tag{4.16}$$

Якщо Ω = [*a; b*], то

$$M(X) = \int_{a}^{в} x\ f(x)dx. \qquad (4.17)$$

Властивості математичного сподівання

1. Математичне сподівання від сталої величини *C* дорівнює самій сталій:

$$M(C) = C.$$

Справді, сталу *C* можна розглядати як випадкову величину, що з імовірністю, яка дорівнює одиниці, набуває значення *C*, а тому $M(C) = C \cdot 1 = C$.

2. $M(CX) = CM(X)$.

Для дискретної випадкової величини маємо

$$M(CX) = \sum_{i=1}^{k} Cx_i p_i = C\sum_{i=1}^{k} x_i p_i = CM(X),$$

для неперервної

$$M(CX) = \int_{-\infty}^{\infty} Cxf\ (x)\,dx = C\int_{-\infty}^{\infty} xf\ (x)\,dx = CM\ (X).$$

3. Якщо *A* і *B* є сталими величинами, то

$$M(AX + B) = AM(X) + B;$$

для дискретної випадкової величини

$$M(AX + B) = \sum_{i=1}^{n} (Ax_i + B)p_i = A\sum_{i=1}^{n} x_i p_i + B\sum_{i=1}^{n} p_i = AM(X) + B,$$

для неперервної випадкової величини

$$M(AX + B) = \int_{-\infty}^{\infty} (Ax + B)f(x)dx = A\int_{-\infty}^{\infty} x\ f(x)dx + B\int_{-\infty}^{\infty} f(x)dx = AM(X) + B.$$

<u>Приклад 1</u>. Закон розподілу дискретної випадкової величини задано у вигляді таблиці:

x_i	-6	-4	2	4	6	8
p_i	0,1	0,1	0,2	0,3	0,1	0,2

Обчислити $M(X)$.

Розв'язання. Скориставшись (4.14), дістанемо

$$M(X)=\sum_{i=1}^{6}x_i p_i = x_1p_1+x_2p_2+x_3p_3+x_4p_4+x_5p_5+x_6p_6 =$$
$$=-6\cdot 0{,}1-4\cdot 0{,}1+2\cdot 0{,}2+4\cdot 0{,}3+6\cdot 0{,}1+8\cdot 0{,}2=$$
$$=-0{,}6-0{,}4+0{,}4+1{,}2+0{,}6+1{,}6=2{,}8.$$

<u>Приклад 2.</u> За заданою щільністю ймовірностей

$$f(x)=\begin{cases}0, & x\le -1;\\ \frac{1}{12}\sqrt[3]{x+1}, & -1<x\le 7;\\ 0, & x>7\end{cases}$$

обчислити $M(X)$.

Розв'язання. Згідно із (4.17) маємо:

$$M(X)=\int_{-1}^{7}x\,f(x)dx=\int_{-1}^{7}x\frac{1}{12}\sqrt[3]{x+1}\,dx=\frac{1}{12}\int_{-1}^{7}x\sqrt[3]{x+1}\,dx=$$
$$=\left|\begin{matrix}x+1=z^3 & \\ x=z^3-1 \;\rightarrow & -1\le x\le 7\\ dx=3z^2dz & 0\le z\le 2\end{matrix}\right|=\frac{1}{12}\int_{0}^{2}(z^3-1)\;z\;3z^2dz=$$
$$=\frac{1}{4}\int_{0}^{2}(z^6-z^3)dz=\frac{1}{4}\left(\int_{0}^{2}z^6dz-\int_{0}^{2}z^3dz\right)=$$
$$=\frac{1}{4}\left(\left.\frac{z^7}{7}\right|_0^2-\left.\frac{z^4}{4}\right|_0^2\right)=\frac{1}{4}\left(\frac{128}{7}-4\right)=\frac{1}{4}\;\frac{128-28}{7}=\frac{100}{28}=\frac{25}{7};$$
$$M(X)=\frac{25}{7}.$$

Приклад 3. За заданою функцією розподілу ймовірностей

$$F(x) = \begin{cases} 0, & x \le -4; \\ \dfrac{\sqrt{x+4}}{3}, & -4 < x \le 6; \\ 1, & x > 6 \end{cases}$$

обчислити $M(X)$.

Розв'язання. Для обчислення $M(X)$ знаходимо щільність імовірностей

$$f(x) = \begin{cases} 0, & x < -4; \\ \dfrac{1}{6\sqrt{x+4}}, & -4 < x \le 6; \\ 0, & x > 6. \end{cases}$$

Тоді:

$$M(X) = \int_{-4}^{6} x \ f(x) dx = \int_{-4}^{6} x \frac{1}{6\sqrt{x+3}} dx = \frac{1}{6}\int_{-4}^{6} \frac{x}{\sqrt{x+3}} \ dx =$$

$$= \left| \begin{array}{lc} x+3 = z^2 & \\ x = z^2 - 3 \quad \rightarrow & \begin{array}{c} -3 \le x \le 6 \\ 0 \le z \le 3 \end{array} \\ dx = 2zdz & \end{array} \right| = \frac{1}{6}\int_{0}^{3} \frac{z^2-3}{z} 2zdz =$$

$$= \frac{1}{3} \ \int_{0}^{3} (z^2 - 3) dz = \frac{1}{3}\left(\int_{0}^{3} z^2 dz - 3 \ \int_{0}^{3} dz \right) =$$

$$= \frac{1}{3}\left(\left.\frac{z^3}{3}\right|_0^3 - \left.3z\right|_0^3 \right) = \frac{1}{3}(9-9) = 0;$$

$$M(X) = 0.$$

Якщо випадкова величина $X \in [a; b]$, то $M(X) \in [a; b]$, тобто математичне сподівання випадкової величини має обов'язково міститися всередині інтервалу $[a; b]$, являючи собою центр розподілу цієї величини.

*Модою (М*о) дискретної випадкової величини X називають те її можливе значення, якому відповідає найбільша ймовірність появи.

Модою для неперервної випадкової величини X називають те її можливе значення, якому відповідає максимальне значення щільності ймовірності:

$$f(\text{Mo}) = \max.$$

Якщо випадкова величина має одну моду, то такий розподіл імовірностей називають *одномодальним*; якщо розподіл має дві моди – *двомодальним* і т. ін. Існують і такі розподіли, які не мають моди. Їх називають *антимодальними*.

Медіаною (Me) неперервної випадкової величини X називають те її значення, для якого виконуються рівність імовірностей подій:

$$\begin{aligned}&P(-\infty < X < \text{Me}) = P(\text{Me} < X < \infty) \to \\ &\to F(\text{Me}) - F(-\infty) = F(\infty) - F(\text{Me}) \to F(\text{Me}) = \\ &= 1 - F(\text{Me}) \to 2F(\text{Me}) = 1 \to F(\text{Me}) = 0{,}5. \qquad (4.18)\end{aligned}$$

Отже, медіану визначають із рівняння (4.18).

Дисперсією випадкової величини X називається математичне сподівання квадрата відхилення цієї величини:

$$D(X) = M(X - M(X))^2. \qquad (4.19)$$

Для дискретної випадкової величини X дисперсія

$$D(X) = \sum_{i=1}^{n} (x_i - M(X))^2 p_i; \qquad (4.20)$$

для неперервної

$$D(X)=\int_{-\infty}^{\infty}(x-M(X))^2 f(x)dx. \qquad (4.21)$$

Якщо $X \in [a; b]$, то

$$D(X)=\int_{a}^{b}(x-M(X))^2 f(x)dx. \qquad (4.22)$$

Властивості дисперсії

1. Якщо C – стала величина, то

$$D(C)=0.$$

Справді, $D(C)=M(C-M(C))^2=M(C-C)^2=M(0)=0$.

2. $D(CX)=C^2D(X)$.

Маємо:

$$D(CX)=M(CX-M(CX))^2=M(CX-CM(X))^2=M(C(X-M(X))^2=$$
$$=C^2M(X-M(X))^2=C^2D(X).$$

3. Якщо A і B – сталі величини, то

$$D(AX+B)=A^2D\left(X\right).$$

Справді,

$$D(AX+B)=M(AX+B-M(AX+B))^2=M(AX+B-AM(X)-B)^2=$$
$$=M(AX-AM(X))^2=A^2M(X-M(X))^2=A^2D(X).$$

Дисперсію можна обчислити і за такою формулою:

$$D(X)=M(X^2)-M^2(X). \qquad (4.23)$$

$$D(X)=M(X-M(X))^2=M(X^2-2M(X)X+M^2(X))=$$
$$=M(X^2)-2M(X)M(X)+M(M^2(X))=$$
$$=M(X^2)-2M^2(X)+M^2(X)=M(X^2)-M^2(X).$$

Для дискретної випадкової величини X

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

для неперервної

$$D(X)=\int_{-\infty}^{\infty}x^2 f(x)dx-M^2(X).$$

Якщо $X\in[a;b]$, то

$$D(X)=\int_{a}^{b}x^2 f(x)dx-M^2(X).$$

Слід пам'ятати, що дисперсія не може бути від'ємною величиною $\left(D(X\geq 0\right)$.

Отже, дисперсія характеризує розсіювання випадкової величини відносно свого математичного сподівання. Якщо випадкова величина виміряна в деяких одиницях, то дисперсія вимірюватиметься в тих самих одиницях, але в квадраті.

Тому доцільно мати числову характеристику такої ж вимірності, як і випадкова величина. Такою числовою характеристикою є середнє квадратичне відхилення.

Середнім квадратичним відхиленням випадкової величини X називають корінь квадратний із дисперсії:

$$\sigma(X)=\sqrt{D(X)} \tag{4.24}$$

<u>*Приклад 4.*</u> Задано щільність імовірностей (рис. 11)

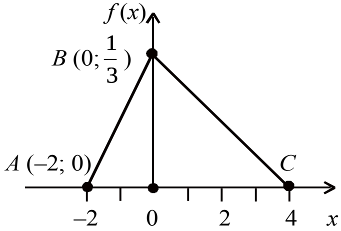


Рис. 11.

Обчислити $D(X)$; $\sigma(X)$; Ме. Знайти Мо.

Розв'язання. За умовою нормування знаходимо ординату точки B:

$$\frac{AB \cdot OB}{2} = 1 \to \frac{6 \cdot 7}{2} = 1 \to y = \frac{1}{3}.$$

На проміжку [–2; 0] $f(x) = \dfrac{x+2}{6}$.

На [0; 4] $f(x) = -\dfrac{(x-4)}{12} = \dfrac{4-x}{12}$.

Отже, щільність імовірностей $f(x) = \begin{cases} 0, & x \le -2; \\ \dfrac{x+2}{6}, & -2 < x \le 0; \\ \dfrac{4-x}{12}, & 0, \ x > 4. \end{cases}$

Знаходимо функцію розподілу ймовірностей:

На проміжку [–2; 0] $F(x) = \int\limits_{-2}^{x} \dfrac{x+2}{6} = \dfrac{(x+2)^2}{12}$.

На [–2; 4] $F(x) = F(0) + \int\limits_{0}^{x} \dfrac{4-x}{12}\,dx = \dfrac{4}{12} - \dfrac{(4-x)^2}{24}\Bigg|_{0}^{x} =$

$$= \frac{4}{12} - \frac{(x-4)^2}{24} + \frac{16}{24} = \frac{1}{3} + \frac{2}{3} - \frac{(x-4)^2}{24} = 1 - \frac{(x-4)^2}{24}.$$

Отже, функцію розподілу ймовірностей можна подати у такому вигляді:

$$F(x)=\begin{cases}0, & x\le -2;\\ \dfrac{(x+2)^2}{12}, & -2<x\le 0;\\ 1-\dfrac{(x-4)^2}{24}, & 0<x\le 4;\\ 1, & x>4.\end{cases}$$

Графік $F(x)$ зображено на рис. 12.

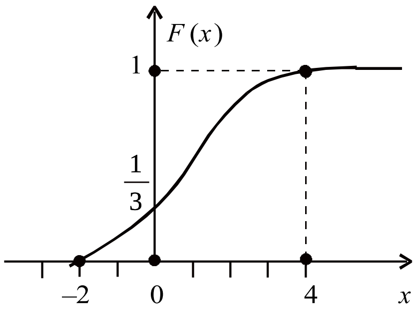


Рис. 12.

Далі обчислюємо $D(X)$:

$$M(X)=\int_{-2}^{0}xf(x)dx+\int_{0}^{4}xf(x)dx=$$

$$=\int_{-2}^{0}x\frac{x+2}{6}dx+\int_{0}^{4}x\frac{4-x}{12}dx=\frac{1}{6}\int_{-2}^{0}(x^2+2x)dx+\frac{1}{12}\int_{0}^{4}(4x-x^2)dx=$$

$$=\frac{1}{6}\left(\int_{-2}^{0}x^2dx+2\int_{-2}^{0}xdx\right)+\frac{1}{12}\left(4\int_{0}^{4}xdx-\int_{0}^{4}x^2dx\right)=$$

$$=\frac{1}{6}\left(\frac{x^3}{3}\Big|_{-2}^{0}+x^2\Big|_{-2}^{0}\right)+\frac{1}{12}\left(2x^2\Big|_{0}^{4}-\frac{x^3}{3}\Big|_{0}^{4}\right)=$$

$$=\frac{1}{6}\left(\frac{8}{3}-4\right)+\frac{1}{12}\left(32-\frac{64}{3}\right)=\frac{1}{6}\left(\frac{8-12}{3}\right)+\frac{1}{12}\left(\frac{96-64}{3}\right)=$$

$$=-\frac{4}{18}+\frac{32}{36}=\frac{-8+32}{36}=\frac{24}{36}=\frac{2}{3}.$$

$$M(X^2)=\int_{-2}^{0}x^2 f(x)dx+\int_{0}^{4}x^2 f(x)dx=\frac{1}{6}\int_{-2}^{0}x^2(x+2)dx+\frac{1}{12}\int_{0}^{4}x^2(4-x)dx=$$

$$=\frac{1}{6}\left(\int_{-2}^{0}x^3dx+2\int_{-2}^{0}x^2dx\right)+\frac{1}{12}\left(4\int_{0}^{4}x^2dx-\int_{0}^{4}x^3dx\right)=$$

$$=\frac{1}{6}\left(-4+\frac{16}{3}\right)+\frac{1}{12}\left(\frac{256}{3}-64\right)=\frac{1}{6}\left(\frac{-12+16}{3}\right)+\frac{1}{12}\left(\frac{256-192}{3}\right)=$$

$$=\frac{4}{18}+\frac{64}{36}=\frac{8+64}{36}=\frac{72}{36}=2;$$

$$D(X)=M(X^2)-M^2(x)=2-\left(\frac{2}{3}\right)^2=2-\frac{4}{9}=\frac{18-4}{9}=\frac{14}{9};$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{14}{9}}=\frac{\sqrt{14}}{9}=\frac{\sqrt{14}}{3}.$$

Для визначення Ме необхідно знайти проміжок, в якому вона міститься. Оскільки $F(0)=\dfrac{1}{3}<0{,}5,$ то медіана належить проміжку [0; 4].

Далі маємо:

$$1-\frac{(Me-4)^2}{24}=\frac{1}{2}\to\frac{(Me-4)^2}{24}=\frac{1}{2}\to(Me-4)^2=12\to$$
$$\to Me-4=\pm2\sqrt{3}\to Me=4\pm2\sqrt{3}\to Me=4+2\sqrt{3}\in[-2;\ 4];$$
$$Me=4-2\sqrt{3}\in[-2;4].$$

Отже, Me = $4-2\sqrt{3}$; Mo = 0.

4.3. Деякі найпоширеніші закони розподілу

Розглянемо деякі закони розподілу випадкових величин, які найчастіше використовуються в економетрії.

4.3.1. Рівномірний розподіл

Неперервна випадкова величина *X розподілена рівномірно на* інтервалі [*a, b*], якщо на цьому інтервалі її щільність імовірності є сталою величиною, а поза даним інтервалом дорівнює нулю:

$$f(x)=\begin{cases}0 \text{ для } x<a,\\ C=const \text{ для } a\le x\le b,\\ 0 \text{ для } x>b.\end{cases} \tag{4.25}$$

Графіки рівномірного розподілу зображені на рис. 13 і 14.

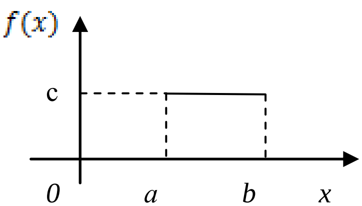


Рис. 13. Щільність імовірності рівномірного розподілу

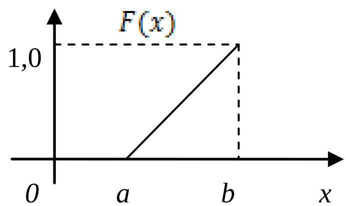


Рис. 14. Інтегральний закон рівномірного розподілу

Основні числові характеристики:

математичне сподівання: $M(X) = \bar{x} = \tilde{x} = \frac{b+a}{2}$;

дисперсія: D(X) = $\sigma_x^2 = \frac{(b-a)^2}{12b-a}$;

середнє квадратичне відхилення: $\sigma_x = \frac{b-a}{2\sqrt{3}}$.

4.3.2. Біноміальний розподіл

Дискретна випадкова величина *X* розподілена за *біноміальним законом,* якщо вона приймає значення *0, 1, 2,… m, … n* з імовірностями

$$P(X = m) = C_n^m p^m q^{n-m}, \qquad (4.26)$$

де $0 < p < 1$, $q = 1- p$, $m = 0, 1,…, n$.

Біноміальний розподіл – це закон розподілу числа *X* = *m* на події *A* в *n* незалежних випробуваннях, у кожному з яких вона може відбутися з однією і тією ж імовірністю *p*.

Формула (4.14.) називається *формулою Бернуллі.*

Основні числові характеристики: *M(X) = np; D(X)= npq*.

4.3.3. Розподіл Пуассона

Дискретна випадкова величина Х має *закон розподілу Пуассона,* якщо вона приймає значення *0, 1, 2,…, m,….*(нескінченна, але зчисленна множина) з імовірностями

$$P(X = m) = \frac{a^m}{m!} e^{-a} \qquad (4.27)$$

де *m*= 0,1,2,…; *a* – середня кількість появи подій за проміжок часу $\Delta t, a = np$; *n* – кількість випробувань; *p* – ймовірність появи події.

Закону Пуассона підпорядковане випадкове число подій m, які відбуваються за певний проміжок часу за умови, що ці події незалежні одна від одної, середня інтенсивність їх появи постійна, а ймовірність їх появи p мала.

Формула (4.15) називається формулою Пуассона.

Основні числові характеристики:

$$\bar{x} = M(X) = a = np;\ \sigma_x^2 = D(x) = a;\ \sigma_x = \sqrt{np}.$$

4.3.4. Показниковий (експоненціальний) розподіл

Неперервна випадкова величина X має *показниковий* або *експоненціальний закон розподілу з* параметром λ, якщо її щільність імовірності визначається залежністю:

$$f(x) = \begin{cases} 0, \text{для } x < 0; \\ \lambda e^{-\lambda x}, \text{для } x \geq 0. \end{cases} \qquad (4.28)$$

Інтегральна функція розподілу випадкової величини X, розподіленої за показниковим законом:

$$F(x) = 1 - e^{-\lambda x}, \text{для } x \geq 0 \qquad (4.29)$$

Основні числові характеристики:

$$\bar{x} = M(x) = \frac{1}{\lambda};\ \sigma_x^2 = \frac{1}{\lambda^2};\ \sigma_x = \frac{1}{\lambda}.$$

4.3.5. Нормальний розподіл

Неперервна випадкова величина X має *нормальний розподіл,* або закон розподілу Гаусса з параметрами a і σ^2, якщо її щільність імовірності визначається за формулою:

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{(x-a)^2 12\sigma^2} \qquad (4.30)$$

Криву нормального закону розподілу називають *нормальною* (*гауссовою) кривою* (рис. 15).

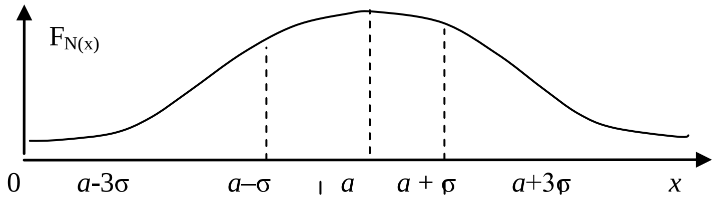


Рис. 15. Крива нормального розподілу

Основні числові characteristics:

$\bar{x} = M(X) = a;\ D(X) = \sigma_x^2 = \sigma^2;\ \sigma_x = \sigma.$

Нормальний закон розподілу може бути заданий у вигляді функції розподілу:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x-a)^2/2\sigma^2} dx. \quad (4.31)$$

Нормальний закон розподілу з параметрами $a = 0$, $\sigma^2 = 1$, тобто *N(0;1)*, називається *стандартним* (або *нормованим)*.

На практиці часто необхідно знаходити ймовірність того, що нормально розподілена випадкова величина знаходиться в певних *межах.*

Розрахунок проводиться за допомогою спеціальної *функції Лапласа* або *інтеграла ймовірності Ф(t):*

$$\Phi(t)= \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt, \quad (4.32)$$

де $t = (x-a)\backslash\sigma;\ x = \sigma t + a;\ dx = \sigma dt.$

Функція розподілу випадкової величини *X*, розподіленої за нормальним законом, виражається через функцію Лапласа за формулами:

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

$$\text{або } F(t) = \frac{1}{2} + \frac{1}{2}\Phi(t). \qquad (4.33)$$

Для значень функції Лапласа складені спеціальні таблиці (додаток 1).

Розглянемо деякі *властивості* випадкової величини, розподіленої за нормальним законом.

1. Імовірність попадання випадкової величини *X*, розподіленої за нормальним законом, в інтервал $[x_1; x_2]$ дорівнює:

$$P(x_1 \le X \le x_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)], \qquad (4.34)$$

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

2. Імовірність того, що відхилення випадкової величини, розподіленої за нормальним законом, від математичного сподівання *a* не перевищить величину $\Delta = t\sigma$ ($\Delta > 0$ за абсолютною величиною), дорівнює:

$$P(|X - a| < t\sigma) = \Phi(t). \qquad (4.35)$$

Отже, значення функції *Ф(t)* визначає імовірність того, що відхилення нормально розподіленої випадкової величини за абсолютною величиною менше від t_σ.

Для $t = 3$ маємо $P(|X - a| < 3\sigma) = \Phi(3) = 0{,}9987$.

Це означає, що розсіяння нормально розподіленої випадкової величини вкладається на діапазоні $M(X) \pm 3\sigma$. Ймовірність того, що випадкова величина *X* лежить поза цим діапазоном, дуже мала, а саме *0,0013*, тобто ця подія може відбутися лише в *0,13%* випадків. Такі події можна вважати практично неможливими.

Звідси випливає *правило трьох сигм:* якщо випадкова величина *X* має нормальний розподіл, то відхилення цієї величини від математичного сподівання за абсолютною величиною не перевищує потрійного середнього квадратичного відхилення, тобто *практично достовірно, що її значення лежать в інтервалі (а-3σ, а+3σ).*

Нормальний розподіл відіграє особливу, центральну роль в теорії і практиці статистичних та економетричних досліджень, яка базується на наслідках *центральної граничної теореми.*

Теорема стверджує: якщо випадкова величина є загальним результатом взаємодії великої кількості інших випадкових величин, жодна з яких не є домінуючою, то вона буде приблизно нормально розподілена, навіть якщо окремі складові не мають нормального розподілу [2, с. 82].

Центральна гранична теорема може бути застосована у випадках, коли: випадкові складові не є однаково розподіленими (теорема Ляпунова); випадкові величини X_i не є незалежними; випадкові величини X_i є багатовимірними.

4.3.6. Логнормальний розподіл

Неперервна випадкова величина X має *логарифмічно нормальний (логнормальний) розподіл,* якщо $\ln X$ підпорядкований нормальному закону.

Основні числові характеристики:

$$\overline{x} = M(X) = a \cdot e^{1/2\sigma^2},\ D(X) = a^2 e^{\sigma^2}\left(e^{\sigma^2} - 1\right).$$

Запитання та завдання для самоконтролю

1. Дайте визначення випадкової величини.
2. Які основні властивості дискретних та неперервних випадкових величин?
3. Що називається розподілом та законом розподілу випадкової величини?
4. У якому вигляді задається закон розподілу дискретної випадкової величини?
5. Які випадкові величини вважаються незалежними?
6. У яких формах задаються закони розподілу?

7. Що називається інтегральним законом або функцією розподілу?

8. Назвіть властивості функції розподілу випадкової величини.

9. Дайте визначення неперервної випадкової величини.

10. Дайте графічну інтерпретацію інтегрального закону для дискретних і неперервних випадкових величин.

11. Дайте визначення щільності ймовірності випадкової величини.

12. Назвіть властивості щільності ймовірності.

13. У чому полягає суть рівномірного розподілу?

14. Дайте визначення біноміального розподілу.

15. Охарактеризуйте суть розподілу Пуассона.

16. Охарактеризуйте суть показникового розподілу.

17. У чому полягає суть нормального розподілу?

18. Побудуйте та охарактеризуйте нормальну криву.

19. Як визначається функція Лапласа?

20. Назвіть властивості випадкових величин, розподілених за нормальним законом.

21. Сформулюйте правило трьох сигм.

22. Чому нормальний розподіл відіграє центральну роль у статистиці та економетрії?

23. У чому суть центральної граничної теореми?

24. У грошовій лотереї розігрується два виграші по 1000 гривень, 10 виграшів по 100 гривень і 100 виграшів по 10 гривень при загальній кількості білетів 10000. Запишіть закон розподілу випадкової величини X – величини виграшу власника одного лотерейного білета.

25. Пристрій складається з трьох незалежно працюючих елементів. Імовірність відмови кожного елемента в одному досліді дорівнює 0,1. Складіть закон розподілу випадкової величини X – числа елементів, які відмовили в одному досліді.

26. Електронна пошта банку підтримує зв’язок із сотнею абонентів. Імовірність того, що на електронну пошту надійде повідомлення від абонента, дорівнює 0,02. Складіть закон розподілу випадкової величини X – числа надходжень сигналів від абонентів.

27. У партії з 10 виробів є 7 стандартних. Навмання взяли 3 вироби.

Побудуйте: 1) закон розподілу випадкової величини X – кількості стандартних виробів серед відібраних; 2) функцію розподілу та її графік; 3) многокутник розподілу.

Обчисліть числові характеристики (математичне сподівання, дисперсію, середнє квадратичне відхилення, моду, коефіцієнт асиметрії та ексцес).

28. З 10 дівчат та 8 хлопців утворюють групу з 3 чоловік.

Побудуйте: 1) закон розподілу випадкової величини X – числа дівчат у даній групі; 2) функцію розподілу та її графік; 3) многокутник розподілу.

Обчисліть числові характеристики (математичне сподівання, дисперсію, середнє квадратичне відхилення, моду).

29. Графік заданої щільності ймовірностей зображено на рис. 16.

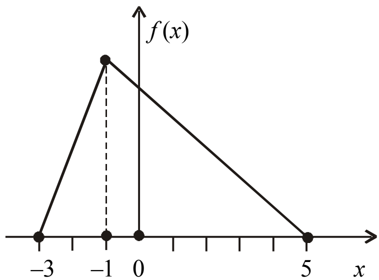


Рис. 16.

Запишіть вирази для $f(x)$ і $F(x)$. Побудуйте графік $F(x)$.

30. Менеджер банку, який займається інвестиціями, припускає, що в недалекому майбутньому можливе зростання ділової активності, застій і спадання активності відповідно з імовірностями 0,1;0,5;0,4. Менеджер сподівається отримати 30% прибутку з активу при високому зростанні активності, 10% у випадку застою і 1% у випадку спадання активності. Визначте (у відсотках) очікуваний рівень дохідності активу та його ризик.

31. Дано функцію розподілу ймовірностей

$$F(x)=\begin{cases}0, & x\le -2;\\ \dfrac{\sqrt{x+2}}{3}, & -2<x\le 7;\\ 1, & x>7.\end{cases}$$

Знайдіть $f(x)$. Побудуйте графіки $F(x)$ і $f(x)$.

32. Знайдіть математичне сподівання та дисперсію випадкової величини $Z=X-2Y$, якщо $M(X)=5$, $M(Y)=3$, $D(X)=4$, $D(Y)=5$.

33. Дискретна випадкова величина X набуває трьох значень: $x_1=4$ з імовірністю $p_1=0{,}5$, $x_2=6$ з імовірністю $p_2=0{,}3$, x_3 з імовірністю p_3. Знайдіть x_3 і p_3, враховуючи, що $M(X)=8$.

34. Дискретна випадкова величина X набуває значення $x_1=-1$, $x_2=0$, $x_3=1$. Відомі математичні сподівання X і X^2: $M(X)==0{,}1$, $M(X^2)=0{,}9$. Знайдіть імовірності p_1, p_2, p_3 для значень x_1, x_2, x_3 відповідно.

35. Знайдіть закон розподілу дискретної випадкової величини X, що може набувати тільки двох значень: x_1 з імовірністю $p_1=0{,}7$ і x_2, причому $x_1<x_2$ і $M(X)=3{,}3$, $D(X)=0{,}21$.

36. Дискретна випадкова величина X набуває трьох значень: $x_1=2$ з імовірністю $p_1=0{,}2$, $x_2=3$ з імовірністю $p_2=0{,}4$, x_3 з імовірністю p_3. Знайдіть x_3 і p_3, знаючи, що $M(X)=6$.

37. Дискретна випадкова величина X набуває значення $x_1=1$, $x_2=2$, $x_3=3$. Відомі математичні сподівання X і X^2: $M(X)=2{,}3$, $M(X^2)=5{,}9$. Знайдіть імовірності p_1, p_2, p_3, які відповідають значенням x_1, x_2, x_3.

38. Знайдіть закон розподілу дискретної випадкової величини X, що може набувати тільки двох значень: x_1 з імовірністю $p_1=0{,}4$ і x_2, причому $x_1<x_2$ і $M(X)=2{,}6$, $D(X)=0{,}24$.

39. Щільність розподілу неперервної випадкової величини X задано функцією

$$f(x)=\begin{cases}0, x\le 0;\\ x^2, 0<x\le 1;\\ 0, x>1.\end{cases}$$

Обчисліть $M(X)$, $D(X)$, $\sigma(X)$.

40. Задано закон розподілу ймовірностей

$$f(x)=\begin{cases}0, & x<-3;\\ \dfrac{1}{4\sqrt{x+3}}, & -3<x\le 1;\\ 0, & x>1.\end{cases}$$

Знайдіть $D(X)$; $\sigma(X)$; Me.

41. Обчисліть $M(Y)$, $D(Y)$, $\sigma(Y)$ для неперервної випадкової величини $Y=X^2$, якщо

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

42. Випадкова величина X має закон розподілу щільності ймовірностей рівнобедреного трикутника (рис. 17).

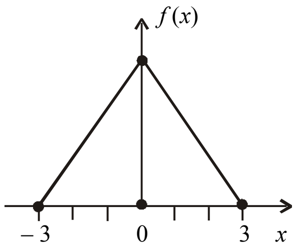


Рис. 17.

Знайдіть $M(X)$; $D(X)$; $\sigma(X)$; μ_3.

43. Задано

$$f(x)=\begin{cases}0, & x\le 0;\\ ax^2e^{-x^2}, & x>0.\end{cases}$$

Знайдіть a; $F(x)$; $M(X)$; $D(X)$.

Тема 5. ТОЧКОВІ ТА ІНТЕРВАЛЬНІ ОЦІНКИ ПАРАМЕТРІВ

1. *Поняття і суть оцінок параметрів.*
2. *Властивості оцінок параметрів.*
3. *Метод максимальної правдоподібності.*
4. *Суть інтервальних оцінок параметрів.*

5.1. Точкові оцінки

Нехай з деякої генеральної сукупності випадкової величини зроблено випадкову вибірку $(x_1, x_2, \dots x_n)$, обчислено середнє арифметичне цієї вибірки $\overline{x}_n$, оцінку дисперсії S_n^2 та інші числові характеристики. Очевидно, що $\overline{x}_n$ та S_n^2 є функціями значень $x_1, x_2, \dots x_n$ і при кожній наступній вибірці можуть змінювати свої значення (тією чи іншою мірою). Такі функції, які залежать від складу вибірки, називаються *вибірковими.*

Як відомо, однією з основних задач математичної статистики є одержання надійної інформації про сукупність в цілому на основі даних випадкових вибірок. Для цього необхідно знати зв’язок між параметрами вибіркових статистик і параметрами сукупностей, з яких здійснюються вибірки. Вибірки беруться з сукупностей випадково. Тому вибіркові статистики є випадковими величинами і, як і будь-які величини, мають свої числові характеристики та закони розподілу.

Важливим є той факт, що розподіл вибіркового середнього $\overline{x}_n$ відповідно до центральної граничної теореми є близьким до нього. Якщо навіть вихідна сукупність X не є нормальною, при $n>30$ розподіл $\overline{x}_n$ близький до нормального. Якщо ж вихідна сукупність є нормальною, то розподіл $\overline{x}_n$ є нормальним і при невеликих вибірках.

ЯК зазначалося вище, практично визначаються не самі характеристики випадкових величин, а їх *випадкові оцінки.* Існують два підходи до одержання значень оцінок: *точкові оцінки* та *інтервальні.*

Точкові оцінки дають єдине значення кожної числової характеристики, кожного параметра сукупності (у вигляді одного числа). Позначимо через β значення параметра, а через b його оцінку.

Оцінкою b *параметра* β називають будь-яку функцію результатів спостережень над випадковою величиною X, за допомогою якої оцінюють значення параметра β.

Оскільки X – випадкова величина, то і b є випадковою величиною.

Для того, щоб оцінка b могла дати реальну характеристику параметра β, вона повинна мати наступні властивості:

1) незміщеність;
2) обґрунтованість;
3) ефективність.

Оцінка b параметра β називається *незміщеною,* якщо її математичне сподівання дорівнює параметру, що оцінюється, тобто

$$\mathrm{M}(b) = \beta. \tag{5.1}$$

У протилежному випадку оцінка називається *зміщеною.* Різниця між математичним сподіванням оцінки і відповідною теоретичною характеристикою генеральної сукупності називається *зміщенням* оцінки.

Якщо рівність (5.1) не виконується, то оцінка b може або завищувати значення β ($M(b) > \beta$), або занижувати його ($M(b) < \beta$). Це призводить до систематичних похибок в оцінці параметра β. Отже, вимога незміщеності гарантує відсутність систематичних похибок в оцінці параметрів. Це означає, що якщо зроблено l різних вибірок і в кожній з них отримані свої оцінки $b_j(j = \overline{1, l})$, то в середньому за результатами всіх вибірок буде отримано значення оцінки, близьке до β, причому чим більше зроблено вибірок і чим більший їх обсяг, тим b ближче до β.

Оцінка b параметра β називається *обґрунтованою,* якщо вона підпорядкована *закону великих чисел,* тобто *збігається за ймовірністю,* а отже, виконується рівність

$$\lim_{n \to \infty} P(|b - \beta| \le \varepsilon) = 1, \tag{5.2}$$

де ε – достатньо мала додатня величина.

Під *законом великих чисел* у широкому аспекті розуміють загальний принцип, згідно з яким, за формулюванням академіка Колмогорова, сукупна дія великої кількості випадкових факторів веде (за деяких загальних умов) до результату, що майже не залежить від випадку, тобто при великій кількості випадкових величин їх середній результат перестає бути випадковим і може бути передбачений з великою ймовірністю [5, с. 41]. Строге формулювання закону великих чисел наведено в роботі С.А. Айвазяна та В.С. Мхитаряна [1, с. 154] й у класичних підручниках з теорії ймовірності та математичної статистики.

Обґрунтованість оцінки означає, що чим більшим є обсяг вибірки, тим більша ймовірність того, що похибка оцінки не перевищить достатньо малого додатнього числа ε. Отже, у випадку використання обґрунтованих оцінок стає доцільним збільшення обсягу вибірки, оскільки при цьому зменшується ймовірність значної похибки в оцінці невідомого параметра.

Незміщена оцінка b, яка має найменшу дисперсію серед усіх незміщених оцінок параметра β, обчислених за одного й того ж обсягу n, називається *ефективною оцінкою.*

Основним методом знаходження оцінок генеральної сукупності за даними вибірки є *метод максимальної правдоподібності.*

Цей метод розроблений видатним англійським математиком Р. Фішером. Розглянемо основну ідею цього методу.

Нехай $x_1, x_2, \dots, x_n$ – *вибіркові дані,* де n – кількість незалежних спостережень над випадковою величиною X, яка може бути як дискретною, так і неперервною, $f(x, \theta)$ – ймовірність значення (функція розподілу), якщо випадкова величина дискретна, або щільність імовірності, якщо випадкова величина неперервна. Функція $f(x, \theta)$ залежить від невідомого параметра θ, який необхідно оцінити за даними вибірки.

Якщо $x_1, x_2, \dots, x_j, \dots, x_n$ – незалежні випадкові величини, то вираз

$$L = f(x_1, \theta) f(x_2, \theta) \dots f(x_i, \theta) \dots f(x_n, \theta) \qquad (5.3)$$

називається *функцією правдоподібності.*

В якості оцінки невідомого параметра θ береться таке значення $\tilde{\theta}$, при підстановці якого у вираз (5.3) замість параметра θ отримаємо максимальне значення функції L.

Оцінку $\tilde{\theta}$ називають *оцінкою максимальної правдоподібності.* Зауважимо, що оцінка максимальної правдоподібності $\tilde{\theta}$ залежить від кількості та числових значень випадкових величин $x_1, x_2, \ldots, x_j, \ldots, x_n$, отже, теж є випадковою величиною.

Якщо функція L диференційована відносно параметра θ, то оцінка цього параметра зводиться до знаходження функції L відносно θ. Якщо L недиференційована відносно параметра θ, то для обчислення максимуму функції L використовуються інші методи, що іноді пов'язано із значними труднощами обчислення.

Якщо функція L диференційовна, то для знаходження оцінки максимальної правдоподібності необхідно розв'язати рівняння

$$\frac{\partial L}{\partial \theta} = 0 \tag{5.4}$$

і вибрати той розв'язок, при якому функція L набуває максимального значення.

Перевагою методу максимальної правдоподібності є те, що знайдені *оцінки обґрунтовані, асимптотично ефективні* (при $n \to \infty$ мають *асимптотично* нормальний розподіл).

Найважливішими числовими характеристиками (параметрами) випадкової величини є математичне сподівання і дисперсія.

Нехай для оцінки параметрів розподілу випадкової величини X підібрана випадкова вибірка ($x_1, x_2, \ldots, x_n$). Обчислені вибіркові характеристики:

- вибіркове середнє $\bar{x} = \frac{\sum_{i=0}^{n} x_i}{n}$;
- вибіркова дисперсія $\sigma^2 = \frac{\sum_{i=0}^{n} (x_i - \bar{x})^2}{n}$.

Властивості параметрів генеральної сукупності і вибіркові статистики базуються на таких основних теоремах.

Теорема 5.1. Арифметичне середнє $\bar{x}$, обчислене за n незалежними спостереженнями над випадковою величиною X, яка має математичне сподівання μ, є *незміщеною оцінкою* цього параметра.

Теорема 5.2. Арифметичне середнє $\bar{x}$, обчислене за n незалежними спостереженнями над випадковою величиною X, яка має математичне сподівання μ і дисперсію σ^2, є *обґрунтованою оцінкою* математичного сподівання.

Теорема 5.3. Якщо випадкова величина X розподілена за нормальним законом з параметрами μ, σ^2, то незміщена оцінка $\bar{x}$ математичного сподівання μ має мінімальну дисперсію σ^2/n, тому середнє арифметичне $\bar{x}$ є *ефективною* оцінкою математичного сподівання μ.

Теорема 5.4. Якщо випадкова вибірка складається з n незалежних спостережень над випадковою величиною X з математичним сподіванням μ і дисперсією σ^2, то вибіркова дисперсія

$$S = 1/n \sum_{i=0}^{n} (x_i - \bar{x})^2$$

не є незміщеною оцінкою дисперсії генеральної сукупності σ^2.

Незміщена оцінка дисперсії генеральної сукупності має такий вигляд:

$$S^{2*} = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=0}^{n} (x_i - \bar{x})^2. \tag{5.5}$$

Оцінка (5.5) називається *виправленою вибірковою дисперсією.*

Дріб *n/(n-1)* називають *поправкою Бесселя.* При малих значеннях n поправка Бесселя досить значно відрізняється від одиниці, із збільшенням n вона швидко прямує до одиниці. При $n > 50$ практично відсутня різниця між оцінками S^2 і S^2*, і оцінки S^2 та S^2* є *обґрунтованими оцінками* σ^2.

Незміщеною, обґрунтованою та *ефективною оцінкою* σ^2 є *оцінка*

$$S^2 = \frac{1}{n} \sum_{i=0}^{n} (x_i - \mu)^2, \tag{5.6}$$

для обчислення якої необхідно знайти математичне сподівання μ випадкової величини X. Оцінки S^2 і S^2* *не є ефективними.* У випадку, коли значення математичного сподівання невідоме, для оцінки дисперсії σ^2 користуються обґрунтованою і незміщеною оцінкою S^2*.

Розглянемо ще декілька важливих понять, необхідних для обробки статистичної інформації.

З теореми 5.1 випливає, що $M(\bar{x}) = \mu$, а з теорем 5.2, 5.3, що $D(\bar{x})=\sigma^2/n$. Відповідно, $\sigma_x = \sigma/\sqrt{n}$.

Центрованою випадковою величиною $\overset{\circ}{X}$ називається відхилення випадкової величини X від її математичного сподівання $M(X)$:

$$\overset{\circ}{X} = X - M(X). \tag{5.7}$$

Нормованою величиною Z називається центрована випадкова величина, *виміряна* у масштабі середніх квадратичних відхилень:

$$Z = \frac{\overset{\circ}{X}}{\sigma_x} = \frac{X - M(X)}{\sigma_x}. \tag{5.8}$$

Враховуючи вираз (5.8) для величини Z, зауважимо, що нормальний закон розподілу (формули (4.18) і (4.19)) можна записати як закон для нормованої випадкової величини, оскільки вираз $\frac{x-a}{\sigma}$ є нормованою випадковою величиною, тобто $Z=\frac{x-a}{\sigma}$.

Тоді формули (4.18) і (4.19) можна записати так:

$$f(Z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2}; \tag{5.9}$$

$$F(Z) = \int_{-\infty}^{z} f(Z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz \tag{5.10}$$

Оскільки з теорем 5.2 і 5.3 випливає, що $\sigma_x=\sigma/\sqrt{n}$, то для випадкової величини X, розподіленої за нормальним законом із середнім μ та дисперсією σ^2, для вибірки обсягом n маємо:

$$Z = \frac{\bar{x} - \mu}{\sigma}\sqrt{n}. \tag{5.11}$$

Оскільки $\bar{x}$ та σ_x – випадкові величини, величина Z буде розподіленою так само, як і X, тобто за нормальним законом. Як відомо, для функції Z складені спеціальні таблиці.

Таблиця функції Лапласа Φ (Z) подана у додатку 1. За допомогою таких таблиць обчислюється ймовірність відхилення випадкової величини від свого математичного сподівання.

Розглянутий Z – розподіл є досить зручним і широко використовується у статистичних розрахунках. Натомість для отримання значень Z необхідно знати дисперсію генеральної сукупності σ^2, яка майже завжди є невідомою. Відомі або зміщена оцінка дисперсії S^2, або виправлена (незміщена) оцінка S^2*. Тому великий практичний інтерес викликає вивчення розподілу статистики

$$t = \frac{\bar{x}-\mu}{s}\sqrt{n-1}. \tag{5.12}$$

Закон розподілу статистики t називається t-*розподілом Стьюдента* з n-1 ступенями свободи. Доведено, що цей розподіл не залежить ні від математичного сподівання μ випадкової величини X, ні від дисперсії σ^2, а залежить лише від обсягу вибірки n.

Для функції розподілу Стьюдента складені спеціальні таблиці (див. додаток 2). Входами у ці таблиці є число ступенів свободи k або df та імовірність p або *рівень значущості.*

Для обсягу вибірки n визначають число ступенів свободи $k = n-1$. Далі обирають значення ймовірності p (рівня значущості), для якої за таблицями визначають величину $t_{p,k}$ або $t_{a,k}$. Ймовірність (рівень значущості) обирають, виходячи з умов розв'язку задачі. На практиці в економетрії використовують імовірність не нижче 0,95 (рівень значущості не більше *0,05*).

Зауважимо, що t-розподіл суттєво відрізняється від Z-розподілу за малих обсягів вибірок. Це обумовлено різницею між величинами σ та S* при малих значеннях n.

За умови $n \geq 50$ розподіл Стьюдента вже незначно відрізняється від нормального.

Статистика $t = \frac{\bar{x}-\mu}{s}\sqrt{n-1}$ має розподіл Стьюдента, якщо випадкова величина X підпорядкована нормальному закону розподілу, а середнє значення $\bar{x}$ обчислюється за вибірковими даними $x_1, x_2, \ldots, x_j, \ldots, x_n$, отриманими в результаті незалежних спостережень.

5.2. Інтервальні оцінки

У реальному житті в статистичних дослідженнях досить часто зручніше вказувати не значення параметра, а *діапазон, тобто інтервал,* у межах якого він може знаходитися.

Принцип одержання інтервальних оцінок базується на положенні, що більшість вибіркових статистик мають нормальний або близький до нормального розподіл. Це дозволяє за допомогою Z-функції знаходити ймовірність, із якою деяка величина β лежить у межах заданого інтервалу $[b_1, b_2]$.

Інтервальною оцінкою параметра β називається числовий інтервал $[b_1,b_2]$, який із заданою імовірністю P_α містить невідоме значення параметра β.

Отже, *довірчим інтервалом* $[b_1,b_2]$ для параметра β називається такий інтервал, для якого з наперед заданою ймовірністю $P_\alpha = 1 - \alpha$ (α – рівень значущості), близькою до одиниці, можна стверджувати, що він містить невідоме значення параметра β, тобто

$$P[b_1 < \beta < b_2] = P_\alpha = 1 - \alpha. \qquad (5.13)$$

Імовірність P_α називається *довірчою ймовірністю або надійністю оцінки.*

Величина довірчого інтервалу залежить від обсягу вибірки n (зменшується із зростанням n) та від значення довірчої ймовірності P_α (збільшується із наближенням P_α до одиниці).

Розглянемо приклади побудови довірчих інтервалів.

Побудуємо *довірчий інтервал для математичного сподівання при відомій дисперсії.*

Нехай випадкова величина X розподілена нормально, при цьому відоме стандартне відхилення σ цього розподілу. Необхідно оцінити невідоме математичне сподівання μ. Найкращою оцінкою математичного сподівання (незміщеною, обґрунтованою, ефективною), як показано у підрозділі 5.1, є вибіркове середнє $\bar{x}$. Також відомо, що вибіркове середнє $\bar{x}$ розподілене нормально з параметрами $M(\bar{x})=\mu$, $D(\bar{x})=\sigma^2/n$; нормоване відхилення $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ розподілене нормально з параметрами $\mu = 0$, $\sigma^2=1$. Тому ймовірність будь-якого відхилення $|\bar{x} - \mu|$ може бути обчислена за формулою

$$p\left(\left|\frac{\bar{x}-\mu}{\sigma}\sqrt{n}\right| < z_p\right) = \Phi(Z)\,. \qquad (5.14)$$

Якщо задати певну довірчу ймовірність $P_\alpha = 1 - \alpha = \Phi(z)$, то з таблиці функції Лапласа у додатку 1 можна визначити значення Z.

Для оцінки μ перетворимо формулу (5.14):

$$p\left(\bar{x} - z_p\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_p\frac{\sigma}{\sqrt{n}}\right) = \Phi(Z).$$

Отже, з імовірністю (надійністю) $\Phi(Z) = 1 - \alpha$ можна стверджувати, що довірчий інтервал

$$\left[\bar{x} - Z_p\frac{\sigma}{\sqrt{n}};\, \bar{x} + Z_p\frac{\sigma}{\sqrt{n}}\right]. \qquad (5.15)$$

<u>*Приклад 5.1.*</u> Випадкова величина X розподілена нормально з відомим стандартним відхиленням σ = 1,5. Визначити довірчий інтервал для оцінки невідомого математичного сподівання $n=25$; $\bar{x} = 4{,}2$; $P_\alpha = 1 - \alpha = 0{,}95$.

<u>*Розв'язання.*</u> *З* таблиць функції Лапласа (додаток 1) знаходимо значення z_P, яке відповідає $P_\alpha = 0{,}95$. Маємо $Z_P = 1{,}96$.

Тоді знаходимо величину Δ:

$$\Delta = Z_p\frac{\sigma}{\sqrt{n}} = (1{,}96 * 1{,}5)/\sqrt{25} = 0{,}59.$$

Отже, $(\bar{x} - 0{,}59;\ \bar{x} + 0{,}59)$ є довірчим інтервалом для оцінки μ з імовірністю $P_a = 0{,}95$, тобто фактичне значення μ лежить у межах $4{,}2 \pm 0{,}59$. Тому можна записати, що

$$4{,}2 - 0{,}59 < \mu < 4{,}2 + 0{,}59, \text{ або } 3{,}61 < \mu < 4{,}79.$$

Побудуємо *довірчий інтервал для математичного сподівання при невідомій дисперсії.*

Нехай випадкова величина X розподілена нормально, при цьому стандартне відхилення цього розподілу невідоме. Необхідно оцінити невідоме математичне сподівання μ. Як відомо, випадкова величина $t = \frac{\bar{x}-\mu}{s}\sqrt{n-1}$ розподілена за законом Стьюдента з n-1 ступенями свободи. Тому якщо взяти ймовірність $P_\alpha = 1 - \alpha$ і знати обсяг вибірки *n*, то можна, користуючись таблицями розподілу Стьюдента (додаток 2), знайти $t_{p,k}$ $(k = n-1)$ таке, що

$$p\left(\frac{|\bar{x}-\mu|}{S}\sqrt{n-1} < t_{p,n-1}\right) = 1-\alpha \qquad (5.16)$$

Проведемо перетворення формули (5.16), яке дозволить оцінити μ:

$$p\left(\bar{x}-t_{p,n-1}\frac{S}{\sqrt{n-1}} < \mu < \bar{x}+t_{p,n-1}\frac{S}{\sqrt{n-1}}\right).$$

Тому з імовірністю (надійністю) $P_\alpha = 1 - \alpha$ можна стверджувати, що інтервал

$$\left[\bar{x}-t_{p,n-1}\frac{S}{\sqrt{n-1}};\bar{x}+t_{p,n-1}\frac{S}{\sqrt{n-1}}\right] \qquad (5.17)$$

є довірчим для оцінки μ.

Між формулами (5.15) і (5.17) є істотна різниця, яка полягає в тому, що у формулі (5.17) коефіцієнт $t_{p,n-1}$ залежить не тільки від довірчої ймовірності, але й від кількості елементів у вибірці. Ця різниця є особливо значною при малій кількості спостережень.

Приклад 5.2. Із великої кількості дерев плантації одновікових лісонасаджень випадково відібрані та виміряні діаметри 5 дерев (див. табл. 5.1).

Таблиця 2

Результати вимірювань

Номер дерева	1	2	3	4	5
Діаметр дерева, см	32	28	31	29	30

Необхідно з довірчою ймовірністю *p = 0,95* визначити інтервальну оцінку для середнього діаметра дерев усієї плантації.

Розв'язання. Визначаємо середнє арифметичне вибірки

$$\bar{x} = \frac{32 + 28 + 31 + 29 + 30}{5} = 30 \text{ см.}$$

Визначаємо оцінку стандартного відхилення

$$S = \sqrt{\frac{(32-30)^2+(28-30)^2+(31-30)^2+(29-30)^2+(30-30)^2}{5}} = \sqrt{\frac{2^2+2^2+1^2+1^2}{5}} = \sqrt{\frac{10}{5}} = 2 \text{ см.}$$

За таблицями розподілу Стьюдента для *p* = 0,95 і числа ступенів свободи $k = n - 1 = 5 - 1 = 4$ знаходимо $t_{0,95;4} = 2{,}776$. Тоді

$$t_{p,n-1}\frac{S}{\sqrt{n-1}} = 2{,}776 * \frac{2}{\sqrt{5-1}} \approx 2{,}8 \text{ см.}$$

Отже, з імовірністю *p*=0,95 можна гарантувати, що середній діаметр дерев усієї плантації лежить у межах:

$$30 - 2{,}8 = 27{,}8 \text{ см.} < \mu < 30 + 2{,}8 = 32{,}8 \text{ см.}$$

Приклад 5.3. На деревообробному підприємстві працює *300* робітників основних професій. Із них випадково відібрано *n* = 26. Середньомісячна заробітна платня відібраних робітників склала $\bar{x} = 920$ грн при стандартному відхиленні *S = 120 грн.*

Необхідно з довірчою імовірністю визначити інтервальну оцінку для:

а) середньомісячної заробітної платні основних робітників підприємства;

б) суми місячних витрат підприємства на заробітну платню основних робітників.

Розв'язання. За таблицями розподілу Стьюдента знаходимо для *p* = 0,99 і *k* = 26 – 1 = 25 $t_{0,99;25} = 2{,}787$. Тоді

$$t_{p,n-1}\frac{S}{\sqrt{n-1}} = 2{,}787\ \frac{120}{\sqrt{25}} = 66{,}89 \text{ грн.}$$

Отже, з імовірністю *p* = 0,99 можна стверджувати, що середньомісячна заробітна плата основних робітників підприємства лежить у межах:

$$920 - 66{,}89 = 853{,}11 \text{ грн} < \mu < 920 + 66{,}89 = 986{,}89 \text{ грн.}$$

Нехай сума місячних витрат підприємства на заробітну платню основних робітників складає $З_0$(грн.). Тоді з імовірністю p = 0,99 можна стверджувати, що ця сума не вийде за межі інтервалу:

$$300 \times 855{,}11 = 255933 \text{ грн} < З_0 < 300 \times 986{,}89 = 296067 \text{ грн}.$$

За допомогою формул (5.15) і (5.17) можна розв'язати обернену задачу: визначити необхідний обсяг вибірки n, який з довірчою ймовірністю P_a забезпечить необхідну точність.

Для випадкової нормально розподіленої величини з відомою дисперсією обсяг вибірки

$$n = \frac{z_p^2 \sigma^2}{\Delta^2}, \tag{5.18}$$

де Δ – половина довжини довірчого інтервалу,

$$\Delta = z_p \frac{\sigma}{\sqrt{n}}.$$

Якщо дисперсія сукупності невідома, то

$$n = \frac{t_{p,k} * S^2}{\Delta^2}, \tag{5.19}$$

де $\Delta = t_{p,k} \frac{s}{\sqrt{n-1}}$.

<u>*Приклад 5.4.*</u> Для умов прикладу 5.2 необхідно визначити обсяг вибірки, який би забезпечував точність вимірювань діаметра дерев *1,5см* з довірчою ймовірністю P_α =0,95.

Розв'язання.

$$n = \frac{(2{,}76)^2 \cdot 2^2}{(1{,}5)^2} = 13{,}69 \approx 14 \text{ спостережень.}$$

Отже, для забезпечення вимірювань діаметрів дерев з точністю 1,5 *см* та довірчою ймовірністю P_α =0,95 необхідно випадково відібрати та виміряти діаметри 14 дерев.

Запитання та завдання для самоконтролю

1. Дайте визначення вибіркових статистик. Охарактеризуйте їх властивості.
2. Які підходи існують до визначення оцінок випадкових величин?
3. Що таке точкова оцінка? Яким властивостям вона повинна задовільняти?
4. У чому полягає суть незміщеності оцінок?
5. У чому полягає суть обґрунтованості оцінок?
6. Яка оцінка вважається ефективною?
7. Охарактеризуйте метод максимальної правдоподібності.
8. Сформулюйте основні теореми щодо параметрів генеральної сукупності і вибіркових статистик.
9. Дайте визначення центрованої та нормованої випадкової величини.
10. У чому полягає суть t-розподілу Стьюдента?
11. Дайте визначення інтервальної оцінки та довірчого інтервалу.
12. Складіть алгоритм побудови довірчого інтервалу для математичного сподівання при відомій дисперсії.
13. Складіть алгоритм побудови довірчого інтервалу для математичного сподівання при невідомій дисперсії.
14. Наведіть приклади побудови довірчих інтервалів.
15. У будинку відпочинку випадковим способом було відібрано 20 осіб і виміряно їхній зріст x_i.

Отримані результати наведено у вигляді інтервального статистичного розподілу:

x_i, см	165,5–170,5	170,5–175,5	175,5–180,5	180,5–185,5
n_i	4	6	8	2

Із надійністю $\gamma = 0,99$ побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 2$.

16. У 25 осіб було виміряно кров'яний тиск x_i (в умовних одиницях). Результати вимірювання наведено у вигляді дискретного статистичного розподілу:

x_i	1,5	1,8	2,3	2,5	2,9	3,3
n_i	2	3	5	8	4	3

Із надійністю γ = 0,999 побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 1$.

17. Виміряно максимальну місткість конденсаторів x_i (у пікофарадах). Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i пФ	4,0-4,2	4,2-4,4	4,4-4,6	4,6-4,8	4,8-5,0
n_i	2	4	5	9	5

Із надійністю γ = 0,999 побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}5$.

18. У 30 телевізорів було виміряно чутливість x_i. Результати вимірювання подано як дискретний статистичний розподіл:

x_i, мкВ	200	250	300	350	400	450	500
n_i	2	7	6	8	4	2	1

Із надійністю γ = 0,99 побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 4$.

19. У 25 випадково відібраних деталей була виміряна відстань у мікронах від центру маси, що міститься на її осі, до зовнішньої поверхні. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

Межі інтервалів x_i, мк	80-96	96-112	112-128	128-144	144-160
n_i	2	5	8	6	4

З надійністю $\gamma = 0{,}999$ побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 3$.

20. З партії однотипних запобіжників відібрано 24 шт. Вимірювання відхилення від номіналу в кілоомах x_i наведено як дискретний статистичний розподіл:

x_i, км	– 1	– 2	1	2	3	4	5
n_i	3	4	4	5	4	3	1

Із надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 3$.

21. З партії однотипних плашок навмання було вибрано 28 шт., і в кожній із них була виміряна глибина пазу (канавки) x_i. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, мм	2,4-2,6	2,6-2,8	2,8-3,0	3,0-3,2	3,2-3,4
n_i	5	8	9	5	1

З надійністю $\gamma = 0{,}999$ побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}8$.

22. 28 однотипних приладів були випробувані щодо їх безвідмовної роботи x_i. Результати вимірювання наведено як дискретний статистичний розподіл:

x_i, год	100	110	120	130	140	150
n_i	10	6	5	4	2	1

З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 4$.

23. У 30 випадково вибраних валиках виміряні відхилення їх діаметрів від номіналу x_i. Результати вимірювань наведено як інтервальний статистичний розподіл:

x_i, мм	5-10	10-15	15-20	20-25	25-30
n_i	2	6	10	8	4

З надійністю $\gamma = 0{,}999$ побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{г} = 0{,}8$ мм.

24. Навмання вибрано 29 різців, які випробувані на знос. Результати експерименту наведено у вигляді дискретного статистичного розподілу:

x_i, год	2	3	4	5	6	7	8
n_i	10	8	6	2	1	1	1

З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{г} = 2$.

25. Залежність собівартості Y одного примірника книги від тиражу X досліджувалась видавництвом. Результати дослідження наведено у вигляді двовимірного статистичного розподілу:

$X = x_i$, тис. прим.	$Y = y_i$, грн.				
	10,15	5,52	4,08	2,85	n_{yi}
1	10	5	5	–	
2	–	15	10	5	
3	–	–	20	10	
4	–	–	5	15	
n_{xj}					

З надійністю $\gamma = 0{,}99$ побудуйте інтервали для $\overline{Y}_{\Gamma}$, σ_{Γ}, r_{xy}. Використовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

26. Залежність річної продуктивності праці в розрахунку на одного робітника y_i від енергомісткості праці x_i на підприємствах однієї галузі наведено в таблиці:

y_i, грн	5,4	5,6	6,2	6,8	7,1	7,8	8,5	9,1	10,5	10,9
x_i, Вт/робітн.	1,8	2,1	2,8	3,0	3,2	3,8	3,9	4,2	4,5	4,8
y_i, грн	11,0	11,6	12,1	12,7	13,2	13,9	14,1	14,6	14,9	15,4
x_i, Вт/робітн.	5,2	5,8	5,9	6,2	6,9	7,2	7,5	8,5	8,8	9,4

Із надійністю $\gamma = 0{,}999$ побудуйте довірчі інтервали для $\overline{Y}_{\Gamma}$, σ_{Γ}, r_{xy}, x_i, y_i. Використовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$

27. Середня температура у квітні у Києві X і Донецьку Y вимірювалась протягом 40 років. Результати вимірювання наведено у вигляді двовимірного статистичного розподілу:

$Y = y_i$	$X = x_i$					
	10	14	18	22	26	n_{yi}
12	2	3	5	–	–	
16	–	8	2	3	2	
18	1	5	2	2	1	
20	–	2	1	1	1	
n_{xj}						

Із надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для $\overline{Y}_{\Gamma}$, σ_{Γ}, r_{xy}.

Використовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

28. На підприємствах однієї галузі промисловості була досліджена залежність річної продуктивності праці одного робітника Y від енергомісткості виробництва X.

Результати дослідження наведено у вигляді парного статистичного розподілу:

y_i, тис. грн/робітн.	2,88	2,91	2,92	2,96	3,01	3,11	3,21	3,25
x_i, кВт/робітн.	2,07	2,12	2,41	2,59	2,89	2,92	3,01	3,12
y_i, тис. грн/робітн.	3,32	3,36	3,42	3,46	3,58	3,88	4,12	
x_i, кВт/робітн.	3,21	3,29	3,31	3,35	3,41	3,48	3,81	

З надійністю $\gamma = 0{,}999$ побудуйте довірчий інтервал для $\overline{Y}_{\Gamma}$, σ_{Γ}, r_{xy}.

Використовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

29. Проводяться випробування міцності 100 волокон залежно від їх товщини.

Результати експериментів задано двовимірним статистичним розподілом:

$Y = y_i$	$X = x_i$, мк					
	4100	4300	4500	4700	4900	n_{yi}
6,75	5	5	10	–	–	
6,25	–	5	10	5	–	
5,75	–	–	5	15	10	
5,25	–	–	5	5	10	
4,75	–	–	–	–	10	
n_{xi}						

З надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для $\overline{Y}_{\Gamma}, \sigma_{\Gamma}, r_{xy}$

Використовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

30. Досліджувалась залежність кількості гризунів y_i, що загинули від концентрації спожитої отрути (в умовних одиницях). Результати досліджень наведено у вигляді парного статистичного розподілу:

y_i	32	36	36	42	46	47	49	55	59
x_i	3	3,5	4	4,5	5	5,5	6	6,5	7
y_i	62	68	70	73	75	88	92	94	98
x_i	7,5	8	8,5	9	9,5	10,5	11	11,5	12

З надійністю $\gamma = 0{,}999$ побудуйте довірчий інтервал для $\overline{Y}_{\Gamma}, \sigma_{\Gamma}, r_{xy}$.

Використовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

31. У навмання вибраних однотипних телевізорах вимірювалась чутливість відео y_i та звукового x_i каналів. Результати перевірки наведено як парний статистичний розподіл:

y_i	250	200	180	160	140	120	110	100	95
x_i	180	230	240	250	300	310	320	330	340
y_i	90	85	80	75	80	70	65	60	55
x_i	350	360	370	380	390	400	410	420	430

З надійністю $\gamma = 0{,}999$ побудуйте довірчий інтервал для $\overline{Y}_{\Gamma}, \sigma_{\Gamma}, r_{xy}$.

Застосовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

32. На заводі «Азовсталь» вимірювався вміст кремнію (у %) у чавуні за різних температур шлаку

$Y = y_i$ ($t°$ C)	S_i (%) $X = x_i$					
	0,27	0,32	0,42	0,51	0,65	n_{yi}
1330	2	1	1	1	–	
1340	–	4	2	3	1	
1345	–	–	3	4	3	
1365	–	–	–	1	4	
n_{xj}						

З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для $\overline{Y}_{\Gamma}, \sigma_{\Gamma}, r_{xy}$.

Застосовуючи нерівність Чебишова, побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

33. У лабораторних умовах здійснювався експеримент з метою визначення залежності кількості речовини, що розчиняється у воді y_i, від температури останньої x_i. Результати експерименту наведено у вигляді двовимірного статистичного розподілу:

$Y = y_i$	$X = x_i$, °C					
	10	20	30	40	50	n_{yi}
48	–	2	3	5	–	
60	2	1	1	1	5	
63	1	2	1	1	–	
71	–	–	2	2	1	
n_{xi}						

З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для $\overline{Y}_{\Gamma}, \sigma_{\Gamma}, r_{xy}$.

Згідно з нерівністю Чебишова побудуйте довірчий інтервал для $\overline{X}_{\Gamma}$.

Тема 6. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

1. *Поняття статистичної гіпотези.*
2. *Загальні правила перевірки статистичних гіпотез;*
3. *Помилки першого і другого роду при перевірці гіпотез.*
4. *Перевірка гіпотез про рівність центрів розподілу двох нормальних генеральних сукупностей при відомій та невідомій дисперсії.*
5. *Перевірка гіпотези про рівність дисперсій двох нормально розподілених генеральних сукупностей.*

6.1. Поняття статистичної гіпотези

Статистичною гіпотезою називається будь-яке *припущення* щодо виду або параметрів розподілу випадкової величини.

Перевірка гіпотези – це оцінювання гіпотези з метою її прийняття або відхилення.

Сформулюємо задачу статистичної перевірки гіпотези у загальному вигляді. Нехай $f(X,\Theta)$ – закон розподілу випадкової величини X, який залежить від одного параметра Θ. Припустимо, що необхідно перевірити гіпотезу про те, що $\Theta = \Theta_0$. Назвемо цю гіпотезу *нульовою* і позначимо H_0.

Гіпотезу про те, що $\Theta = \Theta_1$ назвемо *конкуруючою* або *альтернативною* і позначимо її H_1. Таким чином, необхідно перевірити гіпотезу H_0 відносно конкуруючої гіпотези H_1 на основі вибірки, яка складається з n незалежних спостережень $x_1, x_2, \ldots, x_n$ над випадковою величиною X. Отже, всю можливу множину вибірок обсягу n можна поділити на дві *неперетинні* підмножини O *і* W такі, що *гіпотеза H_0 відхиляється,* якщо вибірка, яка перевіряється, потрапляє у підмножину W, і *приймається,* якщо вибірка належить підмножині O.

Підмножина W – це *критична область, O – область допустимих значень.*

Приймаючи або відхиляючи гіпотезу H_0, можна прийняти правильне рішення або допустити *помилки двох родів.*

Помилка першого роду полягає в тому, що гіпотеза H_0 відхиляється, тобто приймається гіпотеза H_1, коли насправді гіпотеза H_0 все ж є правильною.

Помилка другого роду полягає в тому, що гіпотеза H_0 приймається, коли правильною все ж є гіпотеза H_1.

Імовірність помилок першого і другого роду однозначно визначається вибором критичної області W. Нехай α – ймовірність помилки першого роду, β – ймовірність помилки другого роду.

Тоді частка *хибних висновків* дорівнює α, якщо вірною є гіпотеза H_1.

Загальна схема перевірки статистичних гіпотез.

1. Проводиться вибірка $x_1, x_2, \ldots, x_n$.
2. Формулюється нульова гіпотеза H_0 і альтернативна H_1.
3. Задається рівень значущості α .
4. Обирається тестова статистика (Z – статистика, t-статистика, розподіл Фішера-Снедекора (F-статистика) або ін.).
5. Формулюється правило перевірки гіпотез: гіпотеза H_0 при заданому рівні значущості відхиляється, якщо вибіркове значення x потрапляє у критичну область. У протилежному випадку гіпотеза H_0 приймається.

6.2. Перевірка гіпотези про рівність центрів розподілу двох нормальних генеральних сукупностей при відомій дисперсії

Розглянемо дві випадкові величини X та Y, кожна з яких підпорядкована нормальному закону розподілу. Нехай є дві незалежних вибірки обсягами n_1 та n_2 з генеральних сукупностей X та Y, обчислені вибіркові середні $\bar{x}$ та $\bar{y}$. Необхідно перевірити нульову гіпотезу H_0: $M(X) = M(Y)$ відносно альтернативної гіпотези H_1: $|M(X) - M(Y)| > 0$.

Розглядаємо випадок, коли генеральні дисперсії σ_x^2 та σ_y^2 відомі. Оскільки нам нічого не відомо про математичні сподівання *M(X)* та *M(Y)*, то для перевірки гіпотези H_0 використовують їх найкращі оцінки $\bar{x}$ та $\bar{y}$.

Як відомо (див. розділ 5), середні $\bar{x}$ та $\bar{y}$ мають нормальний закон розподілу з параметрами $[M(X); \sigma_x^2 / n_1]$ та $[M(Y); \sigma_y^2 / n_2]$.

Вибірки є незалежними, тому $\bar{x}$ та $\bar{y}$ також незалежні, і випадкова величина, яка дорівнює різниці між $\bar{x}$ та $\bar{y}$, має нормальний розподіл.

Отже, нормована різниця випадкової величини $\bar{x} - \bar{y}$

$$Z = \frac{|\bar{x} - \bar{y}|}{\sqrt{\sigma_x^2 / n_1 + \sigma_y^2 / n_2}} \tag{6.1}$$

підпорядкована нормальному закону розподілу.

Якщо $\sigma_x^2 = \sigma_y^2$, то

$$Z = \frac{|\bar{x} - \bar{y}|}{\sigma\sqrt{1/n_1 + 1/n_2}}. \tag{6.2}$$

Обираючи ймовірність $p = 1 - \alpha$, можна за таблицями функції Лапласа визначити статистику Z_p, яка поділить множину Z на дві неперетинні підмножини: область допустимих значень і критичну область Z. Ті значення Z, для яких $|Z| \le Z_p$, утворюють область значень; значення Z, для яких $|Z| > Z_p$, визначають критичну область. Тобто критична область визначається нерівністю:

$$\frac{|\bar{x} - \bar{y}|}{\sigma\sqrt{1/n_1 + 1/n_2}} > Z_p. \tag{6.3}$$

Приклад 6.1. Проведемо обстеження двох фанерних підприємств із метою визначення витрат фансировини на виготовлення 1 *м*^{3} фанери. Нехай $n_1 = 40$; $\bar{x} = 2{,}59$ *м*^{3}; $n_2 = 50$; $\bar{y} = 2{,}44$ *м*^{3}. Чи можна з надійністю $p = 0{,}99$ вважати цю різницю випадковою, якщо середнє квадратичне відхилення в обох випадках $\sigma = 0{,}25$ *м*^{3}?

Розв'язання.

1. Нульова гіпотеза $H_0 : M(X) - M(Y) = 0$.

2. Для $p = 0{,}99$, $Z_p = 2{,}58$.

3.

$$Z = \frac{|\bar{x} - \bar{y}|}{\sigma\sqrt{1/n_1 + 1/n_2}} = \frac{2{,}59 - 2{,}44}{0{,}25\sqrt{1/40 + 1/50}} = \frac{0{,}15}{0{,}25\sqrt{0{,}045}} = \frac{0{,}15}{0{,}25 * 0{,}212} = \frac{0{,}15}{0{,}053} = 2{,}83.$$

4. $|Z| = 2{,}83 > Z_p = 2{,}58$.

Отже, з імовірністю 0,99 різницю у витратах фансировини не можна вважати випадковою.

6.3. Перевірка гіпотези про рівність центрів розподілу двох нормально розподілених генеральних сукупностей при невідомій дисперсії

Розглянемо дві випадкові величини *X* та *Y*, кожна з яких підпорядкована нормальному закону розподілу з математичними сподіваннями *M(X)* та *M(Y)*. Будемо вважати, що $\sigma_x^2 = \sigma_y^2 = \sigma^2$. При цьому числове значення σ^2 невідоме.

Нехай є дві незалежні вибірки обсягами n_1 та n_2 відповідно з генеральних сукупностей *X* та *Y*. Необхідно перевірити нульову гіпотезу H_0, яка полягає в тому, що *M(X) = M(Y)*, відносно альтернативної гіпотези H_1, яка полягає в тому, що $|M(X) - M(Y)| > 0$.

Для оцінки *M(X)* та *M(Y)* використовують їх найкращі оцінки за вибіркою $\bar{x}$ та $\bar{y}$, а для оцінки σ^2 використовують вибіркові

$$S_x^{2*} = \frac{\sum_{i=0}^{n}(x_i - \bar{x})^2}{n_1 - 1}, \quad S_y^{2*} = \frac{\sum_{i=0}^{n}(y_i - \bar{y})^2}{n_2 - 1}.$$

Оскільки генеральні сукупності *X* та *Y* мають однакові дисперсії, то для σ^2 доцільно використати результати обидвох вибірок. У математичній статистиці доведено, що найкращою оцінкою для σ^2 в даному випадку є

$$S^{2*} = \frac{S_x^{2*}(n_1 - 1) + S_y^{2*}(n_2 - 1)}{n_1 + n_2 - 1}. \qquad (6.4)$$

Оскільки величина σ^2 невідома, а випадкова величина підпорядкована нормальному закону, то гіпотеза H_0 оцінюється за допомогою t-статистики:

$$t = \frac{\bar{x} - \bar{y}}{S^* \sqrt{1/n_1 + 1/n_2}}. \tag{6.5}$$

Обравши довірчу ймовірність $p = 1 - \alpha$, за таблицями t – розподілу визначають табличне значення t_{p, n_1+n_2-2}, яке визначає межі прийняття гіпотези H_0 та критичної області. Якщо

$$|t| > t_{p, n_1+n_2-2},$$

то з імовірністю $p = 1 - \alpha$ можна вважати розходження середніх значень значущим (невипадковим), і гіпотеза H_0 відхиляється.

Приклад 6.2. Нехай в умовах попередньої задачі σ_x^2 і σ_y^2 не відомі, але ми припускаємо їх рівними. Необхідно перевірити гіпотезу H_0 про рівність середніх з імовірністю *0,99.*

Розв'язання.

1. Нехай $S_x^* = 0{,}28$; $S_y^* = 0{,}22$, тоді

$$S^* = \sqrt{\frac{0{,}28^2(40-1) + 0{,}22^2(50-1)}{40+50-2}} = \sqrt{\frac{0{,}078 * 39 + 0{,}048 * 49}{88}} =$$

$$= \sqrt{\frac{3{,}06 + 2{,}37}{88}} = \sqrt{\frac{5{,}43}{88}} = \sqrt{0{,}062} = 0{,}25$$

2. $t = \frac{2{,}59 - 2{,}44}{0{,}25} = \frac{0{,}15}{0{,}25 * 0{,}212} = 2{,}83.$

3. $t_{0{,}99;88} = 2{,}62.$

4. $|t| = 2{,}83 > 2{,}62$, отже, гіпотеза H_0 про рівність середніх відхиляється.

6.4. F-розподіл і перевірка гіпотези про рівність дисперсій двох нормальних генеральних сукупностей

Гіпотези про дисперсії мають особливо велике значення в техніці та економіці, оскільки величина розсіяння, яка вимірюється дисперсією, характеризує такі надзвичайно важливі показники, як точність машин, приладів, технологічних процесів, похибку показів вимірювальних пристроїв, точність розрахунків тощо.

Сформулюємо гіпотезу про рівність дисперсій двох нормальних генеральних сукупностей. Нехай є дві випадкові величини X та Y, кожна з яких підпорядковується нормальному закону розподілу, з дисперсіями σ_x^2 та σ_y^2. Нехай з генеральних сукупностей X та Y вибрані дві незалежні вибірки обсягами n_1 та n_2. Перевіримо гіпотезу H_0 про те, що $\sigma_x^2 = \sigma_y^2$ відносно альтернативної гіпотези H_1: $\sigma_x^2 > \sigma_y^2$. В якості оцінок σ_x^2 та σ_y^2 використовують вибіркові дисперсії S_x^{2*} та S_y^{2*} відповідно.

Для побудови критичної області необхідно дослідити спільний закон розподілу оцінок S_x^{2*} та S_y^{2*}. Таким спільним законом розподілу статистик S_x^{2*} та S_y^{2*} є *F-розподіл Фішера-Снедекора.* Випадкова величина F, яка визначається відношенням

$$F=\frac{n_1}{n_2}\cdot\frac{S_1^{2*}\cdot n_1/\sigma^2}{S_2^{2*}\cdot n_2/\sigma^2}=\frac{S_1^{2*}}{S_2^{2*}}, \qquad (6.6)$$

називається *випадковою величиною з розподілом Фішера-Снедекора.* Зауважимо, що завжди можна так ввести позначення, щоб $S_x^{2*} \geq S_y^{2*}$, тому випадкова величина F приймає значення, які не менші за одиницю. Для різних значень рівня значущості α (довірчої ймовірності $P = 1 - \alpha$) і ступенів свободи $k_1 = n_1 - 1$ *та* $k_2 = n_2 - 1$ розроблені таблиці *F-розподілу* (див. додаток 3), за допомогою яких визначають межу області прийняття гіпотези та критичної області.

Обравши необхідний рівень значущості α, за таблицями F-розподілу знаходимо число F_{α,k_1,k_2}, яке порівнюється з розрахованим F. Якщо $F > F_{\alpha,k_1,k_2}$, то гіпотеза H_0 відхиляється, якщо $F < F_{\alpha,k_1,k_2}$, то вибіркові спостереження не суперечать гіпотезі, що перевіряється.

Приклад 6.3. Обробка меблевих заготовок відбувається на двох токарних верстатах. Обсяги виборок: $n_1 = 10$ *шт.*, $n_2 = 15$ *шт.* Вибіркові дисперсії діаметрів заготовок: $S_1^{2*} = 0{,}096$ *мм*^{2}, $S_2^{2*} = 0{,}057$ *мм*^{2}. Визначити з імовірністю $P = 0{,}95$, чи мають верстати однакову точність.

Розв'язання.

1. Нульова гіпотеза H_0: $\sigma_x^2 = \sigma_y^2$.

2. Розрахункове значення $F = \frac{0{,}096}{0{,}057} = 1{,}684$.

3. $k_1 = 10 - 1 = 9$; $k_2 = 15 - 1 = 14$. Табличне значення $F_{0,05;9;14} = 2{,}65$.

4. $F = 1{,}684 < 2{,}65$, тобто $F < F_{0,05;\,9;14}$.

5. Немає підстав відхиляти нульову гіпотезу. Верстати вважаємо однаково точними.

Запитання та завдання для самоконтролю

1. Дайте визначення статистичної гіпотези.
2. Сформулюйте задачу статистичної перевірки гіпотези в загальному вигляді.
3. У чому полягають помилки першого і другого роду?
4. У чому полягає загальна схема перевірки статистичних гіпотез ?
5. Складіть алгоритм перевірки гіпотези про рівність центрів розподілу двох сукупностей при відомій дисперсії.
6. Складіть алгоритм перевірки гіпотези про рівність центрів розподілу двох сукупностей при невідомій дисперсії.
7. Охарактеризуйте F-розподіл Фішера-Снедекора.
8. Сформулюйте правила перевірки гіпотези про рівність дисперсій двох сукупностей.
9. Проведено 25 незалежних вимірювань випадкової величини X, що має нормальний закон розподілу зі значенням $\sigma_Г = 2$:

x_i	2,4	5,4	8,4	11,4	14,4	17,4
n_i	2	3	10	6	3	1

При рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : M(X) = 10{,}5$, якщо альтернативна гіпотеза $H_0 : M(X) < 10{,}5$.

10. Маємо дані про розподіл підприємств певної галузі за зростанням виробітку на одного працівника у відсотках до наступного року:

x_i, %	75	85	95	105	115	125
N_i	5	8	10	5	2	1

Враховуючи, що ознака має нормальний закон розподілу зі значенням $\sigma_Г = 6$, перевірте правильність нульової гіпотези при $\alpha = 0{,}01$:

$H_0 : M(X) = 90$, якщо альтернативна гіпотеза $H_a : M(X) \neq 90$.

11. У результаті двадцяти незалежних вимірювань певної величини X отримано статистичний розподіл:

x_i	3,4	6,4	9,4	12,4	15,4	18,4
n_i	2	4	8	3	2	1

Припускаючи, що випадкова величина X має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірте правильність:

$H_0 : M(X) = 10$, якщо альтернативна гіпотеза $H_a : M(X) > 10$.

12. Результати вимірювання зросту дівчаток віком 16 років дали такі показники:

$h = 4$, см	160–164	164–168	168–172	172–176	176–180
n_i	4	6	20	4	2

Вважаючи, що випадкова величина X – зріст дівчаток – має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : M(X) = 180$, якщо альтернативна гіпотеза $H_a : M(X) \neq 180$.

13. Рівноточні вимірювання довжини двадцяти однотипних деталей дали такі результати:

x_i, мм	122,8	128,8	134,8	140,8	146,8
n_i	2	6	8	3	1

Вважаючи, що випадкова величина X – довжина деталі – має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : M(X) = 144$, якщо альтернативна гіпотеза $H_a : M(X) < 144$.

14. Вимірювалась швидкість руху автомобілів x_i на певній ділянці шляху. Результати вимірів наведено в таблиці:

x_i, км/год	56	60	64	68	72	70	80
n_i	2	4	6	8	3	1	1

Вважаючи, що x_i – швидкість автомобіля – є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірте правильність нульової гіпотези $H_0 : M(X) = 70$, якщо альтернативна гіпотеза $H_a : M(X) \neq 70$.

15. Маса 100 шарикопідшипників наведена у вигляді статистичного розподілу:

x_i, мг	148	150	152	154	156	158	160
n_i	2	4	14	30	40	8	2

Беручи до уваги, що випадкова величина X – маса шарикопідшипників – має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : M(X) = 159$, якщо альтернативна гіпотеза $H_a : M(X) \neq 159$.

16. Вимірювання барометром атмосферного тиску протягом 100 діб дали такі результати:

x_i, мм рт. ст.	744,4	746,4	748,4	750,4	752,4	754,4
n_i	10	20	30	20	15	5

Вважаючи, що X – атмосферний тиск – є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості

$\alpha = 0{,}01$ перевірте правильність нульової гіпотези $H_0 : M(X) = 749{,}2$, якщо альтернативна гіпотеза $H_a : M(X) > 749{,}2$.

17. Електролампочки на 220 В виготовлялися двома електроламповими заводами. З першої партії, виготовленої заводом № 1, здійснили вибірку обсягом $n' = 25$, а з другої партії – обсягом $n'' = 36$. Першу і другу партії електролампочок перевірили на тривалість роботи.

Результати перевірки наведено у вигляді статистичних розподілів такого вигляду:

y_i	48	50	52	54	56
n'_i	2	3	14	5	1

x_j	53	56	59	62	65
n''_j	4	6	10	12	4

Відомо, що ознаки Y – тривалість роботи електролампочки першого заводу і X – тривалість роботи електролампочки другого заводу – є випадковими незалежними величинами і мають нормальний закон розподілу зі значеннями $\sigma_y = 50$, $\sigma_x = 72$. При рівні значущості $\alpha = 0{,}01$ перевірте правильність $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) > M(Y)$.

18. У двох партіях містяться однотипні шарикопідшипники, виготовлені двома заводами. Вимірювання їх діаметрів дали результати, які наведено у вигляді двох статистичних розподілів:

y_i, мм	6,64	6,7	6,74	6,78	6,82
n'_i	2	4	8	6	4

x_j, мм	6,58	6,6	6,8	7	7,2
n''_j	6	8	10	4	2

При рівні значущості $\alpha = 0{,}01$ перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) \neq M(Y)$, коли відомі значення $D_x = 50$; $D_y = 60$.

19. З двох партій монет вартістю 5 коп. було вибрано 50 і 60 штук, які зважували на терезах.

Результати цих зважувань подано у вигляді двох статистичних розподілів:

y_i, мг	9,4	9,6	9,8	10	10,2

n_i'	5	15	20	8	2
x_j,мг	9,33	9,63	9,63	10,23	10,53
n_j''	8	12	26	10	4

Припускаючи, що X і Y мають нормальний закон розподілу і незалежні між собою, при рівні значущості $a = 0{,}01$ перевірте $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) < M(Y)$, коли відомі значення $D_x = 10$; $D_y = 14$.

20. Вимірювання зросту дітей віком шість років, випадково вибраних із двох дитячих садків, дало такі результати:

y_i, м	0,52	0,58	0,64	0,72	0,8
n_i'	2	5	10	3	1

x_j, м	0,48	0,56	0,64	0,72	0,8
n_j''	1	4	12	6	2

Беручи до уваги, що випадкові величини X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) > M(Y)$.

21. Кров'яний тиск було виміряно (в умовних одиницях) у 20 осіб віком 40 років із одного району міста і в 18 осіб того самого віку з іншого району міста. Результати вимірювання подано двома статистичними розподілами:

y_i	114	116	118	120	122	124
n_i'	2	4	6	5	2	1
x_j	115	118	121	124	127	130
n_j''	1	3	6	4	3	1

Припускаючи, що випадкові величини X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірте правильність $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) \neq M(Y)$.

22. Пружність вимірювалась на зразках, виготовлених з однієї і тієї самої марки сталі і вибраних із двох партій.

Результати вимірювання подано двома статистичними розподілами:

y_i	36,8	38,8	40,8	42,8	44,8
n'_i	2	4	6	5	3

x_j	34,2	38,2	42,2	46,2	50,2
n''_j	2	5	10	4	4

Зважаючи, що ознаки X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) < M(Y)$.

23. Протягом року вимірювалась продуктивність праці (в тис. грн/працівн.) у двох будівельних фірмах. Результати вимірювання подано статистичними розподілами:

y_i	120	150	180	210	240	270
n'_i	10	20	30	20	15	5
x_j	90	130	170	210	250	290
n''_j	10	20	40	20	5	5

Вважаючи, що ознаки X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) > M(Y)$.

24. Досліджувався місячний прибуток у гривнях робітників двох заводів однієї і тієї самої галузі виробництва. Результати досліджень подано двома статистичними розподілами:

y_i	150,6	160,6	170,6	180,6	190,6
n'_i	12	28	40	18	2
x_j	140,8	160,8	180,8	200,8	220,8
n''_j	2	6	32	8	2

Ознаки X і Y є незалежними і мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}01$ перевірте правильність нульової гіпотези $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) < M(Y)$.

25. Норма витрат на технічне обслуговування і ремонт нових марок тракторів вимірювалась у двох сільських господарствах району. Результати вимірювань показано двома статистичними розподілами:

y_i, грн/га	0,58	0,6	0,62	0,64	0,66
n'_i	2	3	10	4	1

x_j, грн/га	0,56	0,6	0,64	0,7	0,74
n''_j	4	6	3	2	1

Ознаки X і Y (норми витрат) є незалежними випадковими величинами, що мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : D_x = D_y$, якщо альтернативна гіпотеза $H_a : D_x > D_y$.

26. Визначалися річні середні витрати електроенергії на комунально-побутові вимоги для одного мешканця у двох містах. Результати розрахунків подано двома статистичними розподілами для першого і другого міст:

y_i, Вт/м.	700	708	716	724	732	740	
n'_i	5	6	9	6	3	1	
x_j, Вт/м.	706	710	714	718	722	726	730
n''_j	8	10	12	5	2	2	1

Ознаки X і Y (річні витрати в кВт/особу) є незалежними між собою і мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : D_x = D_y$, якщо альтернативна гіпотеза $H_a : D_x < D_y$.

27. Вимірювалось споживання масла за одну добу одним мешканцем у двох регіонах країни.

Результати вимірювання подано двома статистичними розподілами:

y_i, мг	15,99	18,99	21,99	21,99	24,99
n'_i	4	6	20	10	5

x_j, мг	14,55	20,55	26,55	30,55	36,55
n''_j	6	14	16	6	4

Ознаки X і Y (добове споживання масла в мг) є незалежними випадковими величинами, які мають нормальний закон розподілу ймовірностей.

При рівні значущості $\alpha = 0{,}001$ перевірте правильність нульової гіпотези $H_0 : D_x = D_y$, якщо альтернативна гіпотеза $H_a : D_x > D_y$.

28. Вимірювання маси в грамах пухових волокон від овець подано двома статистичними розподілами:

y_i, г	4,44	4,84	5,24	6,64	6,04
n'_i	2	4	5	8	1
x_j, г	4,36	4,96	5,46	5,96	6,46
n''_j	3	5	8	6	4

Ознаки X і Y (маса волокон в грамах) є незалежними випадковими величинами, які мають нормальний закон розподілу.

При рівні значущості $\alpha = 0{,}01$ перевірте правильність нульової гіпотези $H_0 : D_x = D_y$, якщо альтернативна гіпотеза $H_a : D_x > D_y$.

Тема 7. ОСНОВИ КОРЕЛЯЦІЙНОГО АНАЛІЗУ

1. Функціональні, стохастичні (статистичні) та кореляційні залежності.

2. Коефіцієнт кореляції.

3. Оцінка значущості коефіцієнта кореляції.

4. Кореляційне відношення, коефіцієнт детермінації та стандартна похибка оцінки.

5. Основи множинного кореляційного аналізу.

7.1. Функціональні, стохастичні (статистичні) та кореляційні залежності

Зв'язки між різними явищами у природі та суспільстві є складними та різноманітними, натомість їх можна певним чином класифікувати.

У техніці та природознавчих науках часто розглядається *функціональна залежність (зв'язок)* між змінними X та Y, коли кожному можливому значенню однієї змінної поставлено у відповідність деяке значення іншої (наприклад, залежність між тиском та об'ємом газу тощо).

В економіці для більшості явищ між змінними величинами існують залежності, коли кожному значенню однієї змінної відповідає не одне певне значення, а множина можливих значень іншої змінної. Тобто кожному значенню однієї змінної відповідає *деякий (умовний) розподіл* іншої змінної. Така залежність називається *статистичною* (*стохастичною, імовірнісною).*

Виникнення поняття статистичної залежності обумовлено тим, що на залежну змінну можуть впливати деякі неконтрольовані та неврахованi фактори. Наприклад, відомо, що із збільшенням обсягів виробництва продукції X собівартість одиниці продукції Y знижується. Натомість фактичний рівень собівартості залежить також від інших факторів: вартості сировини, транспортних витрат, вартості основних фондів, кваліфікації персоналу тощо. Кожен із цих факторів для різних підприємств досить обґрунтовано можна вважати випадковим. Іншим прикладом статистичного зв'язку є, наприклад, залежність продуктивності праці на підприємстві від фондоозброєності. Проте очевидно, що на продуктивність праці впливають також такі фактори, як енергоозброєність, кваліфікація персоналу тощо.

Найбільший інтерес для дослідника становить частковий випадок статистичного зв'язку, коли *умовне математичне сподівання* однієї випадкової змінної є функцією значення, якого набуває інша випадкова змінна, тобто

$$M\,(Y/X) = x = f(x). \qquad (7.1)$$

Умовне математичне сподівання $M\,(Y/X) = x$ або $M_x(Y)$ – це математичне сподівання випадкової змінної Y, обчислене в припущенні, що змінна X прийняла значення x.

Якщо кожному значенню однієї змінної відповідає певне умовне математичне сподівання (або середнє значення) іншої, то така статистична залежність називається *кореляційною.*

Отже, *кореляційною залежністю* між двома змінними називається функціональна залежність між значеннями однієї з них і умовним математичним сподіванням іншої.

Кореляційна залежність може бути також представлена у вигляді

$$M_x(У) = \varphi(x)$$

або

$$M_y(X) = \Psi(y), \qquad (7.2)$$

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

Питання про те, що вибрати залежною змінною, а що незалежною, необхідно вирішувати у кожному конкретному випадку.

Кореляційна залежність широко застосовується в економічних дослідженнях. Вона характеризується *формою* і *щільністю (тіснотою, силою) зв'язку.*

Кореляційний аналіз – це спеціальний математичний апарат для визначення наявності і щільності зв'язку між випадковими величинами.

Основне завдання кореляційного аналізу полягає у визначенні *коефіцієнта кореляції (парної* при аналізі двох випадкових величин і *множинної,* якщо їх більше ніж дві).

Кореляційне поле – це координатна площина з нанесеними на неї точками, які відповідають значенням випадкових величин *X* та *Y* (рис. 18).

За розташуванням точок на кореляційному полі можна в першому наближенні зробити припущення про наявність, форму і щільність зв'язку між випадковими величинами. Так, очевидно, що із збільшенням значень *x* (обсяги виробництва) значення *y* (собівартість одиниці продукції) зменшується.

Отже, можна зробити попередній висновок, що має місце обернена залежність між *X* та *Y*.

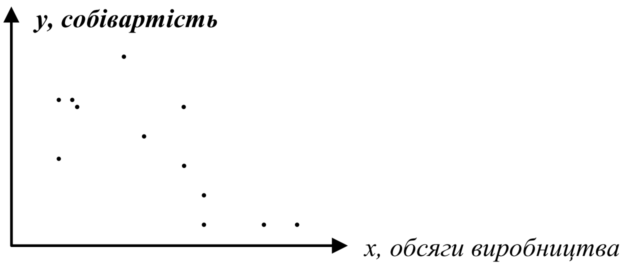


Рис. 18. Кореляційне поле

7.2. Коефіцієнт кореляції

Перейдемо до оцінки *щільності (тісноти, сили)* кореляційної залежності між двома випадковими величинами *X* та *Y*. Перш за все, розглянемо дуже важливий для теорії і практики випадок *лінійного* кореляційного зв'язку між випадковими величинами виду

$$Y = B_0 + B_1X \qquad (7.3)$$

Найбільш точною мірою тісноти зв'язку між величинами *X* та *Y* у даному випадку є *коефіцієнт кореляції*.

Так само, як дисперсія і коваріація, коефіцієнт кореляції має дві форми – *теоретичну* і *вибіркову*. *Теоретичний коефіцієнт кореляції* традиційно позначається буквою *«ρ»* і визначається так:

$$\rho_{xy}=\frac{pop.cov(x,y)}{\sqrt{pop.var(x)\cdot pop.var(y)}}=\frac{\sigma_{x,y}}{\sqrt{\sigma_x^2\sigma_y^2}}. \qquad (7.4)$$

Вибірковий коефіцієнт кореляції r визначається шляхом заміни теоретичних дисперсій і коваріації у виразі (7.4) на їх незміщені оцінки. Як було показано в темі 5, незміщені оцінки можуть бути визначені шляхом множення вибіркових дисперсій та коваріації на поправку Бесселя *n/(n-1)*. Отже,

$$r_{xy}=\frac{\frac{n}{n-1}\operatorname{cov}(x,y)}{\sqrt{\frac{n}{n-1}S_x^{2*}\frac{n}{n-1}S_y^{2*}}}. \qquad (7.5)$$

Звідси

$$r_{xy}=\frac{\sum_{i=0}^{n}(x_i-\bar{x})(y_i-\bar{y})}{nS_xS_y}. \qquad (7.6)$$

Вибірковий коефіцієнт кореляції r (при досить великому обсязі вибірки n) і теоретичний коефіцієнт кореляції ρ мають такі *властивості:*

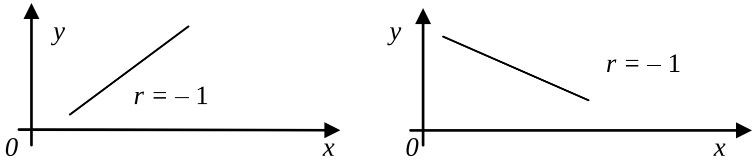
1. Коефіцієнт кореляції приймає значення на відрізку [-1, 1], тобто $-1 \le r \le 1$.

Чим ближче |r| до одиниці, тим сильнішим є зв'язок.

2. При $|r| = 1$ кореляційний зв'язок перетворюється у лінійну функціональну залежність (рис. 19).

3. Додатне значення коефіцієнта кореляції $(r > 0)$ свідчить про *прямий* зв'язок між показниками (рис. 19*а*), від'ємне $(r < 0)$ – про *обернений* (рис. 19*б*).

4. При $r = 0$ лінійний кореляційний зв'язок між спостереженнями x та y у вибірці відсутній; $\rho = 0$, якщо випадкові величини X та Y незалежні. З *незалежності двох випадкових величин випливає їх некорельованість,* тобто рівність $\rho = 0$. Проте некорельованість двох випадкових величин ще не означає їх незалежності. Зрозуміло, той факт, що $r = 0$, не обов'язково означає, що $\rho = 0$, і навпаки.



а) прямий зв'язок *б) обернений зв'язок*

Рис. 19. Кореляційний зв'язок при |r| = *1*

5. Щільність статистичного зв'язку визначається абсолютним значенням коефіцієнта кореляції. Прийнято вважати, що при: $|r|<0,25 \div 0,3$ залежність між x та y слабка; при $|r| = 0,4 \div 0,6$ залежність середня; при $|r|=0,7 : 0,8$ залежність сильна, її необхідно враховувати; при $|r|>0,8$ залежність є достатньо щільною, щоби дослідити її форму.

Необхідно зауважити, що r є безпосередньою оцінкою теоретичного коефіцієнта кореляції ρ лише у випадку двовимірного нормального закону розподілу випадкових величин X та Y. *В* інших випадках вибірковий коефіцієнт кореляції не можна розглядати як строгу міру взаємозв'язку змінних величин.

7.3. Оцінка значущості коефіцієнта кореляції

Значення теоретичного коефіцієнта кореляції ρ, як правило, невідомі. На практиці визначається його *точкова оцінка* – вибірковий коефіцієнт кореляції r на підставі обмеженої випадкової вибірки обсягом n спостережень. Тому необхідно оцінити *значущість коефіцієнта кореляції r.* Для цього формулюється нульова гіпотеза про те, що лінійний кореляційний зв'язок між змінними X та Y відсутній, тобто $H_0 : \rho_{xy} = 0$. Припускається, що випадкові величини X та Y підпорядковані нормальному закону розподілу. У даному випадку використовується тестова статистика Стьюдента з $(n-2)$ ступенями свободи

$$|t| = r\sqrt{\frac{n-2}{1-r^2}}\,. \tag{7.7}$$

Для перевірки нуль-гіпотези обирається рівень значущості α і визначається табличне критичне значення $t_{\alpha,к}$ з $(n\text{-}2)$ ступенями свободи. Табличне значення $t_{\alpha,к}$ порівнюється з розрахунковим t.

Якщо $|t| < t_{\alpha,к}$, то нуль-гіпотеза приймається, тобто з рівнем значущості α (довірчою ймовірністю $p_a=1\text{-}\alpha$) можна стверджувати, що змінні X та Y лінійно некорельовані, $\rho_{xy} = 0$.

Якщо $|t| > t_{\alpha,к}$, то нуль-гіпотеза відхиляється та із заданим рівнем значущості α можна вважати, що величини X та Y є лінійно кореляційно залежними (додатно або від'ємно).

Нехай обрано рівень значущості 5% і знайдено критичне значення t з $(n\text{-}2)$ ступенями свободи. Це означає, що коли вірною є нульова гіпотеза, то величина t буде перевищувати його критичне значення (в додатному або від'ємному напрямку) лише у 5% випадків. Тобто, при виконанні перевірки ймовірність припущення помилки першого роду, що відхиляє нульову гіпотезу, коли вона фактично є правильною, складає 5%.

Якщо ризик припущення помилки 5% є завеликим для даної задачі, то ступінь ризику можна скоротити, здійснюючи розрахунки при рівні значущості, наприклад, 1%.

Критичне значення t тепер буде вищим, ніж у попередньому випадку, тому необхідна більш висока (додатна або від'ємна) t-статистика для відхилення нуль-гіпотези, а це означає, що необхідне більш високе абсолютне значення r і (або) збільшення обсягу вибірки n.

<u>*Приклад 7.1.*</u> Нехай є дані про обсяги продукції X та собівартість одиниці продукції на лісопильних підприємствах Y, $n = 11$, $r_{xy} = -0{,}7$. Необхідно перевірити значущість коефіцієнта кореляції при рівнях значущості 5% та 1%.

<u>*Розв'язання.*</u> Перевіряється нуль-гіпотеза: H_0: $\rho_{xy} = 0$, альтернативна гіпотеза $H_1 : \rho_{xy} \neq 0$. Визначаємо тестову статистику:

$$t = \frac{(-0{,}7)\sqrt{11-2}}{\sqrt{\quad}} = -2{,}937.$$

За таблицями Стьюдента знаходимо теоретичні значення t-статистик з $11 - 2 = 9$ ступенями свободи та 5% і 1% рівнями значущості: $t_{0,05;9} = 2{,}262$; $t_{0,01;9} = 3{,}250$.

Для рівня значущості 5%: $|t| = 2{,}937 > 2{,}262$.

Отже, нуль-гіпотеза відхиляється і приймається альтернативна гіпотеза, що $\rho_{xy} \neq 0$.

Для рівня значущості 1%: $|t| = 2{,}937 < 3{,}250$, тобто нуль-гіпотеза приймається.

Це означає, що стосовно 11 обстежених підприємств можна з довірчою ймовірністю P_a=1-0,05=0,95 стверджувати, що існує сильний обернений лінійний кореляційний зв'язок між обсягами виробництва і собівартістю пиломатеріалів. Натомість із довірчою ймовірністю $P_a = 1 - 0,01 = 0,99$ вже не можна стверджувати, що існує такий зв'язок між досліджуваними показниками.

Збільшимо кількість обстежених підприємств до 18, тобто $n = 18$.

Розрахуємо нове значення тестової статистики

$$t = \frac{(-0,7)\sqrt{18-2}}{\sqrt{}} = -3,9.$$

Теоретичні значення t-статистик для $18 - 2 = 16$ ступенів свободи: $t_{0,05;16} = 2,120$; $t_{0,01;16} = 2,921$.

$$|t| = 3,9 > 2,120; \ |t| = 3,9 > 2,921.$$

Отже, щодо *18* обстежених підприємств із ризиком як *5%*, так і *1%*, можна стверджувати про наявність сильного оберненого лінійного кореляційного зв'язку між досліджуваними показниками.

Необхідно зауважити, що даний тест є прийнятним лише для нульової гіпотези про відсутність лінійної кореляційної залежності.

7.4. Розкладання дисперсій. Кореляційне відношення. Коефіцієнт детермінації. Стандартна похибка оцінки

Значення коефіцієнта кореляції визначається на основі припущення про *лінійність* зв'язку між випадковими величинами. Якщо ж досліджувана залежність відхиляється від лінійної, то коефіцієнт кореляції втрачає свій зміст як характеристика щільності зв'язку. Тому для вивчення та оцінки сили зв'язку між випадковими величинами у *загальному випадку* використовують інші показники – *коефіцієнт детермінації* і *кореляційне відношення.*

Для визначення показників сили зв'язку в загальному випадку розглянемо розкладання (декомпозицію) дисперсій. Як відомо з дисперсійного аналізу, загальна (повна) дисперсія випадкової величини складається з поясненої за допомогою рівняння регресії дисперсії, яка вимірює вплив фактора *X* на *Y*, і непоясненої (залишкової) дисперсії, яка характеризує вплив інших факторів, тобто

$$\sigma_Y^2 = \sigma_R^2 + \sigma_{YX}^2, \qquad (7.8)$$

де σ_Y^2 – повна дисперсія випадкової величини *Y;* σ_R^2 – дисперсія, що пояснюється рівнянням регресії; σ_{YX}^2 – непояснена або залишкова дисперсія.

Частка поясненої дисперсії в загальній дисперсії називається *відношенням* або *коефіцієнтом детермінації (*η^2_{YX}*),* тобто

$$\eta_{YX}^2 = \frac{\sigma_R^2}{\sigma_Y^2}. \qquad (7.9)$$

Величина η_{YX}^2 – це теоретичний коефіцієнт детермінації.

Поділивши обидві частини рівності (7.8) на σ_Y^2, отримаємо:

$$\frac{\sigma_Y^2}{\sigma_Y^2} = \frac{\sigma_R^2}{\sigma_Y^2} + \frac{\sigma_{YX}^2}{\sigma_Y^2}$$

або

$$1 = \eta_{YX}^2 + \frac{\sigma_{YX}^2}{\sigma_Y^2}.$$

З останньої формули отримаємо інший вираз для η_{YX}^2:

$$\eta_{YX}^2 = 1 - \frac{\sigma_{YX}^2}{\sigma_Y^2}. \qquad (7.10)$$

Оскільки $\sigma_{YX}^2 \le \sigma_Y^2$ (σ_{YX}^2 є складовою частиною σ_Y^2), з рівності (7.10) випливає, що значення η_{YX}^2 завжди знаходиться між нулем і одиницею (рис. 20*a*).

З рівності (7.10) також випливає, що η_{YX}^2 = 1 тільки тоді, коли σ_{YX}^2 = 0, тобто відсутній вплив інших факторів і весь розподіл зосереджений на кривій регресії *y(x).*

У цьому випадку між змінними Y та X існує функціональна залежність (рис. 20*б*).

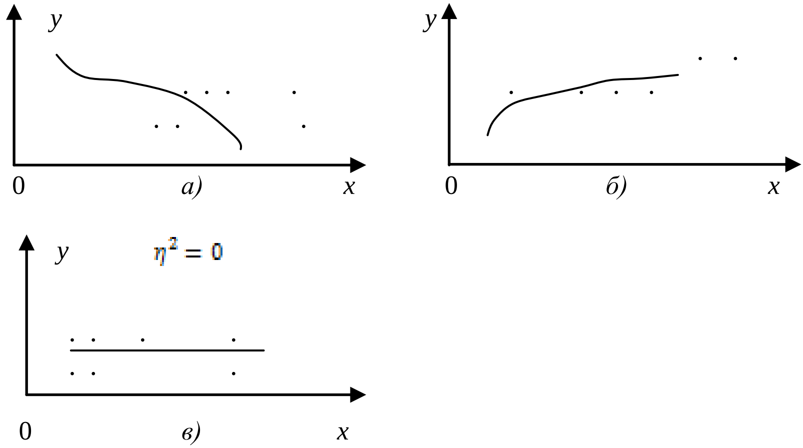


Рис. 20. Коефіцієнт детермінації. Графічна інтерпретація

Далі, з рівності (7.10) випливає, що $\eta_{YX}^2 = 0$ тоді і тільки тоді, коли $y(x)=\mu_Y = const$, тобто лінія регресії Y відносно X є горизонтальною прямою (рис. 20*в*).

Доведено, що у випадку парної лінійної залежності коефіцієнт детермінації дорівнює квадрату коефіцієнта кореляції [1, с. 409; 2, с. 70]:

$$\eta_{YX}^2 = \rho_{YX}^2. \qquad (7.11)$$

Корінь квадратний з коефіцієнта детермінації називається *кореляційним відношенням* η_{YX} залежної змінної Y по незалежній змінній X. Отже, *теоретичне кореляційне відношення* визначається як

$$\eta_{YX} = \sqrt{\frac{\sigma_R^2}{\sigma_Y^2}} = 1- \sqrt{\frac{\sigma_{YX}^2}{\sigma_Y^2}}. \qquad (7.12)$$

Кореляційне відношення, за означенням, є *невід'ємною величиною* і може знаходитися лише в *межах від нуля до одиниці,* оскільки є результатом арифметичного значення квадратного кореня з коефіцієнта детермінації.

На відміну від коефіцієнта кореляції, *кореляційне відношення є несиметричним щодо досліджуваних змінних,* тобто $\eta_{YX} \neq \eta_{XY}$.

Можна показати, що кореляційне відношення η не може бути меншим від абсолютної величини коефіцієнта кореляції ρ, який характеризує залежність між тими ж змінними Y та X $(\eta_{YX} \geq \rho_{XY})$. У випадку лінійної залежності ці дві характеристики збігаються. Це дозволяє використовувати величину різниці $(\eta_{YX}^2 - \rho_{XY}^2)$ як міру відхилення регресійної залежності від лінійного вигляду [1, с. 409].

Нарешті, всі зауваження щодо змістової інтерпретації властивостей коефіцієнта кореляції (ρ та r) стосуються і кореляційного відношення.

До цього часу при розкладанні дисперсій, визначенні коефіцієнта детермінації та кореляційного відношення ми оперували *теоретичними* характеристиками. Натомість у статистичному аналізі їх замінюють на відповідні *вибіркові оцінки.*

Замінимо у формулі (7.8) дисперсії відповідними вибірковими оцінками. В результаті отримаємо:

$$S_Y^2 = S_R^2 + S_{YX}^2. \qquad (7.13)$$

S_Y^2 – оцінка загальної дисперсії:

$$S_Y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}; \qquad (7.14)$$

S_R^2 – оцінка дисперсії, що пояснює регресію:

$$S_R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{n}, \qquad (7.15)$$

де $\hat{y}_i$ – теоретичні значення залежної змінної, визначені за допомогою лінійного рівняння регресії:

$$\hat{y}_i = b_0 + b_1 x_1, i = 1,2 \ldots, n; \qquad (7.16)$$

S^2_{yx} – оцінка залишкової дисперсії (дисперсії помилок):

$$S^2_{yx} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}. \qquad (7.17)$$

З урахуванням формул (7.14), (7.15), (7.17) формулу (7.13) можна записати так:

$$\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}. \qquad (7.18)$$

Помноживши обидві частини рівняння (7.18) на *n,* отримаємо вираз:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \qquad (7.19)$$

Ліва частина рівняння (7.19) – це *загальна сума квадратів відхилень* залежної змінної від її вибіркового середнього значення. Перший член у правій частині рівняння є *поясненою* (за допомогою рівняння регресії) *сумою квадратів,* а другий член – непоясненою, залишковою сумою квадратів або сумою квадратів помилок [14, с. 110].

Отже, у скороченому вигляді рівняння (7.19) можна записати:

$$TSS = ESS + RSS. \qquad (7.20)$$

Вибіркове (емпіричне) значення коефіцієнта детермінації (традиційно позначається R^2) можна записати за допомогою таких еквівалентних формул:

$$R^2 = \frac{S^2_R}{S^2_y} = 1 - \frac{S^2_{yx}}{S^2_y} \qquad (7.21)$$

або

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}. \qquad (7.22)$$

Вибіркове (емпіричне) кореляційне відношення R визначається за формулами:

$$R^2 = \sqrt{\frac{S_R^2}{S_y^2}} = \sqrt{1 - \frac{S_{yx}^2}{S_y^2}} \tag{7.23}$$

або

$$S_Y^2 = S_R^2 + S_{YX}^2, \tag{7.13}$$

$$R^2 = \sqrt{\frac{ESS}{TSS}} = \sqrt{1 - \frac{RSS}{TSS}}. \tag{7.24}$$

У випадку парної лінійної регресійної моделі коефіцієнт детермінації дорівнює квадрату коефіцієнта кореляції, тобто r^2.

Розглянуті вище властивості *теоретичних* значень коефіцієнта детермінації та кореляційного відношення стосуються і їх *вибіркових (емпіричних)* значень.

Розглянемо принципову *модель дисперсійного аналізу.* Для цього повернемося до рівнянь (7.19), (7.20). Кожна сума квадратів пов'язана з *числом ступенів свободи*, при використанні яких отримаємо відповідні *незміщені оцінки дисперсій.*

Нагадаємо, що у математичній статистиці для одержання незміщеної оцінки дисперсії випадкової величини відповідну суму квадратів відхилень від середнього ділять не на кількість спостережень *n,* а на *кількість ступенів свободи,* яка дорівнює різниці між кількістю незалежних спостережень випадкової величини *n* і кількістю зв'язків, які обмежують свободу їх зміни, тобто кількістю рівнянь, що зв'язують ці спостереження, або кількістю констант, установлених в результаті цих спостережень.

Незміщена оцінка загальної дисперсії сукупності має вигляд [див. формулу (5.5)]:

$$S_y^{2*} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{TSS}{n-1}. \tag{7.25}$$

Отже, загальна сума квадратів відхилень залежної змінної має *(n*-1*)* ступінь свободи.

Сума квадратів, яка пояснює парну лінійну регресію, утворюється за допомогою лише однієї одиниці незалежної інформації – константи b_1 [див. рівняння (7.16)]. Тому пояснена сума квадратів має один ступінь свободи, а незміщена оцінка поясненої дисперсії визначається за формулою:

$$S_R^{2*} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{1} = \frac{ESS}{1} \qquad (7.26)$$

Кількість ступенів свободи в даному випадку збігається з кількістю незалежних змінних.

Кількість ступенів свободи суми квадратів помилок дорівнює різниці між кількістю спостережень і кількістю параметрів, що оцінюються. У даному випадку оцінюються два параметри b_0 і b_1 [див. рівняння (7.16)]. Тому сума квадратів помилок має *(n-2)* ступені свободи, а незміщена оцінка залишкової (непоясненої) дисперсії визначається за формулою:

$$S_{yx}^{2*} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \frac{RSS}{n-2}. \qquad (7.27)$$

Зауважимо, що в економетричному аналізі сума квадратів, поділена на відповідний ступінь свободи, також називається *середнім квадратом.*

С*ередній квадрат помилок* позначається *MSE* і має вигляд:

$$\mathrm{MSE} = S_{yx}^{2*} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \frac{RSS}{n-2}. \qquad (7.28)$$

Середній квадрат, який пояснює регресію, позначається *MSR*, збігається з поясненою сумою квадратів і має вигляд:

$$MSR = ESS = S_R^{2*} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{1}. \qquad (7.29)$$

Для загальної суми квадратів *TSS* середній квадрат не обчислюється.

Принципова модель дисперсійного аналізу для випадку однієї незалежної змінної (простої лінійної регресії) зображена у таблиці 3.

Непояснена (залишкова) дисперсія не має реальних одиниць виміру, а отже, і змістової інтерпретації. Тому в математичній статистиці та економетрії використовується також величина S_{yx}^{*}, яка називається *стандартною похибкою оцінки (стандартною похибкою залишків)*

$$S_{yx}^{2*} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2}}. \quad (7.30)$$

Таблиця 3

Модель дисперсійного аналізу для простої лінійної регресії

Компоненти дисперсії	Сума квадратів	Кількість ступенів свободи	Середні квадрати
Регресія	$ESS = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	1	$MSR == \frac{ESS}{1} = S_R^{2*}$
Залишкова	$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	*N*–2	$\mathrm{MSE} = \frac{RSS}{n-2} = S_{yx}^{2*}.$
Загальна	$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$	*N*–1	–

Стандартна похибка оцінки характеризує *варіацію фактичних даних* y_i *навколо теоретичних* $\hat{y}_i$ і є *мірою непоясненої варіації у загальній дисперсії.*

7.5. Багатовимірний кореляційний аналіз

У випадках, коли проводяться багатофакторні експерименти (досліджуються багатофакторні залежності), знаходження лише коефіцієнтів парної кореляції недостатньо.

Більш загальною, інформативною характеристикою є *коефіцієнт множинної кореляції.*

Якщо є сукупність випадкових величин $\overline{X_1, X_m}$, то зв'язок між кожною парою з них визначається за допомогою коефіцієнта парної кореляції. Повний набір коефіцієнтів складає *матрицю коефіцієнтів парної кореляції:*

$$q_m = \begin{vmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{vmatrix}. \qquad (7.31)$$

На підставі матриці (7.31) визначається *вибірковий коефіцієнт множинної кореляції:*

$$R_{i=\overline{1,m}} = \sqrt{1-\frac{|q_m|}{q_{jj}}}, \qquad (7.32)$$

де $|q_m|$ – визначник матриці (7.31); q_{jj} – алгебраїчне доповнення до елемента цієї матриці; *m* – загальне число факторів; *j* – номер фактора, зв'язок якого з усіма іншими аналізується.

Величина q_{jj} визначається як визначник матриці q_m, з якої викреслені рядок і стовпець з номерами j, тобто як мінор елемента r_{jj}.

Величина $R^2_{i=\overline{1,m}}$ називається *коефіцієнтом множинної детермінації (вибірковим).* Обидві величини ($R_{i=\overline{1,m}}$ і $R^2_{i=\overline{1,m}}$) змінюються в межах від 0 до 1. При цьому величина $R_{i=\overline{1,m}}$ говорить про щільність зв'язку фактора *j* з усіма решта факторами сукупності $X_1, \dots, X_m$. Величина $R^2_{i=\overline{1,m}}$ свідчить про те, яка частка зміни (варіації) фактора *j* визначається всіма решта факторами сукупності.

Примітка.

Визначником другого порядку $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ є вираз $\Delta = a_1 b_2 - a_2 b_1$.

Визначником третього порядку $\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$ є вираз $a_1 b_2 c_3 + a_3 b_1 c_2 + a_2 b_3 c_1 - a_3 b_2 c_1 - a_2 b_1 c_3 - a_1 b_3 c_2$,

або $a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}$.

Мінором будь-якого елемента є визначник, який одержується викреслюванням рядка і стовпця, на перетині яких стоїть даний елемент.

Визначники $\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}$, $\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix}$, $\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}$ є мінорами елементів a_1, b_1, c_1.

Алгебраїчним доповненням елемента є його мінор, взятий зі знаком, який визначається за правилом: якщо сума номерів рядка і стовпця, на перетині яких стоїть елемент, парне число, то мінор береться зі своїм знаком, якщо непарне, – то з протилежним.

Запитання та завдання для самоконтролю

1. Дайте визначення та наведіть приклади функціональної залежності між змінними величинами.
2. Які залежності між змінними величинами переважно існують в економіці?
3. Дайте визначення статистичної залежності між двома випадковими величинами.
4. Наведіть приклади статистичного зв’язку.
5. Дайте визначення кореляційної залежності між двома змінними величинами.
6. Як визначається теоретичний та вибірковий коефіцієнт кореляції?
7. Які властивості має коефіцієнт кореляції?
8. Чому необхідно оцінювати (перевіряти) значущість коефіцієнта кореляції?
9. Сформулюйте процедуру перевірки значущості коефіцієнта кореляції.
10. Які показники використовуються для визначення сили зв’язку між випадковими величинами у загальному випадку?
11. Назвіть складові повної дисперсії випадкової величини.
12. Дайте визначення коефіцієнта детермінації.
13. Як визначається теоретичний коефіцієнт детермінації?
14. Охарактеризуйте властивості коефіцієнта детермінації.
15. Як визначається теоретичне кореляційне відношення? В яких межах може знаходитись кореляційне відношення? Чому?
16. Які властивості має кореляційне відношення?
17. Проведіть декомпозицію дисперсій, використовуючи вибіркові оцінки.
18. Запишіть рівняння для суми квадратів відхилень залежної змінної від її вибіркового середнього значення.
19. Як визначається вибірковий коефіцієнт детермінації?

20. Як визначається вибіркове кореляційне відношення?

21. Як отримати незміщені оцінки дисперсії?

22. Поясніть, у чому полягає зміст ступенів свободи.

23. Чому дорівнює кількість ступенів свободи відповідних сум квадратів оцінок дисперсій?

24. Що розуміють під середніми квадратами в економетричному аналізі?

25. Запишіть модель регресійного аналізу для простої лінійної регресії.

26. Як визначається і що характеризує стандартна похибка оцінки?

27. Як визначається вибірковий коефіцієнт множинної кореляції?

28. Результати аналізу 10 зразків сировини, що надійшли з лабораторії, наведені у таблиці, де $\boldsymbol{x}$ та $\boldsymbol{y}$ – відповідний відсотковий вміст потрібних нам речовин у зразках:

$\boldsymbol{x}$	67	54	72	64	39	22	58	43	46	34
$\boldsymbol{y}$	24	15	23	19	16	11	20	16	17	13

Обчисліть коефіцієнт кореляції між статистичними змінними x та y і запишіть рівняння прямої регресії y відносно x.

29. У таблиці наведено результати вимірювання маси x (кг) і зросту y (см) 50 студентів:

$\boldsymbol{y}$	$\boldsymbol{x}$					
	[22,5;25,5]	[25,5;28,5]	[28,5;31,5]	[31,5;34,5]	[34,5;37,5]	Σ
[117,5;122,5]	1	3	–	–	–	4
[122,5;127,5]	–	2	6	1	–	9
[127,5;132,5]	–	1	5	5	–	11
[132,5;137,5]	–	1	6	7	2	16
[137,5;142,5]	–	–	1	4	2	7
[142,5;147,5]	–	–	–	1	1	2

[147,5;152,5]	–	–	–	–	1	1
Σ	1	7	18	18	6	50

Знайдіть вибірковий коефіцієнт кореляції між змінними x та y.

Запишіть рівняння вибіркових прямих регресії y на x, x на y та побудуйте ці прямі.

30. Існує певний взаємозв'язок між успішністю учнів з хімії та фізики. Нехай успішність 60 учнів 10-х класів з хімії за попередній навчальний рік описується змінною x, а з фізики – змінною y. Статистичний матеріал, отриманий в кінці навчального року, подано у кореляційній таблиці:

		Фізика (y)				Сума
		«2»	«3»	«4»	«5»	
Хімія (x)	«2»	0	0	0	0	0
	«3»	1	16	12	0	29
	«4»	0	6	3	4	13
	«5»	0	0	10	8	18
	Сума	1	22	25	12	60

Знайдіть кореляцію між змінними x та y та побудуйте прямі регресії.

31. Залежність між зростом Y та масою дітей X наведена в таблиці:

$Y = y_i$, м	0,620	0,580	0,640	0,650	0,670	0,680	0,695	0,699	0,710
Маса $X = x_i$, кг	0,531	0,524	0,541	0,550	0,559	0,620	0,632	0,672	0,682
$Y = y_i$, м	0,715	0,725	0,781	0,790	0,795	0,800	0,810	0,850	0,860
Маса $X= x_i$, кг	0,689	0,692	0,694	0,698	0,690	0,710	0,720	0,725	0,730

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0,99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0{,}01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

32. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведено в таблиці:

$Y = y_i$, шт.	10	25	68	136	152	162	170	180
$X = x_i$	44	43	42	41	40	39	38	37

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0{,}01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

33. Зі старшого класу навмання вибраної середньої школи було відібрано групу учнів. Дані про їх середньорічні оцінки з математики та решти дисциплін в балах наведено в таблиці:

$Y = y_i$	45	25	48	52	54	51	59	60	62	69
$X = x_i$	30	35	31	38	41	48	50	55	51	58
$Y = y_i$	72	78	76	80	82	85	81	90	93	
$X = x_i$	60	59	65	73	78	71	79	80	81	

1) Знайти точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0{,}01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

34. Конденсатор було заряджено до повної напруги в певний момент часу *t*, після цього він починає розряджатися. Залежність напруги *Y* від часу розряджання *X* наведено в таблиці:

$Y = y_i$	100	85	70	65	60	55	50
$X = x_i$	0	1	2	3	4	5	6
$Y = y_i$	45	40	35	30	25	22	20
$X = x_i$	7	8	9	10	11	12	13

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0{,}01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

35. Залежність урожайності пшениці *Y* від глибини зволоження *X* наведено в таблиці:

$Y = y_i$, ц/га	10	12	14	16	18	20	22	24	26	28
$X = x_i$, см	0	5	8	10	12	14	16	18	20	22
$Y = y_i$, ц/га	30	32	34	36	38	40	42	44	46	48
$X = x_i$, см	24	26	28	30	32	34	36	38	40	42

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0,99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0,01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0,99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0,99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

36. Показники товарообігу *Y* та суми витрат *X*, які досліджувалися в 20-ти магазинах, наведено в таблиці:

$Y = y_i$, грн.	480	510	530	540	555	564	570	575	580	585
$X = x_i$, грн.	30	25	31	32	38	41	40	46	49	54
$Y = y_i$, грн.	590	596	605	618	625	635	640	650	660	
$X = x_i$, грн.	58	60	64	75	78	82	83	85	90	

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0,99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0,01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0,99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0,99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

37. Результати вимірювання чутливості *Y* відеоканалу та звукового каналу *X* наведено в таблиці:

$Y = y_i$	240	200	190	180	170	160	150	140	130	120

$X = x_i$	170	180	200	230	240	250	280	300	310	320
$Y = y_i$	110	100	90	80	70	65	60	55	50	45
$X = x_i$	330	350	380	400	410	420	430	440	450	460

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0{,}01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

38. Залежність величини зносу різця Y від тривалості роботи X показано в таблиці:

$Y = y_i$, мм	30,0	29,1	28,4	28,1	28,0	27,7	27,5	27,2	27,0
$X = x_i$, год	6	7	8	9	10	11	12	13	14
$Y = y_i$, мм	26,8	26,5	26,3	26,1	25,7	25,3	24,3	24,1	24,0
$X = x_i$, год	15	16	17	18	19	20	21	22	23

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0{,}01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

39. Залежність кров'яного тиску Y людини (в умовних одиницях) від довжини руки X наведена в таблиці:

$Y = y_i$, умов. од.	115	116	117	118	119	120	121	122	123
$X = x_i$, см	62,1	61,1	61,0	60,5	60,0	59,0	58,5	58,0	57,5
$Y = y_i$, умов. од.	124	125	126	127	128	129	130	135	150
$X = x_i$, см	56,5	56,0	55,5	55,0	54,5	54,0	53,5	53,0	52,5

1) Знайдіть точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2) З надійністю $\gamma = 0{,}99$ побудуйте довірчі інтервали для параметрів β_0, β_1.

3) При рівні значущості $\alpha = 0{,}01$ перевірте значущість параметра β_1.

4) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5) Обчисліть вибірковий коефіцієнт кореляції.

6) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для прогнозованих індивідуальних значень.

40. Залежність між продуктивністю праці Y та фондозабезпеченістю, тис. грн., X_1, стажем роботи в роках X_2, X_3 – плинністю кадрів (у частках), рівнем оплати праці, тис. грн/рік – X_4 наведено в таблиці.

№ з/п	Y	X_1	X_2	X_3	X_4
1	14,85	60	30	0,15	5,0
2	11,94	48	19	0,02	3,1
3	8,03	39	8	0,14	4,7
4	7,11	28	18	0,11	2,5
5	9,50	45	9	0,12	4,9
6	9,40	37	23	0,10	2,6
7	11,60	58	15	0,13	4,6

8	8,14	27	17	0,09	3,4
9	15,62	58	28	0,07	4,8
10	11,12	47	16	0,12	4,9
11	7,34	38	7	0,08	3,2
12	10,58	44	15	0,11	4,7
13	7,37	23	25	0,15	2,7
14	10,63	57	8	0,13	5,0
15	10,63	38	24	0,07	2,9

1) Визначте статистичні оцінки $\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ для параметрів лінійної множинної регресії: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$.

2) З надійністю γ = 0,99 побудуйте довірчий інтервал для функції регресії.

3) Обчисліть коефіцієнт множинної кореляції R.

41. Залежність між продуктивністю праці Y та фондозабезпеченістю, тис. грн., X_1, стажем роботи в роках X_2, X_3 – плинністю кадрів (у частках), рівнем оплати праці, тис. грн/рік – X_4 наведено в таблиці.

№ з/п	Y	X_1	X_2	X_3	X_4
1	11,12	47	16	0,12	4,9
2	7,34	38	7	0,08	3,2
3	10,58	44	15	0,11	4,7
4	7,37	23	25	0,15	2,7
5	10,63	57	8	0,13	5,0
6	10,63	38	24	0,07	2,9
7	7,85	22	15	0,12	4,6
8	5,73	29	7	0,09	2,8
9	14,84	56	27	0,02	3,5
10	10,30	45	15	0,14	4,9

11	7,85	34	9	0,10	4,1
12	9,68	51	14	0,11	3,3
13	9,49	55	5	0,13	4,8
14	12,53	43	26	0,08	4,0
15	10,29	44	27	0,15	2,9

1) Визначте статистичні оцінки $\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ для параметрів лінійної множинної регресії:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}.$$

2) З надійністю $\gamma = 0,99$ побудуйте довірчий інтервал для функції регресії.

3) Обчисліть коефіцієнт множинної кореляції R.

42. Залежність між продуктивністю праці Y та фондозабезпеченістю, тис. грн., X_1, стажем роботи в роках X_2, X_3 – плинністю кадрів (у частках), рівнем оплати праці, тис. грн/рік – X_4 наведено в таблиці.

№ з/п	Y	X_1	X_2	X_3	X_4
1	7,85	22	15	0,12	4,6
2	5,73	29	7	0,09	2,8
3	14,84	56	27	0,02	3,5
4	10,30	45	15	0,14	4,9
5	7,85	34	9	0,10	4,1
6	9,68	51	14	0,11	3,3
7	9,49	55	5	0,13	4,8
8	12,53	43	26	0,08	4,0
9	10,29	44	27	0,15	2,9
10	8,99	37	8	0,06	4,3
11	12,28	33	24	0,12	5,0
12	8,00	25	18	0,02	2,9

13	7,27	29	4	0,07	3,5
14	7,47	53	13	0,14	2,7
15	10,86	41	9	0,08	4,9
16	5,23	26	12	0,13	3,4

1) Визначте статистичні оцінки $\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ для параметрів лінійної множинної регресії:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}.$$

2) З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії.

3) Обчисліть коефіцієнт множинної кореляції R.

43. Визначте статистичні оцінки $\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ для параметрів лінійної множинної регресії:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}.$$

З надійністю $\gamma = 0{,}99$ побудуйте довірчий інтервал для функції регресії. Обчисліть коефіцієнт множинної кореляції R. За заданими статистичними даними залежності врожайності цукрових буряків Y, ц/га, від:

1) живої праці X_1, людино-днів/г;
2) кількості внесених поживних речовин X_2, кг/га;
3) опадів на рік X_3, мм;
4) суми температур за період активної вегетації X_4, °С:

№ з/п	Y	X_1	X_2	X_3	X_4
1	369	16	83	460	2500
2	457	18	240	503	2621
3	379	13	125	496	2564
4	403	21	86	548	2792
5	439	17	221	472	2672
6	421	12	201	484	2840
7	448	23	217	537	2711

8	407	24	97	461	2638
9	419	18	144	493	2578
10	441	19	205	539	2617
11	418	20	156	526	2835
12	401	15	175	467	2693
13	451	17	189	542	2691
14	381	21	86	472	2532
15	432	18	204	483	2783

Розділ III

КЛАСИЧНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ. ОСНОВИ РЕГРЕСІЙНОГО АНАЛІЗУ

Тема 8. ПАРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

1. *Поняття функції регресії.*
2. *Узагальнена модель лінійної регресії.*
3. *Оцінка параметрів регресії методом найменших квадратів.*
4. *Передумови застосування регресійного аналізу і теореми Гаусса-Маркова.*
5. *Стандартні похибки оцінок коефіцієнтів регресії.*
6. *Статистичні критерії та порядок перевірки значущості оцінок параметрів і моделі в цілому.*
7. *Довірчі інтервали коефіцієнтів регресії.*
8. *Довірчі інтервали функції регресії.*

8.1. Модель парної лінійної регресії

Коефіцієнт кореляції показує, що дві змінні пов'язані між собою, проте не дає уяви про те, яким чином вони пов'язані. При побудові статистичних моделей важливо одержати *аналітичний вираз* для залежностей між випадковими величинами, якщо вони досить сильно корельовані. Такі залежності для випадкових величин одержують методами *регресійного аналізу.*

Нехай є вибірки з двох сукупностей X та Y, і коефіцієнт кореляції свідчить про достатньо сильний зв'язок між ними. Необхідно знайти аналітичний вираз виду

$$M_x(Y)=f(X), \qquad (8.1)$$

який описує зв'язок між цими величинами. Рівняння виду (8.1) називається *рівнянням регресії Y* по *X*, а лінія, яка описується даним рівнянням, *– лінією регресії*. Отже, рівняння регресії описує не точну залежність між випадковими величинами, а статистичну, тобто тенденцію зміни в середньому величини *Y* при зміні величини *X*.

Неможливо очікувати точного співвідношення між будь-якими економічними показниками, за винятком тих випадків, коли воно існує за означенням. У результаті впливу неврахованих випадкових факторів окремі спостереження змінної *Y* будуть відхилятися від функції регресії *f(X)*. У цьому випадку *парна регресійна модель* може бути представлена у вигляді:

$$Y = f(X) + \varepsilon, \qquad (8.2)$$

де ε – *випадкова змінна* (випадковий член), яка характеризує відхилення від функції регресії.

Цю змінну називають також збуренням, помилкою, залишком, шумом [2, с. 53–55; 5, с. 60]. Отже, в *регресійній моделі залежна змінна Y є деякою функцією f(X) з точністю до випадкового збурення* ε.

Розглянемо найпростішу модель *парного регресійного аналізу,* де функція *f(X)* є *лінійною* відносно параметрів, які оцінюються:

$$f(X) = M_x(Y)=\beta_0 + \beta_1 X, \qquad (8.3)$$

де β_0, β_1 – коефіцієнти регресії, тобто правильні (істинні) параметри для генеральних сукупностей *X* і *Y*. Тоді загальна лінійна парна регресійна модель має вигляд:

$$Y= \beta_0 + \beta_1 X+ \varepsilon, \qquad (8.4)$$

де ε – неспостережувана випадкова складова.

Існування випадкової складової може бути зумовлене: невключенням у модель істотних пояснювальних змінних; агрегуванням змінних, тобто спроба об'єднати декілька мікроекономічних співвідношень; неправильною функціональною специфікацією, тобто співвідношення між *y* та *x* математично може бути визначене неправильно; похибками вимірювань; неправильним описом структури моделі.

Залишкова складова є сумарним проявом усіх зазначених факторів. Оскільки генеральні сукупності *X* і *Y* та їх параметри, як правило, невідомі, на практиці визначаються не «істинні» коефіцієнти регресії β_j, а їх вибіркові оцінки b_j (вони також називаються коефіцієнтами регресії) та оцінки конкретних значень помилок *e*.

Нехай для оцінки параметрів лінійної функції регресії (8.4) зроблена вибірка з *n* спостережень (x_j, y_j), $j = 1,2,\dots,n$. Тоді у загальному вигляді парна *вибіркова* регресійна модель запишеться так:

$$y_i = b_0 + b_1 x_i + e_i, \tag{8.5}$$

де y_i – вектор спостережень за залежною змінною,

$y_i = \{y_1, y_2, \dots y_n\}$; x_i – вектор спостережень за незалежною змінною, $x_i = \{x_1, x_2, \dots, x_n\}$; b_0, b_1 – вибіркові оцінки коефіцієнтів регресії; e_i – вектор помилок (випадкових складових), $e_i = \{e_1, e_2, \dots, e_n\}$.

Оскільки невипадкова складова або теоретичне рівняння регресії *y* від *x* має вигляд (див. тему 7):

$$\hat{y}_i = b_0 + b_1 x_i, \tag{8.6}$$

то рівняння (8.5) перепишеться так:

$$y_i = \hat{y}_i + e_i, \tag{8.7}$$

На рис. 21 показано, що комбінація двох складових (невипадкової і випадкової) визначають величину y_i.

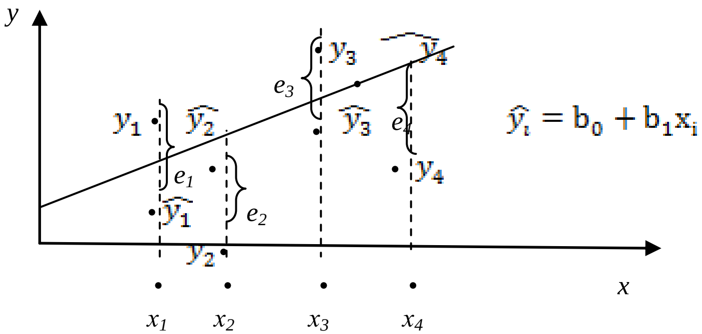


Рис. 21. Відхилення теоретичних значень залежної змінної від фактичних

Показники x_1, x_2, x_3, x_4 – це фактичні значення незалежної змінної, y_1, y_2, y_3, y_4 – це точки на прямій теоретичного рівняння регресії (8.6).

Наявність випадкового члена $e_i = \{e_1, e_2,..., e_n\}$ приводить до того, що реальні значення залежної змінної відповідають точкам y_1, y_2, y_3, y_4. Припускалося, що випадковий член e_i є додатним у першому і третьому спостереженнях; від'ємний – у другому і четвертому.

8.2. Метод найменших квадратів

Для того, щоби записати рівняння регресії і побудувати лінію регресії, необхідно визначити *оцінки* b_j *коефіцієнтів регресії* β_j. Найчастіше для цього використовується *метод найменших квадратів* (МНК), оскільки він забезпечує незміщені та ефективні оцінки β_j.

За МНК невідомі параметри b_0 і b_1 обираються так, щоб сума квадратів відхилень емпіричних значень y_i від значень $\hat{y}_i$, знайдених за теоретичним рівнянням регресії (8.6), була мінімальною, тобто

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \to \min. \quad (8.8)$$

Підставимо рівняння (8.6) у формулу (8.8)

$$S = \sum_{i=1}^{n} \left(y_i - b_0 + b_1 x_i\right)^2 \to \min. \quad (8.9)$$

Для досягнення екстремуму функції двох змінних $S = f(b_0, b_1)$, необхідно прирівняти до нуля її перші частинні похідні:

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum\limits_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0; \\ \dfrac{\partial S}{\partial b_1} = -2\sum\limits_{i=1}^{n}(y_i - b_0 - b_1 x_1) x_i = 0. \end{cases} \quad (8.10)$$

Звідси, врахувавши, що $\sum_{i=1}^{n} b_0 = nb_0$, одержимо *систему нормальних рівнянь:*

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases} \quad (8.11)$$

Розв'язок цієї системи дає значення оцінок коефіцієнтів регресії b_0 та b_1:

$$b_0 = \bar{y} - b_1 \bar{x}, \quad (8.12)$$

$$b_1 \frac{cov(x,y)}{S_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \quad (8.13)$$

Зауважимо, що існують також інші методи визначення оцінок коефіцієнтів регресії: метод найменших модулів, метод максимальної правдоподібності [5, с. 53, 63–64]. Проте метод найменших квадратів є значно простішим і дає, за певних умов, найкращі оцінки. Цим і пояснюється його широке застосування у статистиці та економетрії.

8.3. Передумови застосування регресійного аналізу

Після визначення оцінок коефіцієнтів регресії b_0 і b_1 постає питання, наскільки визначені оцінки відповідають дійсним значенням коефіцієнтів регресії β_0 і β_1, а також наскільки теоретичне значення *y* відповідає дійсному значенню математичного сподівання $M_X(Y)$. Для відповіді на ці запитання необхідно зробити певні припущення.

Ці припущення, відомі як *умови Гаусса-Маркова,* є передумовами класичного регресійного аналізу. Розглянемо ці умови для простої лінійної регресії [2, с. 80–82; 5, с. 69–76].

Перша умова Гаусса-Маркова. Математичне сподівання випадкової складової ε_j дорівнює нулю для будь-якого спостереження, тобто

$$M(\varepsilon_j) = 0. \quad (8.14)$$

Для деяких спостережень випадкова складова буде додатною, для інших – від'ємною, але вона не повинна мати систематичного зсуву в жодному з двох можливих напрямків.

Припущення (8.14) означає, що математичне сподівання залежної змінної y_i дорівнює лінійній функції регресії, тобто $M_x(y_i) = \beta_0 + \beta_1, x_i$.

Друга умова Гаусса-Маркова. Дисперсія випадкової складової ε_j є сталою (однаковою) для будь-якого спостереження x_j

$$D(\varepsilon_j) = \sigma_\varepsilon^2, \tag{8.15}$$

(σ_ε^2 – дисперсія випадкової складової).

Математично ця умова записується [див. формулу (3.19)]:

$$D(\varepsilon_j) = M[\varepsilon_j - M(\varepsilon_j)]^2.$$

Оскільки $M(\varepsilon_j) = 0$ і $D(\varepsilon_j) = M(\varepsilon_j^2)$, то

$$M(\varepsilon_j^2) = \sigma_\varepsilon^2. \tag{8.16}$$

Величина σ_ε невідома. Одна із задач регресійного аналізу полягає в оцінці стандартного відхилення випадкового члена.

Умова сталості дисперсії випадкової величини ε_i називається *гомоскедастичністю.*

Третя умова Гаусса-Маркова. Некорельованість випадкових величин ε_i та ε_j (або змінних y_i та y_j). Умова некорельованості означає, що випадкові складові повинні бути незалежними між собою, тобто коефіцієнт коваріації між випадковими складовими ε_i та ε_j повинен дорівнювати нулю

$$pop.cov(\varepsilon_i, \varepsilon_j) = 0 \tag{8.17}$$

або

$$pop.cov(\varepsilon_i, \varepsilon_j) = M([\varepsilon_i - M(\varepsilon_i)]\,[\varepsilon_j - M(\varepsilon_j)]) = M[(\varepsilon_i - 0)(\varepsilon_j - 0)] = M(\varepsilon_i \varepsilon_j). \tag{8.18}$$

Четверта умова Гаусса-Маркова. Випадкова складова ε_i розподілена незалежно від пояснювальної змінної x_i. Виконання цієї умови означає, що теоретична коваріація між незалежною змінною і випадковою складовою дорівнює нулю.

$$pop.cov(x_i, \varepsilon_i) =$$
$$= M\left([x_i - M(x_i)]\,[\varepsilon_i - M(\varepsilon_j)]\right) = M(x_i, \varepsilon_i) - M(x_i)M(\varepsilon_i) =$$
$$= M(x_i, \varepsilon_i) = 0 \text{ (оскільки} M(\varepsilon_i) = 0\text{).}$$

Отже, дану умову можна записати у вигляді:

$$M(x_i, \varepsilon_i) = 0. \qquad (8.19)$$

Крім умов Гаусса-Маркова, передумовою регресійного аналізу є припущення про *нормальність розподілу випадкової складової,* яка базується на *центральній граничній теоремі*. Випадкова складова ε визначається декількома факторами, які не входять явно до рівняння регресії. Тому навіть якщо ми нічого не знаємо про розподіл цих факторів, то маємо право зробити припущення, що вони нормально розподілені.

Умова нормальності розподілу випадкової складової є необхідною при перевірці гіпотез і побудові довірчих інтервалів для β_0 та β_1, залежної змінної *Y* та інших параметрів регресійної моделі. Справа в тому, що коли випадкова складова *е* розподілена нормально, то так само будуть розподілені і коефіцієнти регресії та залежна змінна.

Якщо виконуються всі умови Гаусса-Маркова та припущення про нормальність розподілу випадкової складової, то модель (8.4) називається *класичною нормальною лінійною регресійною моделлю* [5, с. 61].

Для одержання рівняння регресії достатньо передумов 1–4. Вимога нормального розподілу необхідна для оцінки *точності* рівняння регресії та його параметрів.

Як зазначалося вище, оцінкою моделі (8.4) за вибіркою є рівняння регресії (8.6). Параметри цього рівняння b_0 та b_1 визначаються на основі методу найменших квадратів. Вплив неврахованих випадкових факторів і помилок спостережень у моделі (8.4) визначається за допомогою *дисперсії помилок або залишкової дисперсії* σ_ε^2. Незміщеною оцінкою цієї дисперсії є вибіркова залишкова дисперсія (див. розділ 7)

$$S_{yx}^{2*} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}\ \ )^2}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}. \qquad (8.20)$$

Виникає питання, чи є оцінки b_0, b_1, S_{yx}^{2*} параметрів β_0, β_1, σ_ε^2 найкращими. Відповідь на нього дає теорема Гаусса-Маркова.

Теорема Гаусса-Маркова. Якщо регресійна модель (8.4) *задовольняє умовам Гаусса-Маркова 1–4, то оцінки* b_0 (8.12) і b_1 (8.13) *мають найменшу дисперсію в класі всіх лінійних незміщених оцінок.*

Теорема Гаусса-Маркова показує, що оцінки b_0 та b_1 одержані методом найменших квадратів, є *найкращими лінійними незміщеними оцінками.* Тому за означенням є *ефективними* лінійними оцінками параметрів β_0 та β_1.

8.4. Незміщеність і точність оцінок коефіцієнтів регресії

За означенням, оцінка *b* параметра *β* є незміщеною, якщо її математичне сподівання дорівнює параметру, що оцінюється, тобто

$M(b) = \beta$.

Доведено, що за виконання *1-ї* та *4-ї* умов Гаусса-Маркова [2, с. 82–83; 6, с. 81]

$$M(b_0)=\beta_0, \qquad (8.21)$$

а отже, b_0 є незміщеною оцінкою істинного коефіцієнта регресії β_0.

Також доведено, що коли виконується 4-та умова Гаусса-Маркова [2, с. 82; 6, с. 78–80], то

$$M(b_1)=\beta_1, \qquad (8.22)$$

тобто b_1 є незміщеною оцінкою істинного коефіцієнта регресії β_1.

Безумовно, для будь-якої конкретної вибірки фактор випадковості буде приводити до розходження оцінок та істинних значень коефіцієнтів регресії, але з урахуванням співвідношень (8.21) та (8.22) не буде систематичних похибок, які завищують або занижують оцінки b_0 та b_1 коефіцієнтів регресії β_0 та β_1.

Для визначення точності оцінок коефіцієнтів регресії розглянемо теоретичні дисперсії оцінок b_0 та b_1:

$$\mathrm{pop.var}\,(\mathrm{b}_0) = \sigma_\varepsilon^2 \frac{\sum_{i=0}^{n} x_i^2}{\sum_{i=1}^{n}(\mathrm{x_i} - \bar{\mathrm{x}})^2}; \tag{8.23}$$

$$\mathrm{pop.var}\,(\mathrm{b}_1) = \frac{\sigma_\varepsilon^2}{n \cdot s_x^2} = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(\mathrm{x_i} - \bar{\mathrm{x}})^2}. \tag{8.24}$$

На основі (8. 23) та (8. 24) можна зробити наступні висновки. По-перше, дисперсії b_0 та b_1 прямо пропорційні до дисперсії залишкового члена σ_ε^2. Це означає, що чим більшим є фактор випадковості, тим менш точними за інших рівних умов, будуть оцінки b_0 та b_1.

По-друге, чим більша кількість спостережень *n,* тим меншою є дисперсія оцінок. Це означає, що із збільшенням обсягів інформації підвищується, ймовірно, точність оцінок.

По-третє, чим більшою є дисперсія незалежної змінної *x*, тим меншою буде дисперсія оцінок коефіцієнтів регресії.

На практиці ми не можемо обчислити теоретичні дисперсії b_0 та b_1, тому що випадкову величину *e* не можемо спостерігати, а отже, значення дисперсії σ_ε^2 не можемо визначити.

Замість дисперсії σ_ε^2 використовується її оцінка S_{yx}^{2*} .

Доведено, що

$$M(S_{yx}^{2*}) = \sigma_\varepsilon^2. \tag{8.25}$$

Отже, S_{yx}^{2*} є незміщеною оцінкою σ_ε^2.

Використовуючи формули (8.23), (8.24), (8.25) та (8.20), можна визначити *оцінки теоретичних дисперсій для* b_0 і b_1 та *оцінки* *стандартних відхилень оцінок коефіцієнтів регресії або просто стандартні похибки оцінок коефіцієнтів регресії* S_{b_i} .

Для парного регресійного аналізу:

$$S_{b_0} = S^*_{yx}\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}}; \tag{8.26}$$

$$S_{b_1} = \frac{S^*_{yx}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}, \tag{8.27}$$

де S^{2*}_{yx} – стандартна похибка оцінки за рівнянням регресії.

Стандартна похибка оцінки коефіцієнта регресії S_{b_i} дає лише загальну оцінку точності оцінок коефіцієнтів регресії. Чим більшою є дисперсія випадкової складової, тим більшою буде вибіркова дисперсія залишків, а отже, більш істотними будуть стандартні похибки оцінок коефіцієнтів регресії у теоретичному рівнянні регресії.

Це дозволяє з високою ймовірністю зробити висновок про низьку точність отриманих оцінок коефіцієнтів регресії. Проте, такий висновок буде ймовірнісним.

8.5. Перевірка значущості оцінок коефіцієнтів регресії

Оцінки коефіцієнтів регресії визначаються на основі обмеженої вибірки, тобто b_0 та b_1 є точковими оцінками параметрів β_0 та β_1 відповідно. Тому необхідно перевірити значущість одержаних оцінок.

Така перевірка є особливо важливою, якщо рівняння регресії планується використати для аналізу та прогнозування.

Для перевірки значущості перевіряється нуль-гіпотеза

$$H_0 : \beta_j = 0 \tag{8.28}$$

проти альтернативної

$$H_j : \beta_j \neq 0. \qquad (8.29)$$

Як тестова використовується t-статистика з $k = n$-2 ступенями свободи

$$|t_j| = \frac{|b_j|}{s_{b_j}}, \qquad (8.30)$$

яка підпорядкована розподілу Стьюдента.

Перевірка гіпотез (8.28), (8.29) здійснюється наступним чином.

За формулою (8.30) визначаються розрахункові значення t-статистики t_j, j = 0,1. Потім обирається рівень значущості α (або $\alpha \times 100\%$) або довірчої ймовірності $P_a = 1 - \alpha$ (або $p_\alpha = (1 - \alpha)\ 100\%$). Далі визначається за таблицями Стьюдента відповідне табличне значення $t_{\alpha,\kappa=n-2}$, яке порівнюється з розрахунковим значенням t_j.

Якщо $|t_j| < t_{\alpha,\kappa}$, то нуль-гіпотеза приймається, тобто можна стверджувати, що з імовірністю P_a =1-α оцінка b_j є статистично незначущою.

Якщо $|t_j| > t_{\alpha,\kappa}$, то приймається альтернативна гіпотеза H_j: $\beta_j \neq 0$, а оцінка b_j з імовірністю P_a вважається статистично значущою і нею можна користуватися. Якщо значущою є оцінка b_j, то у випадку парної лінійної регресії це означає, що зміна x впливає на зміну y.

8.6. Довірчі інтервали для коефіцієнтів регресії b_0 і b_1

До того, щоб визначити, як оцінки коефіцієнтів регресії b_0 та b_1 пов'язані з істинними параметрами β_0 та β_1, необхідно провести їх інтервальне оцінювання, тобто побудувати довірчі інтервали для цих параметрів.

З метою побудови довірчих інтервалів для коефіцієнтів регресії β_0 та β_1 використовується *t*-статистика [2, с. 96; 5, с. 67; 6, с. 93]:

$$t_j = \frac{b_j - \beta_j}{S_{b_j}}, j = 0{,}1 \qquad (8.31)$$

з $k = n - 2$ ступенями свободи.

Далі обирається рівень значущості α (або α – 100%) і відповідний рівень довірчої ймовірності $P_\alpha = 1 - \alpha$. В економетрії для побудови довірчих інтервалів найчастіше використовується 95%-й рівень довірчої ймовірності.

За *t*-таблицями Стьюдента знаходять табличне значення $\pm\, t_{\alpha,k}$ з $k = n\text{-}2$ ступенями свободи. За означенням довірчого інтервалу можна записати, що

$$P(t_{\alpha,k} < t_j < t_{\alpha,k}) = P_\alpha = 1\text{-}\alpha = 0{,}95.$$

Підставимо замість t_j значення (8.31) і отримаємо:

$$P\left(t_{\alpha,k} < \frac{b_j - \beta_j}{S_{b_j}} < t_{\alpha,k}\right) = 0{,}95.$$

Якщо провести відповідні перетворення виразу в дужках та помножити всі його члени на *(-1)*, отримаємо формулу для інтервальної оцінки коефіцієнта регресії β_j на рівні значущості $\alpha = 0{,}05$ з $k = n\text{-}2$ ступенями свободи:

$$b_j - t_{\alpha,k} \cdot S_{b_j} < \beta_j < b_j + t_{\alpha,k} \cdot S_{b_j} \qquad (8.32)$$

або

$$\beta_j = b_j \pm t_{\alpha,k} \cdot S_{b_j}, j = 0{,}1. \qquad (8.33)$$

Введемо поняття *граничної похибки оцінки коефіцієнта регресії* ΔS_{b_j}:

$$\Delta S_{b_j} = \mathrm{t}_{\alpha,\mathrm{k}} \cdot S_{b_j}, j = 0,1. \quad (8.34)$$

Тоді формули (8.32) та (8.33) для визначення довірчого інтервалу, відповідно, матимуть вигляд:

$$b_j - \Delta S_{b_i} < \beta_j < b_j + \Delta S_{b_i}, \beta_j = b_j \pm \Delta S_{b_i}, j = 0,1.$$

8.7. Перевірка значущості рівняння регресії в цілому. Прогнозування

Перевірити значущість рівняння регресії – означає з'ясувати, чи відповідає математична модель, яка виражає залежність між змінними, експериментальним даним і чи достатньо включених у рівняння пояснювальних змінних (однієї чи декількох) для опису залежної змінної [5, с. 70].

Перевірка значущості рівняння регресії здійснюється на основі дисперсійного аналізу.

Одним з найефективніших показників адекватності регресійної моделі, мірою якості рівняння регресії, характеристикою прогностичної сили регресійної моделі є коефіцієнт детермінації R^2 [5, с. 74–75].

Величина R^2 показує, яка частка варіації залежної змінної обумовлена варіацією пояснювальної змінної. Чим ближче R^2 до одиниці, тим краще регресія апроксимує емпіричні дані, тим більш щільно дані спостережень близькі до лінії регресії. Якщо $R^2=1$, то емпіричні точки (x_i,y_i) лежать на лінії регресії і між змінними X та Y існує лінійна функціональна залежність (рис. 19). Отже, якщо значення коефіцієнта детермінації близьке до одиниці, то можна вважати рівняння регресії адекватним. Якщо ж значення R^2 близьке до нуля, то модель неадекватна, тобто лінійний зв'язок між залежною та незалежною змінними відсутній. Проте якщо значення R^2 не наближається ні до нуля, ні до одиниці, важко зробити однозначний висновок щодо адекватності регресійної моделі.

Крім того, вибірковий (емпіричний) коефіцієнт детермінації R^2 визначається на основі обмежених вибірок (x_i, y_i), $i = \overline{1, n}$ з випадкових величин X та Y. Тому ми можемо судити про адекватність «істинного» рівняння регресії $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, лише перевіривши значущість самого коефіцієнта детермінації, а отже, і значущість рівняння в цілому.

Для перевірки значущості коефіцієнта детермінації і рівняння регресії в цілому використовується F-статистика Фішера-Снедекора, яка визначається як відношення середнього квадрата, що пояснює регресію MSR (формула (7.29)), до середнього квадрата помилок MSE (формула (7.28)).

Для випадку парної регресійної моделі F- критерій дорівнює:

$$F = \frac{MSR}{MSE} = \frac{ESS/1}{RSS/(n-2)}. \qquad (8.35)$$

Якщо чисельник і знаменник співвідношення (8.35) поділити на загальну суму квадратів відхилень TSS ((7.19), (7.20)), то F-статистика може бути еквівалентно виражена на основі коефіцієнта детермінації R^2 (7.21), (7.22):

$$F = \frac{(ESS/TSS)/1}{(RSS/TSS)/(n-2)} = \frac{R^2/1}{(1-R^2)/(n-2)}. \qquad (8.36)$$

Перевірка значущості рівняння регресії (перевірка моделі на адекватність) за F-критерієм Фішера-Снедекора здійснюється у такій послідовності.

1. Перевіряється нуль-гіпотеза

$$H_0: \hat{y}_1 = \beta_0 + \beta_1 x_i = \bar{y} \ (\text{або } H_0: \beta_1 = 0)$$

відносно альтернативної

$$H_1: \hat{y}_i = \beta_0 + \beta_1 x_i = \bar{y}$$ [6, с. 64].

2. За формулами (8.35) або (8.36) обчислюється фактичне значення F-статистики.

3. Обирається рівень значущості α (або α : 100%). Зауважимо, що критичні (табличні) значення F-статистики Фішера-

Снедекора обчислені для рівнів значущості 10%, 5% та 1%. За статистичними таблицями F-розподілу визначається табличне значення F-критерію на рівні значущості α.

4. При $k_1 = 1$ та $k_2 = n - 2$ ступенях свободи знаходимо $F_{\alpha,k1,k2}$.

5. Фактичне значення F-статистики порівнюється з табличним критичним.

Якщо $F > F_{\alpha,k1,k2}$, то нуль-гіпотеза про те, що $\hat{y}_i = \bar{y}$ ($\beta_1 = 0$) відхиляється з ризиком помилитися не більше ніж у α% випадків. Рівняння регресії вважається значущим (адекватним) на рівні α.

Враховуючи зміст величин *MSR* та *MSE,* а також зміст нуль-гіпотези H_0 та альтернативної гіпотези H_1, можна стверджувати, що F-критерій показує, якою мірою регресія краще оцінює значення залежної змінної y порівняно з її середнім значенням.

Необхідно зауважити, що значущість рівняння парної лінійної регресії може бути перевірена також іншим способом. Для цього необхідно оцінити значущість коефіцієнта регресії b_1 (тобто перевірити нуль-гіпотезу $H_0 : \beta_1 = 0$), який має t-розподіл Стьюдента з $k = n$-2 ступенями свободи

$$|t_1| = \frac{|b_1|}{s_{b_1}}. \qquad (8.37)$$

Рівняння парної лінійної регресії або коефіцієнт регресії b_1 значущі на рівні α (тобто гіпотеза $H_0 : \beta_1 = 0$ відхиляється), якщо фактичне значення t-статистики (8.37) за абсолютною величиною більше від критичного: $|t| > t_{\alpha,k}$.

Крім того, доведено, що для випадку парної регресійної моделі (і тільки парної регресійної моделі) t-критерій для перевірки значущості коефіцієнта кореляції $H_0 : \rho_{x,y} = 0$ (7.7), F-критерій для перевірки значущості коефіцієнта детермінації R^2 (8.35), (8.36) і t-критерій для перевірки значущості коефіцієнта регресії b_1, $H_0 : H_0 : \beta_1 = 0$ (8.37) еквівалентні, тобто дають однаковий результат і пов'язані співвідношенням $F = t^2$.

Отже, коефіцієнт кореляції r між змінними x та y буде вказувати на значущу залежність тоді і тільки тоді, коли й рівень коефіцієнта детермінації R^2 в регресії між x та y буде свідчити про таку залежність.

Крім того, величина b_1 буде значущо відрізнятися від нуля при застосуванні t-критерію тоді і тільки тоді, коли F-критерій буде значущим на даному рівні α.

Усі три тести дають відповідь на питання про адекватність парної регресійної моделі.

Якщо регресійна модель адекватна, її можна використовувати для аналізу і прогнозування. Можна побудувати два типи прогнозів: точковий та інтервальний.

Точковий прогноз дозволяє визначити значення залежної змінної y на основі побудованої вибіркової моделі теоретичного рівняння регресії (8.6) для нового значення незалежної змінної x_{n+1} за формулою

$$\hat{y}_{n+1} = b_0 + b_1 x_{n+1} . \qquad (8.38)$$

«Істинне» значення y_{n+1} для спостереження x_{n+1}, виходячи із загальної лінійної парної регресійної моделі (8.4) – (8.5), матиме вигляд:

$$y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \varepsilon_{n+1}, \qquad (8.39)$$

де ε_{n+1} – неспостережуваний *(n+*1*)-й* випадковий член.

Прогноз $\hat{y}_{n+1}$ є точковою оцінкою «істинного» значення залежної змінної y_{n+1}.

Необхідно також побудувати довірчі інтервали для «істинного» значення залежної змінної, в які б вона потрапляла з заданою довірчою ймовірністю P_a. Визначають довірчі інтервали двох видів: для функції регресії, тобто для умовного математичного сподівання $M_x(Y)$, і для індивідуальних значень залежної змінної.

Довірчий інтервал для математичного сподівання у. Побудуємо довірчий інтервал для функції регресії, або для умовного математичного сподівання $M_{x_{n+1}}(Y)$, який із заданою довірчою ймовірністю $P_a = 1 - \alpha$ містить невідоме значення $M_{x_{n+1}}$ (Y) (8.3):

$$M(y_{n+1})=\beta_0 + \beta_1\, x_{n+1}. \qquad (8.40)$$

Визначимо дисперсію групового середнього значення $\hat{y}_{n+1}$, яка є вибірковою оцінкою $M(y_{n+1})$. Для цього у рівняння для визначення $\hat{y}_{n+1}$ (8.38) підставимо значення $b_0 = \bar{y} - b_1\bar{x}$ (8.12). У результаті одержимо:

$$\hat{y}_{n+1} = \bar{y} - b_1\bar{x} + b_1 x_{n+1} = \bar{y} + b_1(x_{n+1} - \bar{x}). \qquad (8.41)$$

Дисперсія групового середнього значення $\hat{y}_{n+1}$ дорівнює сумі дисперсій двох незалежних доданків виразу (8.41)

$$\sigma^2_{\hat{y}_{n+1}} = \sigma^2_{\bar{y}} + \sigma^2_{b_1}(x_{n+1} - \bar{x})^2. \qquad (8.42)$$

Дисперсія вибіркового середнього значення $\bar{y}$

$$\sigma^2_{\bar{y}} = \frac{\sigma^2_{\varepsilon}}{n}. \qquad (8.43)$$

Дисперсія $\sigma^2_{b_1}$ визначається за формулою (8.24)

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

Якщо підставити вираз (8.43) і $\sigma^2_{b_1}$ у формулу (8.42) та замінити σ^2_{ε} її оцінкою S^{2*}_{yx} (8.20), знайдемо оцінку дисперсії групових середніх значень:

$$S^2_{\hat{y}_{n+1}} = \frac{S^{2*}_{yx}}{n} + \frac{S^{2*}_{yx}(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = S^{2*}_{yx}\left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]. \qquad (8.44)$$

Для побудови довірчого інтервалу використовується статистика

$$t = \frac{\hat{y}_{n+1} - M_x(Y)}{S_{\hat{y}_{n+1}}}, \qquad (8.45)$$

яка має *t*-розподіл Стьюдента з $k = n\text{-}2$ ступенями свободи, де $S_{\hat{y}_{n+1}} = \sqrt{S^2_{\hat{y}_{n+1}}}$ – стандартна похибка групового середнього значення $\hat{y}_{n+1}$.

За означенням довірчий інтервал можна записати у такому вигляді:

$$P(-t_{\alpha,k} < t < t_{\alpha,k}) = P_a.$$

Якщо підставити у попередній вираз значення *t*-статистики для x_{n+1}та $\hat{y}_{n+1}$, одержимо формулу для визначення довірчого інтервалу для математичного сподівання:

$$P(-t_{\alpha,k} < \frac{\hat{y}_{n+1} - M(y_{n+1})}{S_{\hat{y}_{n+1}}} < t_{\alpha,k}) = P_a,$$

$$\hat{y}_{n+1} - t_{\alpha,k} \cdot S_{\hat{y}_{n+1}} < M(y_{n+1}) < \hat{y}_{n+1} + t_{\alpha,k} \cdot S_{\hat{y}_{n+1}}, \qquad (8.46)$$

або

$$M(y_{n+1}) = \hat{y}_{n+1} \pm t_{\alpha,k} \cdot S^*_{yx} \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}. \qquad (8.47)$$

З формул (8.44) та (8.47) видно, що величина довірчого ін.-тервалу залежить від значення незалежної змінної. Оцінка дисперсії помилки прогнозу буде мінімальною, якщо $x_{n+1} = \bar{x}$, а беручи до уваги віддалення x_{n+1} від середнього значення *x,* вона нелінійно зростає. Тому величина довірчого інтервалу мінімальна при $x_{n+1} = \bar{x}$, а у міру віддалення *x* від $\bar{x}$ вона збільшується. Таким чином, прогноз значень залежної змінної *Y* за рівнянням регресії буде виправданим, якщо значення *x* незалежної змінної *X* не виходять за діапазон значень вибірки. При цьому точність прогнозу підвищується з наближенням x до $\bar{x}$.

Отже, екстраполяція кривої регресії (використання її за межами обстежуваного діапазону значень незалежної змінної) може призвести до значних похибок (рис. 22).

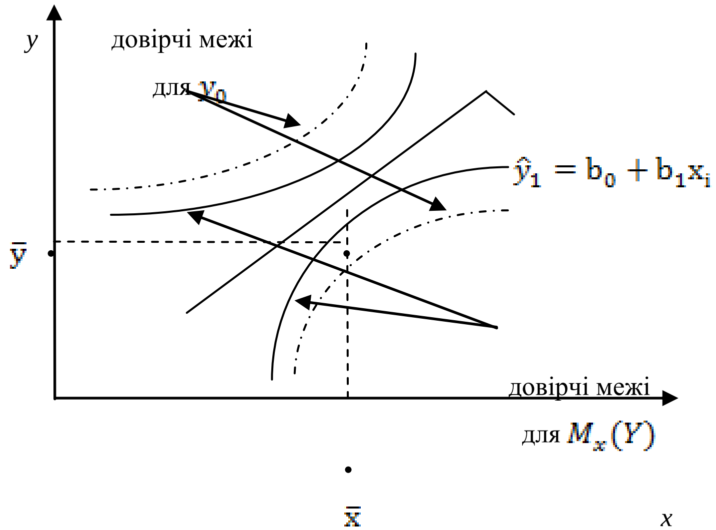


Рис. 22. Інтервальна оцінка функції регресії

Довірчий інтервал для індивідуальних значень залежної змінної. Побудована довірча область для $M_x(Y)$ визначає місцеположення умовного математичного сподівання, тобто модельної лінії регресії, але не окремих можливих значень залежної змінної, які відхиляються від середнього. Тому при визначенні довірчого інтервалу для індивідуальних значень y_0 залежної змінної необхідно враховувати ще одне джерело варіації – розсіяння навколо лінії регресії, тобто в оцінку сумарної дисперсії $S^2_{\hat{y}_0}$ необхідно включити додатково величину S^{2*}_{yx}. В результаті оцінка дисперсії індивідуальних значень y_0 при $x = x_0$ буде виражатися так:

$$S_{\hat{y}_0}^2 = S_{yx}^{2*} + S_{yx}^{2*}\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] = S_{yx}^{2*}\left[1 + \frac{1}{n} + \right.$$
$$\left. + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] \qquad (8.48)$$

.

Графічна інтерпретація інтервальної оцінки функції регресії зображена на рис. 22.

Довірчий інтервал для прогнозів індивідуальних значень y_0 визначається за формулами:

$$\hat{y}_0 - t_{\alpha,k} \cdot S_{\hat{y}_0} < y_0 < \hat{y}_0 + t_{\alpha,k} \cdot S_{\hat{y}_0} \qquad (8.49)$$

або

$$y_0 = \hat{y}_0 \pm t_{\alpha,k} \cdot S_{yx}^{*}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\,, \qquad (8.50)$$

де

$$\hat{y}_0 = b_0 + b_1 x_0;\; S_{\hat{y}_0} = \sqrt{S_{\hat{y}_0}^2}.$$

Запитання та завдання для самоконтролю

1. Запишіть аналітичний вираз рівняння регресії. Яку залежність між випадковими величинами описує рівняння регресії?
2. Запишіть загальну лінійну парну регресійну модель.
3. Які причини існування випадкової складової у парній регресійній моделі?
4. Запишіть вибіркову лінійну парну регресійну модель.
5. Дайте графічну інтерпретацію вибіркової лінійної парної регресійної моделі.
6. Сформулюйте критерій методу найменших квадратів.
7. Виведіть систему нормальних рівнянь за методом найменших квадратів.

8. Запишіть принципові формули для визначення оцінок коефіцієнтів регресії за методом найменших квадратів.

9. Сформулюйте першу та другу умову Гаусса-Маркова. Поясніть їх суть.

10. Сформулюйте третю та четверту умову Гаусса-Маркова та запишіть їх математичні вирази.

11. На чому базується припущення про нормальність розподілу випадкової складової?

12. Сформулюйте поняття класичної нормальної лінійної регресійної моделі.

13. Сформулюйте теорему Гаусса-Маркова. Поясніть її зміст.

14. Сформулюйте умови незміщеності оцінок b_0 та b_1 «істинних» коефіцієнтів регресії β_0 та β_1 відповідно.

15. Запишіть формули для визначення теоретичних дисперсій оцінок коефіцієнтів регресії.

16. Запишіть формули для визначення стандартних похибок оцінок коефіцієнтів регресії.

17. Сформулюйте порядок перевірки значущості оцінок коефіцієнтів регресії.

18. Як визначаються довірчі інтервали для коефіцієнтів регресії?

19. Сформулюйте поняття граничної похибки оцінки коефіцієнта регресії.

20. У чому полягає суть перевірки значущості рівняння регресії?

21. Запишіть формули F-критерію для перевірки значущості рівняння регресії та коефіцієнта детермінації.

22. Сформулюйте порядок перевірки значущості рівняння регресії за F-критерієм.

23. Яким чином значущість рівняння регресії може бути перевірена за t-критерієм Стьюдента?

24. Як пов’язані критерії перевірки значущості для парної регресійної моделі?

25. У чому полягає суть точкового прогнозу?

26. Як побудувати довірчий інтервал для математичного сподівання залежної змінної?

27. За яких умов точність прогнозу на основі функції регресії буде збільшуватися та зменшуватися?

28. Як побудувати довірчий інтервал для індивідуальних значень залежної змінної?

29. Використовуючи оператор оцінювання 1МНК, знайдіть оцінки параметрів моделі $Y = a_0 + a_1 x + u,$ якщо задано вектори Y і X.

Y	5	7	6	9	10	8	11	12
X	3	4	3	5	6	4	8	8

30. Визначте вектор коваріації параметрів моделі, базуючись на результатах завдання 32.

31. Порівняйте значення оцінок стандартної помилки на основі вихідних даних завдання 32. Чи мають зміщення оцінки параметрів?

32. Знайдіть оцінки параметрів 1МНК на основі вихідних даних завдання 32, до яких приєднується ще одне спостереження:

Y	13
X	9

33. На основі даних дев’яти металобаз побудуйте економетричну модель, яка характеризує залежність між витратами обігу та вантажооборотом. Проаналізуйте достовірність моделі та її параметрів. Зробіть економічні висновки.

Таблиця 1

№ з/п	Витрати обігу	Вантажо-оборот
1	2,7	15,6
2	3,0	15,3
3	2,8	14,9
4	2,9	15,1
5	2,6	16,1
6	2,5	16,7
7	2,8	15,4
8	2,6	17,1
9	2,5	16,8

Таблиця 1

№ з/п	Витрати обігу	Вантажо-оборот
1	2,6	16,9
2	2,9	16,1
3	2,7	15,0
4	2,5	18,0
5	2,7	17,2
6	2,6	17,1
7	2,7	16,4
8	2,6	16,7
9	2,8	16,9

Таблиця 1

№ з/п	Витрати обігу	Вантажо-оборот
1	2,9	14,1
2	2,.6	17,2
3	2,8	17,1
4	2,7	17,8
5	2,7	16,2
6	2,9	17,2
7	2,4	16,8
8	2,9	14,8
9	2,3	19,6

34. Знайдіть оцінки параметрів моделі на основі 1МНК, якщо до вихідних даних завдання 32 приєднати ще два спостереження:

Y	13	15
X	9	10

35. Використовуючи оцінки параметрів моделі задачі 32, 35 і 37, визначте величину зміщення оцінок.

36. Знайдіть стандартні помилки оцінок параметрів моделей завдання 37 і порівняйте їх з помилками оцінок завдання 34.

37. Покажіть, чи будуть оцінки параметрів моделей, які обчислено в завданні 37, обґрунтованими?

38. Порівняйте дисперсії оцінок завдань 32, 35, 37 і визначте, які з оцінок є найефективнішими.

39. Побудуйте економетричну модель залежності витрат на одиницю продукції від рівня фондомісткості продукції за методом найменших квадратів. Вихідні дані (у грошових одиницях) наведені у таблиці:

x_i	90	75	120	100	80	78	110	115	115	125
y_i	50	40	65	55	45	42	56	60	64	65

Визначте залишки $\boldsymbol{u_i}$ моделі та знайдіть дисперсію залишків.

40. На підставі статистичних даних:

x_i	25	21	15	15	10	12	18	6	12	8
G_i	90	80	60	70	40	60	70	50	50	50

методом найменших квадратів знайдіть оцінки коефіцієнтів моделі

$G = a_0 + a_1x + u,$

де G_i – рейтинг i-го студента (шкала зміни від 1 до 100), x_i – дохід його батьків у тис. грн.

41. Побудуйте економетричну модель залежності індивідуальних доходів та витрат на споживання населення протягом десяти років на основі наступної таблиці:

Рік	Дохід (x)	Споживання (y)
1970	751,6	672,1
1971	779,2	696,8
1972	810,3	737,1
1973	864,7	767,9
1974	857,5	762,8
1975	874,9	779,4
1976	906,8	823,1
1977	942,9	864,3
1978	988,8	903,2
1979	1015,7	927,6

Тема 9. МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

1. *Класична нормальна лінійна модель множинної регресії та її формулювання у матричній формі.*
2. *Передумови застосування множинного регресійного аналізу.*
3. *Оцінювання параметрів множинної регресії методом найменших квадратів.*
4. *Коваріаційна матриця оцінок параметрів регресійної моделі.*
5. *Статистичні критерії і порядок перевірки значущості оцінок множинної регресії.*
6. *Довірчі інтервали для параметрів множинної регресії.*
7. *Коефіцієнт множинної кореляції і детермінації.*
8. *Скоригований коефіцієнт множинної детермінації.*
9. *Статистичні критерії і порядок перевірки значущості множинної регресії.*
10. *Прогнозування та довірчі інтервали для функції множинної регресії.*
11. *Методи побудови множинної регресійної моделі.*

9.1. Класична нормальна модель множинної регресії

Економічні явища і процеси визначаються, як правило, великою кількістю одночасно діючих факторів. Тому виникає задача дослідження залежності однієї залежної змінної Y від декількох пояснювальних (незалежних) змінних $X_1, X_2, .., X_p$. Таке дослідžення проводиться за допомогою *множинного (багатофакторного) регресійного аналізу.* Отже, множинний регресійний аналіз є розвитком парного регресійного аналізу стосовно випадків, коли залежна змінна гіпотетично пов'язана з більш ніж однією незалежною змінною. Більша частина множинного аналізу є безпосереднім розширенням парної регресійної моделі, але при цьому виникають нові якісні та кількісні проблеми та особливості.

Загальна багатофакторна лінійна регресійна модель має вигляд:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon, \qquad (9.1)$$

де Y – *залежна* змінна; $X_1, X_2, .., X_p$ – незалежні змінні (фактори); $\beta_0, \beta_1, \ldots, \beta_p$ – істинні коефіцієнти регресії (параметри моделі для генеральних сукупностей), які необхідно оцінити; ε – неспостережувана випадкова складова.

Узагальнена регресійна модель дійсна для генеральних сукупностей Y та $\overline{X_1, X_p}$, а випадкова величина ε є неспостережуваною. Для оцінки параметрів узагальненої моделі на основі випадкових вибірок з відповідних генеральних сукупностей будується *вибіркова лінійна багатофакторна регресійна модель,* яка має вигляд:

$$y = b_0 + b_1 x_1 + \ldots + b_p x_p + e, \qquad (9.2$$

де $x_1, x_2, \ldots, x_p$ – незалежні змінні; $b_0, b_1, \ldots, b_p$ – оцінки невідомих параметрів загальної моделі; e – оцінка випадкової складової.

Включення у регресійну модель великої кількості пояснювальних змінних ускладнює формули і відповідні розрахунки. Тому розглядаючи множинні регресійні моделі, доцільно використовувати теорію матриць. Матричний опис регресії полегшує необхідні розрахункові процедури.

Позначимо *i-е* спостереження залежної змінної через y_i, а незалежних змінних – через $x_{i1}, x_{i2}, \ldots, x_{ip}$. Тоді модель (9.1) для кожного окремого спостереження має вигляд:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, \mathrm{i} = 1, 2, \ldots, \mathrm{n}, \qquad (9.3)$$

де ε_i – i-е значення випадкової складової.

Рівняння (9.3) є компактним записом системи:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} + \varepsilon_2 \\ \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np} + \varepsilon_n. \end{cases} \qquad (9.4)$$

Введемо наступні позначення.

Нехай

$Y = \begin{pmatrix} y_2 \\ \cdots \end{pmatrix} = (y_1, y_2 \ldots y_n)'$ – матриця-стовпець, або вектор спостережень за залежною змінною розмірності n (знаком « $'$ » позначається операція транспонування матриць).

$X = \begin{pmatrix} 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$ – матриця спостережень за p незалежними змінними розмірності $n*(p+1)$. При цьому до матриці X додатково включений стовпець, усі елементи якого дорівнюють 1, тобто вважається, що у моделі (9.3) вільний член β_0 множиться на фіктивну змінну x_{i0}, яка приймає одиничне значення для всіх i ($x_{i0}=1$ $(i=1, 2,\ldots,n)$).

β =(β_0, β_1, ..., β_p – матриця-стовпець, або вектор невідомих параметрів розмірності $(p+1)$.

$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ – матриця-стовпець, або вектор випадкових помилок (залишків, збурень) розмірності n.

На підставі введених позначень систему (9.4) можна записати так:

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix}, \qquad (9.5)$$

а в матричній формі:

$$Y = X\beta + \varepsilon. \qquad (9.6)$$

Оцінкою даної моделі за вибіркою обсягом n спостережень або вибірковою лінійною багатофакторною моделлю (9.2) у матричній формі буде рівняння

$$Y = Xb + \varepsilon, \qquad (9.7)$$

де $b = (b_0, b_1, \ldots, b_p)'$; $e = (e_0, e_1, \ldots, e_p)'$.

Оскільки множинна регресійна модель є розвитком та узагальненням парної регресійної моделі, то всі основні передумови застосування регресійного аналізу (умови Гаусса-Маркова) для багатофакторної моделі зберігаються, але при цьому дещо модифікуються.

Розглянемо ці умови у матричній формі.

Перша умова. Математичне сподівання *і-го* значення випадкової величини ε_i дорівнює нулю, тобто вектор математичних сподівань випадкових величин дорівнює нулю:

$$\begin{pmatrix} M(\varepsilon_1) \\ M(\varepsilon_2) \\ \ldots\ldots \\ M(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}, \text{ або } M(\varepsilon) = O_n, \qquad (9.8)$$

де O_n – нульовий вектор розмірності n.

Друга умова. Значення ε_i вектора залишків незалежні між собою і мають сталу дисперсію *(гомоскедастичні)*. Дана умова передбачає незалежність випадкових величин між собою.

$$M(\varepsilon\varepsilon') = M\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}(\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_n)$$
$$= \mathrm{M}\begin{pmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \ldots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \ldots & \varepsilon_2\varepsilon_n \\ \ldots & \ldots & \ldots & \ldots \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \ldots & \varepsilon_n^2 \end{pmatrix}. \qquad (9.9)$$

Якщо застосувати оператор математичного сподівання до кожного елемента матриці (9.9) та врахувати гомоскедастичність, або другу умову Гаусса-Маркова для парної регресії $M(\varepsilon_i)^2=\sigma^2$ і незалежність випадкових величин, або третю умову Гаусса-Маркова для парної регресії ($M(\varepsilon_i, \varepsilon_j) = 0, i \neq j$), то отримаємо:

$$M(\varepsilon\varepsilon') = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & \dots \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & \dots \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & 1 \end{pmatrix} = \sigma^2 E_n, \qquad (9.10)$$

де E_n – одинична матриця n-го порядку.

Матриці (9.9) або (9.10) називаються дисперсійно-коваріаційною матрицею випадкових величин ε_i. Діагональні елементи цієї матриці – це дисперсії, а всі інші – відповідні коваріації [5, с. 86; 6, с. 189–190; 9, с. 89–90].

Третя умова. Незалежні змінні моделі не пов'язані із залишками, тобто

$$M(X'\varepsilon) = O_{p+1}, \qquad (9.11)$$

де O_{p+1} – нульовий вектор розмірності *(p+1).*

Третя умова передбачає незалежність між залишками та пояснювальними змінними. Вона означає, що у моделі (9.6) ε – випадковий вектор, а *X* – невипадкова (не стохастична, детермінована) матриця, утворена з фіксованих елементів $x_{ij}, i = \overline{1,n}; j = \overline{1,p}$ [5, с. 86; 6, с. 190; 9, с. 89-90].

Четверта умова. Стовпці матриці *X* повинні бути лінійно незалежними. Це означає, що ранг матриці *X* повинен дорівнювати кількості стовпців матриці, тобто *r(X)= p+1,* де *r(X)* – ранг матриці пояснювальних змінних *X*.

Четверта умова означає, що всі пояснювальні змінні, які входять до економетричної моделі, повинні бути лінійно незалежними між собою.

Незалежність пояснювальних змінних означає відсутність *мультиколінеарності.*

П'ята умова. Вектор випадкових величин *е* має нормальний закон розподілу, тобто $\varepsilon \sim N_n(0;\sigma^2 E_n)$(параметри нормального закону розподілу див. у розділі 4).

Модель (9.6), яка задовольняє умовам 1–5, називається класичною нормальною лінійною моделлю множинної регресії.

9.2. Оцінювання параметрів багатофакторної регресії методом найменших квадратів

Необхідно знайти оцінки b_j істинних коефіцієнтів регресії β_j, (*j = 0, 1,…, p)* у вибірковій лінійній багатофакторній регресійній моделі виду

$$y_j = b_0 + b_1 x_{i1} + \ldots + b_p x_{ip} + e_i \tag{9.12}$$

або у матричній формі:

$$Y = Xb + e.$$

Невідомі параметри b_j, як і у випадку парної лінійної регресійної моделі, визначаються методом найменших квадратів, мінімізуючи суму квадратів відхилень фактичних даних y_j від теоретичних $\widehat{y_j}$ (знайдених за рівнянням регресії (9.12)).

$$S = \sum_{i=1}^{n}(y_j - \widehat{y_j})^2 = \sum_{i=1}^{n}(y_j - b_0 - b_1 b_{i1} - b_2 b_{i2} - \ldots - b_p b_{ip})^2 = \sum_{i=1}^{n} e_i^2 \to \min. \tag{9.13}$$

З рівняння (9.7) отримаємо:

$$e = Y - Xb. \tag{9.14}$$

Сума квадратів помилок $\sum_{i=1}^{n} e_i^2$ у матричній формі:

$$\sum_{i=1}^{n} e_i^2 = = e_1^2 + e_2^2 + \dots + e_n^2 = (e_i e_2 \dots e_n)'(e_i e_2 \dots e_n) = e'e. \quad (9.15)$$

Якщо в цільову функцію (9.13) підставити значення (9.14) і (9.15), провести необхідні перетворення та врахувати властивості транспонування матриць, добутків матриць, отримаємо вираз для суми квадратів [5, с. 84; 6, с. 191; 9, с. 90]:

$$\begin{gathered} \min S = e'e = (Y - Xb)'(Y - Xb) = \\ = Y'Y - 2b'X'Y + b'X'Xb. \end{gathered} \quad (9.16)$$

Для досягнення необхідного екстремуму функції декількох змінних $S(b_0, b_1, \dots, b_p)$ необхідно знайти нулі перших частинних похідних цих змінних або у матричній формі – вектор частинних похідних:

$$\frac{\partial S}{\partial b} = \left(\frac{\partial S}{\partial b_0}, \frac{\partial S}{\partial b_1}, \dots, \frac{\partial S}{\partial b_p}\right).$$

В результаті отримуємо:

$$\frac{\partial S}{\partial b} = 2X'Y + 2X'Xb = 0, \text{або}$$

систему нормальних рівнянь у матричній формі для визначення вектора *b*:

$$X'Xb = X'Y. \quad (9.17)$$

Матриця $X'X$ – це матриця сум перших степенів, квадратів та попарних добутків значень пояснювальних змінних у *n* спостереженнях.

Тоді

$$X'Xb = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_p \end{pmatrix} = \quad (9.18)$$

$$= \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1} \sum x_{ip} \\ \dots & \dots & \dots & \dots \\ \sum x_{ip} & \sum x_{i1} \sum x_{ip} & \dots & \sum x_{ip}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_p \end{pmatrix}.$$

Матриця $X'Y$ є вектором добутків n значень пояснювальних та незалежної змінних в n спостереженнях:

$$X'Y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \dots \\ \sum y_i x_{ip} \end{pmatrix}. \quad (9.19)$$

З урахуванням умов 1–4 для множинного регресійного аналізу розв'язком системи нормальних рівнянь (9.17) є вектор [5, с. 87; 6, с. 192]

$$B = (X'X)^{-1} X'Y, \quad (9.20)$$

де $(X'X)^{-1}$– матриця, обернена до матриці коефіцієнтів системи (9.17) (див. (9.18)); $X'Y$ – матриця-стовпець або вектор вільних членів системи (9.17) (див. (9.19)).

Теорема Гаусса-Маркова, яка була розглянута у розділі 8 для парної регресійної моделі, має місце і для моделі множинної регресії.

За виконання передумов множинного регресійного аналізу 1–4 оцінка за методом найменших квадратів $b = (X'X)^{-1} \cdot X'Y$ *є найбільш ефективною, тобто має найменшу дисперсію в класі лінійних незміщених оцінок.*

Доведення теореми Гаусса-Маркова див. у роботі [5, с. 94–95]. Підставимо у вектор оцінок (9.20) значення Y (9.6):

$$b = (X'X)^{-1} \cdot X'(X\beta + \varepsilon) = (X'X)^{-1} (X'X)\,\beta + (X'X)^{-1} X'\varepsilon = \beta + (X'X)^{-1} X'\varepsilon. \quad (9.21)$$

Вираз (9.21) означає, що оцінки параметрів (9.20), знайдені за вибіркою, будуть містити випадкові помилки.

Покажемо, що математичне сподівання оцінки b дорівнює параметру β, що оцінюється, підставивши замість b вираз (9.21):

$$M(b) = M[\beta + (X'X)^{-1} X' \varepsilon] = M(\beta) + (X'X)^{-1} X' \mathrm{M}(\varepsilon).$$

Оскільки $\mathrm{M}(\varepsilon) = 0$ за першою передумовою множинного регресійного аналізу (9.8), то $M(b)=M(\beta)$, і вектор b є незміщеною оцінкою істинного параметра β.

Маючи вектор b, вибіркове рівняння множинної регресії можна подати у вигляді:

$$\hat{y} = X_0' b, \qquad (9.22)$$

де y – групове (умовне) середнє значення змінної Y при заданому векторі значень пояснювальної змінної

$X_0' = (1, x_{10}, x_{20}, \dots x_{p0})$.

Крім того, на практиці виникає необхідність визначення впливу на залежну змінну різних пояснювальних змінних, особливо коли останні виражаються різними одиницями виміру.

У таких випадках використовують *стандартизовані коефіцієнти регресії* b_0' *та коефіцієнти еластичності* E_j *(j = 1, 2, ... p):*

$$b_0' = b_j \frac{s_{x_j}}{s_y}; \qquad (9.23)$$

$$E_j = b_j \frac{\bar{x}_j}{\bar{y}}. \qquad (9.24)$$

Стандартизований коефіцієнт регресії $b_j{}'$ показує, на яку частину S_y (стандартного відхилення залежної змінної) зміниться в середньому залежна змінна Y у разі зміни тільки j-ї пояснювальної змінної на одиницю (стандартного відхилення незалежної змінної X_j) при фіксованих значеннях інших факторів.

Коефіцієнт еластичності E_j показує, на скільки відсотків (від середнього арифметичного) зміниться в середньому Y у разі збільшення тільки X_j на *1%* при фіксованих значеннях інших факторів.

9.3. Коваріаційна матриця оцінок параметрів регресійної моделі

Оцінки b параметрів β є випадковими величинами [див. (9.21)], тому варіації оцінок параметрів будуть визначати точність рівняння множинної регресії. Для вимірювання оцінок коефіцієнтів регресії у багатофакторному регресійному аналізі розглядається так звана *коваріаційна матриця* вектора оцінок параметрів регресії, яка є матричним аналогом дисперсії однієї змінної.

За означенням *коваріація двох змінних* – це математичне сподівання добутку відхилень цих змінних від їх математичних сподівань, тобто

$$pop.cov(b_i b_j) = \sigma_{b_i b_j} M\left[\left(b_i - M(b_i)\right)\left(b_j - M(b_j)\right)\right]. \quad (9.25)$$

Коваріаційна матриця оцінок параметрів регресійної моделі *pop.cov(b)* має вигляд:

$$\mathrm{pop.cov(b)} = \begin{pmatrix} \sigma_{b_0 b_0} & \sigma_{b_0 b_1} & \cdots & \sigma_{b_0 b_p} \\ \sigma_{b_1 b_0} & \sigma_{b_1 b_1} & \cdots & \sigma_{b_1 b_p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{b_p b_0} & \sigma_{b_p b_1} & \cdots & \sigma_{b_p b_p} \end{pmatrix}. \quad (9.26)$$

Оцінки b_j, одержані методом найменших квадратів, є незміщеними оцінками параметрів β_j, тобто $M(b_i) = \beta_j$, тому вираз (9.25) можна записати так:

$$\sigma_{b_i b_j} = M\left[(b_i - \beta_i)(b_j - \beta_j)\ \right] \qquad (9.27)$$

або у скороченому вигляді:

$$pop.cov(b) = M\left[(b - \beta)(b - \beta)'\right]. \qquad (9.28)$$

На головній діагоналі коваріаційної матриці містяться *дисперсії* оцінок параметрів регресії, оскільки

$$\sigma_{b_i b_j} = M[(b_j - \beta_j)] = M(b_j - \beta_j)^2 = \sigma^2_{b_j}. \qquad (9.29)$$

Доведено [5, с. 92–93], що коваріаційна матриця визначається за формулою:

$$pop.cov(b) = \sigma^2_\varepsilon (X'X)^{-1}, \qquad (9.30)$$

де σ^2_ε – дисперсія випадкового члена ε.

Отже, за допомогою оберненої матриці $(X'X)^{-1}$ (див. (9.18)) визначається не тільки вектор b оцінок параметрів (9.20), але й дисперсії та коваріації його компонентів.

Незміщеною оцінкою дисперсії залишкового члена σ^2_ε є вибіркова залишкова дисперсія S^2 [5, с. 97; 9, с. 97]:

$$S^2 = \frac{e'e}{n-p-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1}. \qquad (9.31)$$

Тоді *матриця оцінок* дисперсій і коваріацій оцінок параметрів регресії визначається за формулою

$$cov(b) = S^2 (X'X)^{-1}. \qquad (9.32)$$

На головній діагоналі матриці оцінок $cov(b)$ містяться оцінки дисперсій $S_{b_i}^2$ (оцінки параметрів). Елементи, які лежать поза головною діагоналлю, є оцінками коваріацій між оцінками b_j, $j = \overline{0,p}$. Матриця оцінок дисперсій і коваріацій має вигляд:

$$\mathrm{cov(b)} = \begin{pmatrix} S_{b_0}^2 & cov(b_0 b_1) & \dots & cov(b_0 b_p) \\ cov(b_0 b_1) & S_{b_1}^2 & \dots & cov(b_0 b_p) \\ \dots\dots\dots\dots & \dots\dots\dots & \dots & \dots\dots\dots\dots. \\ cov(b_p b_0) & cov(b_p b_1) & \dots & S_{b_p}^2 \end{pmatrix} \quad (9.33)$$

Згідно з (9.29), (9.30) та (9.32) оцінка $S_{b_i}^2$ дисперсії $\sigma_{b_i}^2$

$$S_{b_i}^2 = S^2[(X'X)^{-1}]_{jj}, \quad (9.34)$$

де $[(X'X)^{-1}]_{jj}$ – діагональний елемент матриці $(X'X)^{-1}$.

Оцінка стандартного відхилення *(стандартна похибка оцінки коефіцієнта множинної регресії)* має вигляд:

$$S_{b_j} = S\sqrt{(X'X)^{-1}{}_{jj}}. \quad (9.35)$$

9.4. Значущість оцінок параметрів множинної регресії

Значущість оцінок b_j коефіцієнтів регресії β_j можна перевірити на основі *t*-статистики Стьюдента з $k = n - p - 1$ ступенями свободи:

$$t_j = \frac{b_j - \beta_j}{s_{b_j}}. \quad (9.36)$$

Розподіл Стьюдента (9.36) дозволяє перевірити гіпотези щодо кожного параметра β_j множинної регресії та побудувати довірчі інтервали.

Для перевірки значущості параметрів регресії необхідно обрати рівень значущості α (або $\alpha \cdot 100\%$) або довірчої ймовірності $P_\alpha = 1 - \alpha$ [або $P_\alpha = =(1 - a)100\%$] і побудувати *t*-статистики для кожного параметра окремо:

$$|t_j| = \frac{|b_j - \beta_{j0}|}{s_{b_j}}, \qquad (9.37)$$

де β_{j0} – задане гіпотетичне значення, якого може набувати параметр β_j, тобто H_0: $\beta_j = \beta_{j0}$.

В економетрії, як правило, використовується нуль-гіпотеза H_0: $\beta_{j0} = 0$ проти альтернативної H_1: $\beta_{j0} \neq 0$ [5, с. 98; 6, с. 195; 9, с. 119].

Тому *t*-статистика для перевірки значущості оцінок параметрів має вигляд:

$$|t_j| = \frac{|b_j|}{s_{b_j}}. \qquad (9.38)$$

Фактичне значення *t*-статистики (9.33) порівнюється з табличним $t_{1-\alpha,n-p-1}$.

Якщо $|t_j| > t_{1-\alpha,n-p-1}$, то оцінка b_j відповідного параметра β_j відрізняється від нуля на рівні значущості α (тобто нуль-гіпотеза $H_0 : \beta_j = 0$ відхиляється).

Якщо $|t_j| < t_{1-\alpha,n-p-1}$,, то оцінка є статистично незначущою з довірчою ймовірністю P_α (нуль-гіпотеза приймається).

Для того, щоб визначити, як оцінки коефіцієнтів регресії b_j пов’язані з істинними параметрами β_j, необхідно побудувати довірчі інтервали для цих параметрів.

Довірчі інтервали будуються на основі *t*-статистики (9.36) і стандартної похибки оцінки параметра моделі S_{b_j} для даного рівняння з довірчою ймовірністю $P_a = 1 - \alpha$.

За означенням довірчий інтервал можна записати так:

$$P(-t_{1-\alpha,n-p-1} < t_j < t_{1-\alpha,n-p-1}) = P_\alpha = 1 - \alpha \, .$$

Якщо підставити замість t_j його значення (9.36), отримаємо:

$$P(-t_{1-\alpha,n-p-1} < \frac{b_j - \beta_j}{S_{b_j}} < t_{1-\alpha,n-p-1}) = P_\alpha$$

або

$$b_j - t_{1-\alpha,n-p-1} \cdot S_{b_j} < \beta_j < b_j + t_{1-\alpha,n-p-1} \cdot S_{b_j} \qquad (9.39)$$

або

$$\beta_j = b_j \pm t_{1-\alpha,n-p-1} \cdot S_{b_j}, j = 0, 1, \ldots, \mathrm{p} \qquad (9.40)$$

з $k = n - p - 1$ ступенями свободи.

Нагадаємо, що за аналогією з парною регресією добуток табличного значення *t*-статистики і стандартної похибки оцінки параметра моделі називається *граничною похибкою оцінки коефіцієнта множинної регресії,* тобто

$$\Delta S_{b_j} = t_{1-\alpha,n-p-1} \cdot S_{b_j}. \qquad (9.41)$$

9.5. Коефіцієнт множинної кореляції та детермінації. Скоригований коефіцієнт детермінації. Оцінка значущості рівняння множинної регресії

Мірою тісноти кореляційного зв'язку всіх незалежних змінних *Xj* ($j = \overline{1, p}$) із залежною змінною *Y*, а також показником ступеня відповідності теоретичних даних $\hat{y}_i$ (отриманих на основі регресійної моделі) фактичним даним y_i є *коефіцієнт множинної кореляції.*

Коефіцієнт множинної кореляції може бути визначений: по-перше, як коефіцієнт кореляції між фактичними (y_i) і теоретичними ($\hat{y}_i$) значеннями залежної змінної [6, с. 180; 13, с. 111]:

$$R_{\hat{y}y} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \overline{\hat{y}_i})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2(\hat{y}_i - \overline{\hat{y}_i})}}; \qquad (9.42)$$

по-друге, як арифметичне значення кореня квадратного з коефіцієнта множинної детермінації [6, с. 180; 9, с. 113; 13, с. 101]:

$$R = \sqrt{R^2} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}; \qquad (9.43)$$

по-третє, через парні коефіцієнти кореляції [11, с. 107] (див. також формули (7.31), (7.32)):

$$R_{y;\overline{1,p}} = \sqrt{1 - \frac{|R_y|}{R_{yy}}}, \qquad (9.44)$$

де $|R_y|$ – визначник матриці вибіркових парних коефіцієнтів кореляції.

$$R_y = \begin{array}{c} \\ y \\ x_1 \\ x_2 \\ \cdots \\ x_p \end{array} \begin{bmatrix} y & x_1 & x_2 & \cdots & x_p \\ r_{yy} & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_p} \\ r_{x_1y} & r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_p} \\ r_{x_2y} & r_{x_2x_1} & r_{x_2x_2} & \cdots & r_{x_2x_p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_py} & r_{x_px_1} & r_{x_px_2} & \cdots & r_{x_px_p} \end{bmatrix}; \qquad (9.45)$$

де R_{yy} – алгебраїчне доповнення елемента r_{yy} матриці вибіркових парних коефіцієнтів кореляції.

Як і у випадку парної регресії, у моделі множинної регресії загальна варіація *TSS*, тобто сума квадратів відхилень залежної змінної від її середньої, може бути розкладена на дві складові – суму квадратів відхилень, зумовлену регресією *ESS*, та залишкову суму квадратів *RSS*, що характеризує вплив неврахованих факторів:

$$TSS = ESS + RSS.$$

Якщо використати позначення, які були введені для парної лінійної регресії, то величини *TSS, ESS* та *RSS* визначаються за формулами (7.19), (7.20).

Проведемо у матричній формі розрахунки, необхідні для визначення коефіцієнта множинної детермінації R^2.

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = YY' - n\bar{y}^2. \qquad (9.46)$$

Оскільки

$$\sum^{n}(y_i - \bar{y})^2 = \sum^{n} y_i^2 - \overline{2y}\sum^{n} y_i + \overline{ny}^2 =$$

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

$$\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n}(y_i)^2}{n}$$

;

$$\sum_{i=1}^{n} y_i^2 = y_1^2 + y_2^2 + \ldots + y_n^2 = (y_1 y_2 \ldots y_1)'(y_1 y_2 \ldots y_1)' = Y'Y.$$

Беручи до уваги (9.13) – (9.16),

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = e'e = Y'Y - 2b'Y'Y + b'X'Xb = Y'Y - 2b'X'Y + +b'X'Y = Y'Y - b'X'Y \text{ (тому що в силу (9.17)} X'Xb = X'Y, b'^{X'}Xb = b'^{X'}Y\text{)}. \quad (9.47)$$

$$ESS = TSS - RSS = Y'Y - n\bar{y}^2 - (Y'Y - b'X'Y) = = b'X'Y - n\bar{y}^2. \quad (9.48)$$

Тоді коефіцієнт множинної детермінації у матричній формі має вигляд:

$$R^2 = \frac{ESS}{TSS} = \frac{b'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2} \quad (9.49)$$

або

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{(Y - Xb)'(Y - Xb)}{(Y - \bar{Y})'(Y - \bar{Y})} = 1 - \frac{e'e}{y'y}, \quad (9.50)$$

де $e = Y - Xb, \bar{Y} = (\bar{y}, \bar{y}, \ldots \bar{y},), y = (Y - \bar{Y})$ – n-мірні вектори;

$$e'e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

Нагадаємо, що коефіцієнт детермінації R^2 характеризує частку варіації залежної змінної, обумовленої мінливістю пояснювальних змінних. При цьому чим ближче R^2 до одиниці, тим краще регресія описує залежність між пояснювальними і пояснюваними змінними.

Водночас використання для вибору найкращого рівняння регресії лише одного коефіцієнта детермінації недостатньо.

При додаванні до рівняння регресії нової пояснювальної змінної коефіцієнт детермінації R^2 ніколи не зменшується, а, як правило, зростає, хоча це не обов’язково означає підвищення якості моделі. Тому для врахування кількості факторів, які входять у різні моделі, використовується *скоригований (адаптований, виправлений) коефіцієнт детермінації* $\overline{R}^2$.

На відміну від R^2, величина $\overline{R}^2$ коригується на кількість ступенів свободи суми квадратів залишків та загальної суми квадратів множинної регресійної моделі.

Розглянемо принципову модель дисперсійного аналізу для випадку множинної регресії (див. табл. 4) [6, с. 184; 9, с. 112].

Отже, скоригований коефіцієнт детермінації має вигляд:

$$\overline{R}^2 = 1 - \frac{\frac{RSS}{(n-p-1)}}{\frac{TSS}{(n-1)}} = 1 - \frac{S^2}{S_{yx}^{2*}} = 1 - \frac{(n-1)}{(n-p-1)}(1-R^2) \qquad (9.51)$$

$(R^2 = 1 - RSS/TSS \rightarrow RSS/TSS = 1 - R^2)$ або, з урахуванням (9.50),

$$\overline{R}^2 = 1 - \frac{(n-1e'e)}{(n-p-1)y'y}. \qquad (9.51)$$

Таблиця 4

Модель дисперсійного аналізу для багатофакторної регресії

Джерело варіації	Сума квадратів	Кількість ступенів свободи	Середні квадрати
Множинна регресійна модель	$ESS = \sum_{i=1}^{n}(y_i - \bar{y})^2$	*p*	$MSR = \frac{ESS}{p} = S_R^{2*}$
Залишок (похибка)	$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = Y'Y - b'X'Y$	*n-p-1*	$MSE = \frac{RSS}{n-p-1} = S_{yx}^{2*}$.
Загальна варіація	$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2 == Y'Y - n\bar{y}^2$	*n-1*	$S_{yx}^{2*} = \frac{TSS}{n-1}$

З формули (9.51) випливає, що зі збільшенням кількості незалежних змінних p $\overline{R}^2$ зменшується у порівнянні з R^2. На відміну від R^2, скоригований коефіцієнт $\overline{R}^2$ може зменшуватись при введенні у модель нових пояснювальних змінних, які суттєво не впливають на залежну змінну. Проте збільшення $\overline{R}^2$ при введенні у модель нової незалежної змінної x_{p+1} не обов'язково означає, що її коефіцієнт регресії b_{p+1} значущо відрізняється від нуля. Можна показати, що коефіцієнт регресії нової пояснювальної змінної є статистично значущим тоді і тільки тоді, коли відповідна t-статистика за абсолютною величиною більша від одиниці ($|t_{p+1}|>1$). Отже, збільшення $\overline{R}^2$ ще не означає покращення специфікації моделі [2, с. 164; 5, с. 104].

Для регресії з двома і більше незалежними змінними коефіцієнт детермінації лежить на відрізку [0,1]. Чим ближчий він до одиниці, тим більше варіація залежної змінної визначається варіацією незалежних змінних.

З формули (9.51) бачимо, що $\overline{R}^2<R^2$, якщо $(p+1)>1$, тобто $p>0$. Якщо кількість незалежних змінних p зростає, то скоригований коефіцієнт детермінації $\overline{R}^2$ зростає повільніше, ніж R^2. Отже, вплив кількості факторів на величину коефіцієнта детермінації зменшується.

Крім того, з формули (9.51) видно, що коли $R^2=1$, то і $\overline{R}^2=1$,а якщо $R^2\to 0$, то $\overline{R}^2$ прямує до від'ємної величини. Отже, скоригований коефіцієнт детермінації $\overline{R}^2$, на відміну від звичайного R^2, може приймати *від'ємні* значення.

Для перевірки значущості (адекватності) множинного рівняння регресії, як і у випадку парної регресійної моделі, використовується F-критерій Фішера-Снедекора (відношення середнього квадрата, що пояснює регресію, до середнього квадрата помилок). Для випадку множинної регресії F- критерій, у відповідності з даними табл. 4, має вигляд:

$$F = 1 - \frac{\frac{ESS}{p}}{\frac{RSS}{(n-p-1)}} \qquad (9.53)$$

з кількістю ступенів свободи $k_1 = p$, $k_2 = n - p - 1$.

Поділивши чисельник та знаменник співвідношення (9.53) на загальну суму квадратів *TSS* та зауваживши, що $ESS/RSS = R^2$, а $RSS/TSS = (1 - R^2)$, отримаємо:

$$F = 1 - \frac{\frac{R^2}{p}}{\frac{(1-R^2)}{(n-p-1)}} . \qquad (9.54)$$

Отже, співвідношення (9.54) показує, що *F*-критерій є як мірою ефективності регресійної моделі, так і критерієм значущості коефіцієнта детермінації R^2. Таким чином, ми можемо оцінювати значущість рівняння регресії, виходячи лише із відомого значення R^2.

Перевірка значущості множинного рівняння регресії здійснюється у такій послідовності:

1. Перевіряється нуль-гіпотеза: $H_0: \beta_1 = \beta_2 = \ldots = \beta_p = 0$ проти альтернативної гіпотези H_1: не всі коефіцієнти регресії одночасно дорівнюють нулю (принаймні один з параметрів β_j не дорівнює нулю).
2. Розраховується фактичне значення *F*-статистики.
3. За *F*-таблицями Фішера-Снедекора визначається табличне значення *F*-статистики з $k_1 = p$ та $k_2 = n - p - 1$ ступенями свободи для заданого рівня значущості α (довірчої ймовірності $P_\alpha = 1 - \alpha$) – $F_{\alpha,k1,k2}$.
4. Фактичне значення *F*-статистики порівнюється з табличним.

Якщо $F > F_{\alpha,p,n-p-1}$, то нуль-гіпотеза відхиляється, що свідчить про значущість (адекватність) рівняння множинної регресії. У протилежному випадку модель вважається неадекватною (нуль-гіпотеза приймається) з довірчою ймовірністю $P_\alpha = 1 - \alpha$.

Коефіцієнт множинної кореляції R є вибірковою характеристикою, яка може відхилятись від істинного значення (теоретичного коефіцієнта множинної кореляції). Тому необхідно перевірити його значущість. Перевірка значущості коефіцієнта множинної кореляції здійснюється на основі t-критерію Стьюдента з $k = n - p - 1$ ступенями свободи [9, с. 117]:

$$t = \frac{R\sqrt{n-p-1}}{\sqrt{1-R^2}}. \qquad (9.55)$$

Коефіцієнт множинної кореляції R між залежною та незалежними змінними буде статистично значущим (нуль-гіпотеза H_0: $p = 0$ відхиляється), якщо

$$t > t_{a,p,n-p-1},$$

де $t_{a,p,n-p-1}$ – табличне значення t-розподілу для рівня значущості α з $k = n\text{-}p\text{-}1$ ступенями свободи.

9.6. Прогнозування за множинною регресійною моделлю

Якщо множинна регресійна модель адекватна, то її можна використовувати для прогнозування залежної змінної.

Як відомо з попереднього розділу, можна отримати точкові та інтервальні прогнози.

Нехай для незалежних змінних $X_1, X_2, \ldots, X_p$ ми отримали (n+1)-і спостереження, які задаються вектором X'_{n+i} = $(1, x_{n+1,1} x_{n+1,2} \ldots x_{n+1,p})$.

Тоді точковий прогноз $\hat{y}_{n+1}$ дорівнює

$$\hat{y}_{n+1} = b_0 + b_1 x_{n+1,1} + b_2 x_{n+1,2} + \ldots + b_p x_{n+1,p}, \qquad (9.56)$$

або у матричній формі:

$$\hat{y}_{n+1} = X'_{n+1} b. \qquad (9.57)$$

«Істинне» значення *n+1* для спостережень X'_{n+1} у матричній формі:

$$y_{n+1} = X'_{n+1}\beta + \varepsilon_{n+1}. \qquad (9.58)$$

Прогноз $\hat{y}_{n+1}$ є точковою оцінкою істинного значення змінної y_{n+1} для спостережень незалежних змінних $\overline{X_1, X_p}$ у (*n+1*)-й період.

Як і у випадку парної лінійної регресії, інтервальні прогнози будуються для умовного математичного сподівання $M_x(Y)$ та для індивідуальних значень залежної змінної.

Довірчий інтервал для математичного сподівання у. У розділі 8 такий інтервал побудовано для рівняння парної регресії. Узагальнюючи відповідні вирази для множинної регресії, можна отримати довірчий інтервал для M(y_{n+1}/X'_{n+1}).

У матричному вигляді дисперсія такого прогнозу та оцінка цієї дисперсії має вигляд [6, с. 197; 9, с. 100]:

$$\sigma^2_{\hat{y}_{n+1}} = \sigma^2_\varepsilon X'_{n+1}(X'X)^{-1}X_{n+1}, \qquad (9.59)$$

$$S^2_{\hat{y}_{n+1}} = S^2 X'_{n+1}(X'X)^{-1}X_{n+1}, \qquad (9.60)$$

де S^2 – незміщена оцінка дисперсії залишкової величини σ^2_ε .

Стандартна похибка прогнозу:

$$S_{\hat{y}_{n+1}} = S\sqrt{X'_{n+1}(X'X)^{-1}X_{n+1}}. \qquad (9.61)$$

Довірчий інтервал для математичного сподівання при рівні значущості *a* і *k = n-p-1* ступенях свободи дорівнює:

$$\hat{y}_{n+1} - t_{a,k}\cdot S\sqrt{X'_{n+1}(X'X)^{-1}X_{n+1}} \le M(y_{n+1}/X'_{n+1}) \le$$
$$\le \hat{y}_{n+1} + t_{a,k}\cdot S\sqrt{X'_{n+1}(X'X)^{-1}X_{n+1}} \qquad (9.62)$$

або

$$\hat{y}_{n+1} - t_{a,k} \cdot S_{\hat{y}_{n+1}} \le M(y_{n+1}) \le \hat{y}_{n+1} + t_{a,k} \cdot S_{\hat{y}_{n+1}}. \qquad (9.63)$$

Довірчий інтервал для індивідуальних значень залежної змінної y_0.

Дисперсія інтервального прогнозу індивідуальних значень визначається за формулою [6, с. 197; 9, с. 101]:

$$\sigma^2_{\hat{y}_0} = \sigma^2_\varepsilon \left(X'_0 (X'X)^{-1} X_0 \right). \qquad (9.64)$$

Стандартна похибка прогнозу індивідуальних значень:

$$S_{\hat{y}_0} = S\sqrt{1 + X'_0 (X'X)^{-1} X_0}. \qquad (9.65)$$

Тоді довірчий інтервальний прогноз для індивідуальних значень залежної змінної матиме вигляд:

$$\hat{y}_0 - t_{a,k} \cdot S_{\hat{y}_0} \le y_0 \le \hat{y}_0 + t_{a,k} \cdot S_{\hat{y}_0} \qquad (9.66)$$

або

$$\hat{y}_0 - t_{a,k} \cdot S\sqrt{1 + X'_0 (X'X)^{-1} X_0} \le y_0 \le \hat{y}_0 + t_{a,k} S\sqrt{1 + X'_0 (X'X)^{-1} X_0}, \qquad (9.67)$$

де $\hat{y}_0 = X'_0 b$;

$X'_0 = (x_{01}, x_{02}, \dots, x_{0p})$;
$y_0 = X'_0 \beta + \varepsilon$.

9.7. Методи побудови множинної регресійної моделі

При побудові рівнянь множинної регресії необхідно перевірити значущість факторів, які включені в дану модель і впливають на результативну ознаку. Кількісна статистична оцінка таких факторних змінних проводиться за допомогою *покрокового (багатокрокового, крокового) регресійного аналізу,* згідно з яким формулюються кількісні критерії значущості факторів множинної регресії і послідовність (алгоритм, процедура) їх приєднання (включення) або вилучення (виключення) з остаточного рівняння.

Розглянемо найбільш поширені *процедури приєднання-вилучення* пояснювальних змінних.

Процедура покрокового відбору найбільш інформативних змінних [5, с. 112–115; 6, с. 200–201].

1. Будується початкова регресійна модель, яка включає всі p факторів:

$$y = b_0 + b_1x_1 + ... + b_pX_p + e.$$

Для даної моделі обчислюється матриця парних коефіцієнтів кореляції (9.45).

2. Із пояснювальних змінних $\overline{X_1, X_p}$ обирається змінна X_k, яка має із залежною змінною Y найбільший коефіцієнт де термінації R^2_{yk}. Для парної моделі він дорівнює квадрату коефіцієнтта кореляції $R^2_{yk} = r^2_{yk}$. Потім визначається скоригований коефіцієнт детермінації $\overline{R}^2_{yk}$.

Крім того, перевіряється значущість оцінки коефіцієнта регресії відібраної змінної за t-критерієм Стьюдента:

$$|t_k| = \frac{|b_k|}{s_k} > t_{a,k}, k = n - 2.$$

Регресійна модель має вигляд:

$$y = b_0 + b_kx_k.$$

3. У регресійну модель вводиться нова пояснювальна змінна X_1, яка разом із початковою відібраною змінною $X_к$ утворює пару пояснювальних змінних, що має найвищий скоригований коефіцієнт детермінації з $\overline{R}^2_{yk}$. При цьому його значення повинно зростати: $\overline{R}^2_{yk1} > \overline{R}^2_{yk}$.

Далі перевіряється значущість оцінок коефіцієнтів регресії відібраних змінних:

$$|t_k| = \frac{|b_k|}{S_k} > t_{a,k}; |t_1| = \frac{|b_1|}{S_1} > t_{a,k},$$

$$k = n\text{-}2\text{-}1.$$

Відтак регресійна модель має такий вигляд:

$$y = b_0 + b_k x_k + b_1 x_1.$$

Зауважимо, що при включенні у модель нової змінної визначені на попередньому кроці оцінки коефіцієнтів регресії b_0 та b_k можуть змінюватись.

4. У модель вводиться ще одна пояснювальна змінна X_m, яка разом із двома початково відібраними утворює трійку пояснювальних змінних, які мають з Y найбільший скоригований коефіцієнт детермінації $\overline{R}^2_{yk1m}$ (якщо $\overline{R}^2_{yk1m}$>$\overline{R}^2_{yk1}$).

Аналогічно за t-критерієм перевіряється значущість оцінок коефіцієнтів регресії відібраних змінних b_k, b_1, b_m.

Тоді регресійна модель має вигляд:

$$y = b_0 + b_k x_k + b_1 x_1 + b_m x_m.$$

5. Процедура включення у модель нових змінних триває доти, поки буде збільшуватись відповідний скоригований коефіцієнт детермінації R^2.

Процедура покрокового вилучення змінних передбачає перевірку значущості для кожного незалежного фактора окремо. Для *лінійних* моделей значущість окремих факторних змінних перевіряється *одночасно* з перевіркою значущості оцінок коефіцієнтів регресії [10].

Перевірка гіпотези про значущість j-го параметра $(H_0: \beta_j = 0$; $H_1: \beta_j = 0)$ може здійснюватись на основі t-статистики (3.37) ($|t_j| = b_j/S_j$) або з використанням *частинного F-критерію для* b_j (можна показати, що $t = (F)^{1/2}$ [6, с. 200–204; 13, с. 119–120]). У багатьох комп'ютерних програмах використовуються обидві статистики.

Розглянемо алгоритм вилучення змінних на основі *t*-статистики [10].

1. Будується регресійне рівняння, яке включає всі *p* фактори. Для всіх факторних змінних визначаються статистики t_j, $j = \overline{1, p}$.

2. Визначається мінімальне за абсолютною величиною значення t_j

$$\min_{j=1,p} |t_j| = t_{S_1},$$

яке порівнюється з табличним значенням *t*-статистики – $t_{a,n-p-1}$.

3. Якщо $|t_{S_1}| < P_\alpha = 1\text{-}\alpha$, то із заданою ймовірністю $P_a = 1\text{-}\alpha$ не може бути відхилена нуль-гіпотеза про те, що відповідний коефіцієнт регресії у генеральній сукупності дорівнює нулю, а S_1 фактор є незначущим. Фактор S_1 вилучається з рівняння регресії. Одночасно на одну змінну зменшується система нормальних рівнянь.

4. Проводиться побудова нового рівняння регресії, визначаються нові статистики t_j, $j = \overline{1, p-1}$, серед яких знаходиться мінімальне за модулем значення

$$\min_{j=\overline{1,p-1}} |t_j| = t_{S_2},$$

яке порівнюється з новим табличним значенням $t_{a,n-p-2}$.

Якщо $|t_{S_2}| < t_{a,n-p-2}$, то фактор S_2 вилучається з рівняння регресії.

5. Процес виключення факторних змінних триває доти, доки не буде виконуватись умова значущості:

$$|t_{min}| > t_{a,k}.$$

Примітка. Не можна виключати з рівняння регресії одночасно декілька факторів, для яких $|t_j| > t_{a,k}$, тому що після вилучення одного незначущого фактора оцінки коефіцієнтів регресії можуть змінюватись, і незначущі фактори можуть перетворитись у значущі.

Зауважимо, що для практичної реалізації моделей множинної регресії доцільно, крім статистичних функцій *Excel*, використання економетричних комп’ютерних пакетів прикладних програм, таких як STATGRAPHICS [7], *Econometric Views* [5] та інших. Характеристика найвідоміших пакетів подається у роботі І.Т. Лук’яненко та Л.І. Краснікової [7, с. 10–11].

Запитання та завдання для самоконтролю

1. Запишіть загальну багатофакторну лінійну регресійну модель та її вибіркову оцінку.
2. Запишіть багатофакторну регресійну модель та її оцінку в матричній формі.
3. Сформулюйте передумови застосування регресійного аналізу для багатофакторної моделі. Поясніть їх особливості.
4. Сформулюйте поняття класичної нормальної лінійної моделі множинної регресії.
5. Запишіть критерій МНК та систему нормальних рівнянь множинної регресії у матричній формі.
6. Запишіть принципові формули для визначення параметрів множинної регресії МНК у матричній формі.
7. Сформулюйте теорему Гаусса-Маркова для множинної регресії.
8. Доведіть незміщеність оцінок параметрів множинної регресії.
9. Дайте визначення стандартизованих коефіцієнтів регресії та коефіцієнтів еластичності. Поясніть їх суть.
10. Запишіть коваріаційну матрицю оцінок параметрів регресійної моделі.
11. Запишіть у матричній формі формули для визначення стандартної похибки оцінок коефіцієнтів множинної регресії.
12. Сформулюйте порядок перевірки значущості оцінок параметрів множинної регресії.
13. Як визначаються довірчі інтервали для параметрів множинної регресії?
14. Запишіть формули для визначення коефіцієнта множинної кореляції. Які підходи існують до його визначення?
15. Проведіть у матричній формі розрахунки, необхідні для визначення коефіцієнта множинної детермінації.

16. Запишіть модель дисперсійного аналізу для багатофакторної регресії.

17. Запишіть формули для визначення скоригованого коефіцієнта детермінації. Поясніть його властивості.

18. Сформулюйте правила перевірки значущості множинного рівняння регресії за *F*-критерієм.

19. Поясніть порядок перевірки значущості коефіцієнта множинної кореляції.

20. Як визначається точковий прогноз за множинною регресійною моделлю?

21. Як визначається інтервальний прогноз для математичного сподівання залежної змінної?

22. Як визначається інтервальний прогноз для індивідуальних значень?

23. Сформулюйте процедуру покрокового відбору найбільш інформативних змінних та вилучення змінних.

24. Оцініть параметри економетричної моделі, яка характеризує залежність між тижневими витратами на харчування, загальними витратами та розміром сім'ї. Вихідні дані наведено в таблиці:

№ з/п	Витрати на харчування Y, гр.од.	Загальні витрати X_1, гр.од.	Розмір середньої сім'ї X_2, к-ть членів
1	22	45	1,5
2	34	75	1,6
3	50	125	1,9
4	67	223	1,8
5	47	92	3,4
6	66	146	3,6
7	81	227	3,4
8	106	358	3,5
9	70	135	5,5
10	95	218	5,4
11	119	331	5,4
12	147	490	5,3
13	93	175	8,5
14	133	305	8,3
15	169	468	8,1
16	197	749	7,3

25. Визначте кількісну залежність між прибутком фірми й основними видами ресурсів, які вона вкладає у свою господарську діяльність:

а) інвестиції;

б) основні виробничі фонди;

в) фонд робочого часу.

Для побудови економетричної моделі використайте статистичну інформацію, наведену в таблиці:

Місяць	Прибуток ***Y***, гр.од.	Інвестиції X_1, гр.од.	ОВФ X_2, гр.од	ФРЧ X_3, людино-дні
1-й	39	62	22	104
2-й	41	65	25	109
3-й	38	57	17	99
4-й	42	66	27	114
5-й	44	69	28	116
6-й	49	58	20	110
7-й	44	72	32	119
8-й	45	70	30	116
9-й	48	75	34	114
10-й	51	79	35	120
11-й	49	77	33	124
12-й	54	82	37	119
13-й	55	80	37	129
14-й	57	75	39	129
15-й	56	83	38	132
16-й	54	81	36	130
17-й	59	87	40	124
18-й	61	92	42	134
19-й	62	95	43	137
20-й	64	97	42	139

26. Побудуйте економетричну модель, що характеризує залежність між витратами обігу, обсягом вантажообороту та фондомісткістю бази.

Визначте стандартні помилки параметрів.

Дайте змістове тлумачення взаємозв’язку.

Вихідні дані наведені в табл. A–D.

Таблиця A

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,72	15,6	106,3
2	3,04	13,5	128,5
3	2,84	15,3	118,0
4	2,89	14,9	121,2
5	2,58	15,1	120,0
6	2,64	16,1	118,4
7	2,52	16,7	108,4
8	2,75	15,4	110,0
9	2,63	17,1	105,9

Таблиця B

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,58	15,1	120,0
2	2,64	16,1	118,4
3	2,52	16,7	108,4
4	2,75	15,4	110,0
5	2,63	17,1	105,9
6	2,48	16,8	117,7
7	2,62	16,9	97,5
8	2,88	16,1	113,7
9	2,68	15,0	122,3

Таблиця C

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,48	16,8	117,7
2	2,62	16,9	97,5
3	2,88	16,1	113,7
4	2,68	15,0	122,3
5	2,52	18,0	102,0
6	2,74	17,2	106,7
7	2,56	17,1	108,5
8	2,68	16,4	114,3
9	2,55	16,7	94,3

Таблиця D

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,75	16,8	110,0
2	2,63	15,5	105,9
3	2,48	17,0	117,7
4	2,62	16,8	97,5
5	2,88	16,9	113,7
6	2,68	16,1	122,3
7	2,56	15,0	102,0
8	2,74	18,0	106,7
9	2,60	17,2	108,5

27. Для моделі, яка побудована для даних, наведених у табл. E–H, виконати дисперсійний аналіз, зробити висновки щодо достовірності моделі та її параметрів.

Таблиця E

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,92	14,1	87,8
2	2,64	17,2	72,0
3	2,79	17,1	72,4
4	2,67	17,8	69,5
5	2,68	16,2	75,0
6	2,85	17,2	70,6
7	2,40	16,8	73,4
8	2,91	14,8	80,7
9	2,29	19,6	62,2

Таблиця F

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,45	17,1	71,3
2	2,38	19,5	61,7
3	3,04	12,5	96,2
4	2,67	16,5	72,9
5	2,70	16,0	75,0
6	2,65	16,1	74,6
7	2,79	16,2	74,1
8	2,49	18,0	66,9
9	3,27	11,4	98,6

Таблиця G

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,85	17,2	70,6
2	2,40	16,8	73,4
3	2,91	14,8	80,7
4	2,29	19,6	62,2
5	3,27	11,4	98,6
6	2,45	17,1	71,3
7	2,38	19,5	61,7
8	3,04	12,5	96,2
9	2,67	16,5	72,9

Таблиця H

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,93	18,1	71,2
2	2,50	17,2	73,4
3	2,95	14,9	81,2
4	2,39	20,1	63,7
5	3,25	11,4	96,6
6	2,65	17,1	72,2
7	2,42	19,5	61,7
8	3,14	17,5	96,2
9	2,75	16,7	72,9

28. За даними, які наведені в табл. I–J, побудуйте економетричну модель за методом 1МНК на основі покрокової регресії. Оцініть параметри даної економетричної моделі. Дайте змістове тлумачення оцінок параметрів, зробіть висновки.

Таблиця I

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,14	17,4	80,3
2	2,94	13,8	102,5
3	2,67	15,0	94,3
4	2,44	18,6	76,0
5	2,83	16,2	87,3
6	2,92	15,7	90,1
7	2,61	17,9	82,8
8	2,72	15,3	96,9
9	2,68	16,3	83,7

Таблиця J

№ з/п	Витрати	Вантажо-оборот	Фондо-місткість
1	2,67	15,0	94,0
2	2,45	18,6	78,0
3	2,86	16,2	87,5
4	2,90	15,7	90,2
5	2,60	17,9	84,8
6	2,72	16,3	95,9
7	2,68	17,7	91,0
8	2,50	16,8	84,7
9	2,74	17,5	88,2

ЛІТЕРАТУРА

1. Айвазян С.А. Прикладная статистика и основы эконометрики: учебник для вузов / С.А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022 с.

2. Доугерти К. Введение в эконометрику: пер. с англ. / К. Доугерти. – М.: ИНФРА-М, 1999. – XIV. – 402 с.

3. Єлейко В. Основи економетрії / В. Єлейко. – Львів: «Марка» ЛТД, 1995. – Ч. 1. – 192 с.

4. Козловський С.О. Економетрія: методичні вказівки, програма і контрольні завдання з курсу / С.О. Козловський, Л.Д. Загвойська. – Львів: УкрДЛТУ, 2003. – 32 с.

5. Кремер Н.Ш. Эконометрика: учебник для вузов / Н.Ш. Кремер, Б.А. Путко; под. ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ-ДАНА, 2003. – 311 с.

6. Лук'яненко І.Г. Економетрика: підручник / І.Г. Лук'яненко, Л.І. Краснікова. – К.: Товариство «Знання»; КОО, 1998. – 220 с.

7. Лук'яненко І.Г. Економетрика: практикум з використання комп'ютера / І.Г. Лук'яненко, Л.І. Краснікова. – К.: Товариство «Знання», КОО, 1998. – 220 с.

8. Мардас А.Н. Эконометрика: учебное пособие / А.Н. Мардас. – СПб.: Питер, 2001. – 144 с.

9. Наконечний С. Економетрія: підручник / С.І. Наконечний, Т.О. Терещенко, Т.П. Романюк. – Вид. 2-ге, доп. та перероб. – К.: КНЕУ, 2000. – 296 с.

10. Перепелицкий С.Н. Применение методов математической статистики в экономических расчетах и исследованиях: учебное пособие / С.Н. Перепелицкий. – М.: МЛТИ, 1987. – 96 с.

11. Перепелицкий С.Н. Экономико-математические методы и модели в планировании и управлении на предприятиях лесной промышленности: учебник для вузов / С.Н. Перепелицкий. – М.: Лесная промышленность, 1989. – 360 с.

12. Ржевський С.В. Вступ для економетрії: навч. посібник для студентів економічних спеціальностей / С.В. Ржевський. – К.: Вид-во Європ. ун-ту фінансів, інформ. систем, менедж. і бізнесу, 1999. – 93 с.

13. Толбатов Ю.А. Економетрика: підручник для студентів економічних спеціальностей вищ. навч. закл. / Ю.А. Толбатов. – К.: Четверта хвиля, 1997. – 320 с.

14. Dougherty Cristopher. Introduction to Econometrics / Dougherty Cristopher. – New York. – Oxford University Press, 1992.

15. Green William M. Econometric Analysis. Fourth Edition / Green William M. – New York: Prentice Hall International, 2000.

16. Gujarati D.N. Basic Econometrics. Third Edition / D.N. Gujarati. – New York: McGraw Hill, 1995.

17. Pindyk Robert S. Econometric models and Economic Forecasts / Pindyk Robert S., Rubinfeld Daniel L. – Fourth Edition. – New York: McGraw Hill, 1998.

ДОДАТКИ

Додаток 1

Значення функції Лапласа

	0	1	2	13	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029

	0	1	2	13	4	5	6	7	8	9
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	9915	9917	9920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972

	0	1	2	13	4	5	6	7	8	9
3,0	0,9973	0,9974	0,9975	0,9976	0,9977	0,9978	0,9979	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,2	9986	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,3	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,4	9993	9994	9994	9994	9994	9994	9995	9995	9995	9995
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9997	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	-	-	-	-	-	-	-	-	-
5,0	0,99999994	-	-	-	-	-	-	-	-	-

Додаток 2

Критичні точки розподілу Стьюдента

Кількість ступенів вільності	Рівень значущості α (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,29	636,58
2	2,92	4,30	6,96	9,92	22,33	31,60
3	2,35	3,18	4,54	5,84	1021	12,92
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,02	2,57	3,36	4,03	5,89	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,41
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,02	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,97
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,72	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,50	3,79
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	249	2,79	3,45	3,73
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	202	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Рівень значущості α (одностороння критична область)					

Додаток 3

Критичні точки F-розподілу Фішера – Снедекора

Рівень значущості $\alpha = 0{,}01$												
$k_1 \backslash k_2$	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,41	99,42
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84
31	7,53	5,36	4,48	3,99	3,67	3,45	3,28	3,15	3,04	2,96	2,88	2,82
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,86	2,80
33	7,47	5,31	4,44	3,95	3,63	3,41	3,24	3,11	3,00	2,91	2,84	2,78
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,82	2,76
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,80	2,74
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,79	2,72
37	7,37	5,23	4,36	3,87	3,56	3,33	3,17	3,04	2,93	2,84	2,77	2,71
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,75	2,69
39	7,33	5,19	4,33	3,84	3,53	3,30	3,14	3,01	2,90	2,81	2,74	2,68
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66
41	7,30	5,16	4,30	3,81	3,50	3,28	3,11	2,98	2,87	2,79	2,71	2,65
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,70	2,64
43	7,26	5,14	4,27	3,79	3,48	3,25	3,09	2,96	2,85	2,76	2,69	2,63
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62

Рівень значущості $\alpha = 0{,}01$												
k_1 \ k_2	1	2	3	4	5	6	7	8	9	10	11	12
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74	2,67	2,61
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,66	2,60
47	7,21	5,09	4,23	3,75	3,43	3,21	3,05	2,92	2,81	2,72	2,65	2,59
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,71	2,64	2,58
49	7,18	5,07	4,21	3,73	3,42	3,19	3,03	2,90	2,79	2,71	2,63	2,57
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63	2,56
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69	2,61	3,53	2,47
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45
75	6,99	4,90	4,05	3,58	3,27	3,05	2,89	2,76	2,65	2,57	2,49	2,43
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42
85	6,94	4,86	4,02	3,55	3,24	3,02	2,86	2,73	2,62	2,54	2,46	2,40
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39
95	6,91	4,84	3,99	3,52	3,22	3,00	2,83	2,70	2,60	2,51	2,44	2,38
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2.55	2,47	2,39	2,33
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31
175	6,78	4,73	3,90	3,43	3,12	2,91	2,74	2,61	2,51	2,42	2,35	2,29
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,31	2,24
400	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	245	2,37	2,29	2,23
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22
700	6,67	4,64	3,81	3,35	3,04	2,83	2,66	2,54	2,43	2,35	2,27	2,21
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2.20
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18

Рівень значущості α = 0,01												
k_1 \ k_2	14	16	20	24	30	40	50	75	100	200	500	∞
1	6143	6170	6209	6234	6260	6286	6302	6324	6334	6350	6360	6366
2	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,48	99,49	99,49	99,50	99,50
3	26,92	26,83	26,69	26,60	26,50	26,41	26,35	26,28	26,24	26,18	26,15	26,13
4	14,25	14,15	14,02	13,93	13,84	13,75	13,69	13,61	13,58	13,52	13,49	13,46
5	9,77	9,68	9,55	9,47	9,38	9,24	9,29	9,17	9,13	9,08	9,04	9,02
6	7,60	7,52	7,40	7,31	7,23	7,14	7,09	7,02	6,99	6,93	6,90	6,88
7	6,36	6,28	6,16	6,07	5,99	5,91	5,86	5,79	5,75	5,70	5,67	5,65
8	5,56	5,48	5,36	5,28	5,20	5,12	5,07	5,00	4,96	4,91	4,88	4,86
9	5,01	4,92	4,81	4,73	4,65	4,57	4,52	4,45	4,41	4,36	4,33	4,31
10	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	4,29	4,21	4,10	4,02	3,94	3,86	3,81	3,74	3,71	3,66	3,62	3,60
12	4,05	3,97	3,86	3,78	3,70	3,62	3,57	3,50	3,47	3,41	3,38	3,36
13	3,86	3,78	3,66	3,59	3,51	3,43	3,38	3,31	3,27	3,22	3,19	3,17
14	3,70	3,62	3,51	3,43	3,35	3,27	3,22	3,14	3,11	3,06	3,03	3,00
15	3,56	3,49	3,37	3,29	3,21	3,13	3,08	3,01	2,98	2,92	2,89	2,87
16	3,45	3,37	3,26	3,18	3,10	3,02	2,97	2,90	2,86	2,81	2,78	2,75
17	3,35	3,27	3,16	3,08	3,00	2,92	2,87	2,80	2,76	2,71	2,68	2,65
18	3,27	3,19	3,08	3,00	2,92	2,84	2,78	2,71	2,68	2,62	2,59	2,57
19	3,19	3,12	3,00	2,92	2,84	2,76	2,71	2,64	2,60	2,55	2,51	2,49
20	3,13	3,05	2,94	2,86	2,78	2,69	2,64	2,57	2,54	2,48	2,44	2,42
21	3,07	2,99	2,88	2,80	2,72	2,64	2,58	2,51	2,48	2,42	2,38	2,36
22	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,36	2,33	2,31
23	2,97	2,89	2,78	2,70	2,62	2,54	2,48	2,41	2,37	2,32	2,28	2,26
24	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,37	2,33	2,27	2,24	2,21
25	2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,33	2,29	2,23	2,19	2,17
26	2,86	2,78	2,66	2,58	2,50	2,42	2,36	2,29	2,25	2,19	2,16	2,13
27	2,82	2,75	2,63	2,55	2,47	2,38	2,33	2,26	2,22	2,16	2,12	2,10
28	2,79	2,72	2,60	2,52	2,44	2,35	2,30	2,23	2,19	2,13	2,09	2,06
29	2,77	2,69	2,57	2,49	2,41	2,33	2,27	2,20	2,16	2,10	2,06	2,03
30	2,74	2,66	2,55	2,47	2,39	2,30	2,25	2,17	2,13	2,07	2,03	2,01
31	2,72	2,64	2,52	2,45	2,36	2,27	2,22	2,14	2,11	2,04	2,01	1,98
32	2,70	2,62	2,50	2,42	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96
33	2,68	2,60	2,48	2,40	2,32	2,23	2,18	2,10	2,06	2,00	1,96	1,93
34	2,66	2,58	2,46	2,38	2,30	2,21	2,16	2,08	2,04	1,98	1,94	1,91
35	2,64	2,56	2,44	2,36	2,28	2,19	2,14	2,06	2,02	1,96	1,92	1,89
36	2,62	2,54	2,43	2,35	2,26	2,18	2,12	2,04	2,00	1,94	1,90	1,87
37	2,61	2,53	2,41	2,33	2,25	2,16	2,10	2,03	1,98	1,92	1,88	1,85
38	2,59	2,51	2,40	2,32	2,23	2,14	2,09	2,01	1,97	1,90	1,86	1,84
39	2,58	2,50	2,38	2,30	2,22	2,13	2,07	1,99	1,95	1,89	1,85	1,82
40	2,56	2,48	2,37	2,29	2,20	2,11	2,06	1,98	1,94	1,87	1,83	1,80
41	2,55	2,47	2,36	2,28	2,19	2,10	2,04	1,97	1,92	1,86	1,82	1,79
42	2,54	2,46	2,34	2,26	2,18	2,09	2,03	1,95	1,91	1,85	1,80	1,78
43	2,53	2,45	2,33	2,25	2,17	2,08	2,02	1,94	1,90	1,83	1,79	1,76
44	2,52	2,44	2,32	2,24	2,15	2,07	2,01	1,93	1,89	1,82	1,78	1,75
45	2,51	2,43	2,31	2,23	2,14	2,05	2,00	1,92	1,88	1,81	1,77	1,74
46	2,50	2,42	2,30	2,22	2,13	2,04	1,99	1,91	1,86	1,80	1,76	1,73
47	2,49	2,41	2,29	2,21	2,12	2,03	1,98	1,90	1,85	1,79	1,74	1,71
48	2,48	2,40	2,28	2,20	2,12	2,02	1,97	1,89	1,84	1,78	1,73	1,70
49	2,47	2,39	2,27	2,19	2,11	2,02	1,96	1,88	1,83	1,77	1,72	1,69

Рівень значущості α = 0,01												
k_1 \ k_2	14	16	20	24	30	40	50	75	100	200	500	∞
50	2,46	2,38	2,27	2,18	2,10	2,01	1,95	1,87	1,82	1,76	1,71	1,68
55	2,42	2,34	2,23	2,15	2,06	1,97	1,91	1,83	1,78	1,71	1,67	1,64
60	2,39	2,31	2,20	2,12	2,03	1,94	1,88	1,79	1,75	1,68	1,63	1,60
65	2,37	2,29	2,17	2,09	2,00	1,91	1,85	1,77	1,72	1,65	1,60	1,57
70	2,35	2,27	2,15	2,07	1,98	1,89	1,83	1,74	1,70	1,62	1,57	1,54
75	2,33	2,25	2,13	2,05	1,96	1,87	1,81	1,72	1,67	1,60	1,55	1,52
80	2,31	2,23	2,12	2,03	1,94	1,85	1,79	1,70	1,65	1,58	1,53	1,49
85	2,30	2,22	2,10	2,02	1,93	1,83	1,77	7,69	1,64	1,56	1,51	1,47
90	2,29	2,21	2,09	2,00	1,92	1,82	1,76	1,67	1,62	1,55	1,49	1,46
95	2,28	2,20	2,08	1,99	1,90	1,81	1,75	1,66	1,61	1,53	1,48	1,44
100	2,27	2,19	2,07	1,98	1,89	1,80	1,74	1,65	1,60	1,52	1,47	1,43
125	2,23	2,15	2,03	1,94	1,85	1,76	1,69	1,60	1,55	1,47	1,41	1,37
150	2,20	2,12	2,00	1,92	1,83	1,73	1,66	1,57	1,52	1,43	1,38	1,33
175	2,19	2,10	1,98	1,90	1,81	1,71	1,64	1,55	1,50	1,41	1,35	1,30
200	2,17	2,09	1,97	1,89	1,79	1,69	1,63	1,53	1,48	1,39	1,33	1,28
300	2,14	2,06	1,94	1,85	1,76	1,66	1,59	1,50	1,44	1,35	1,28	1,22
400	2,13	2,05	1,92	1,84	1,75	1,64	1,58	1,48	1,42	1,32	1,25	1,19
500	2,12	2,04	1,92	1,83	1,74	1,63	1,57	1,47	1,41	1,31	1,23	1,16
600	2,11	2,03	1,91	1,82	1,73	1,63	1,56	1,46	1,40	1,30	1,22	1,15
700	2,11	2,03	1,90	1,82	1,72	1,62	1,55	1,45	1,39	1,29	1,21	1,14
800	2,10	2,02	1,90	1,81	1,72	1,62	1,55	1,45	1,39	1,29	1,20	1,13
1000	2,10	2,02	1,90	1,81	1,72	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	2,08	2,00	1,88	1,79	1,70	1,59	1,52	1,42	1,36	1,25	1,15	1,00

Рівень значущості α = 0,05

k_1 \ k_2	1	2	3	4	5	6	7	8	9	10	11	12
1	161	199	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38
18	4,41	3,55	316	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09
31	4,16	3,30	2,91	2,68	2,52	2,41	2,32	2,25	2,20	2,15	2,11	2,08
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07
33	4,14	3,28	2,89	2,66	2,50	2,39	2,30	2,23	2,18	2,13	2,09	2,06
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03
37	4,11	3,25	2,86	2,63	2,47	2,36	2,27	2,20	2,14	2,10	2,06	2,02
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02
39	4,09	3,24	2,85	2,61	2,46	2,34	2,26	2,19	2,13	2,08	2,04	2,01
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00
41	4,08	3,23	2,83	2,60	2,44	2,33	2,24	2,17	2,12	2,07	2,03	2,00
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99
43	4,07	3,21	2,82	2,59	2,43	2,32	2,23	2,16	2,11	2,06	2,02	1,99
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97
47	4,05	3,20	2,80	2,57	2,41	2,30	2,21	2,14	2,09	2,04	2,00	1,96
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96
49	4,04	3,19	2,79	2,56	2,40	2,29	2,20	2,13	2,08	2,03	1,99	1,96

Рівень значущості α = 0,05												
k_1 \ k_2	1	2	3	4	5	6	7	8	9	10	11	12
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,97	1,93
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,94	1,90
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,92	1,88
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88
85	3,95	3,10	2,71	2,48	2,32	2,21	2,12	2,05	1,99	1,94	1,90	1,87
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,98	1,93	1,89	1,86
95	3,94	3,09	2,70	2,47	2,31	2,20	2,11	2,04	1,98	1,93	1,89	1,86
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82
175	3,90	3,05	2,66	2,42	2,27	2,15	2,06	1,99	1,93	1,89	1,84	1,81
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,81	1,78
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85	1,80	1,77
700	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,77
800	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,76
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75

Рівень значущості $\alpha = 0{,}05$												
k_2 / k_1	14	16	20	24	30	40	50	75	100	200	500	∞
1	245	246	248	249	250	251	252	253	253	254	254	254
2	19,42	19,43	19,45	19,45	19,46	19,47	19,48	19,48	19,49	19,49	19,49	19,50
3	8,71	8,69	8,66	8,64	8,62	8,59	8,58	8,56	8,55	8,54	8,53	8,53
4	5,87	5,84	5,80	5,77	5,75	5,72	5,70	5,68	5,66	5,65	5,64	5,63
5	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,41	4,39	4,37	4,37
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,73	3,71	3,69	3,68	3,67
7	3,53	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,27	3,25	3,24	3,23
8	3,24	3,20	3,15	3,12	3,08	3,04	3,02	2,99	2,97	2,95	2,94	2,93
9	3,03	2,99	2,94	2,90	2,86	2,83	2,80	2,77	2,76	2,73	2,72	2,71
10	2,86	2,83	2,77	2,74	2,70	2,66	2,64	2,60	2,59	2,56	2,55	2,54
11	2,74	2,70	2,65	2,61	2,57	2,53	2,51	2,47	2,46	2,43	2,42	2,40
12	2,64	2,60	2,54	2,51	2,47	2,43	2,40	2,37	2,35	2,32	2,31	2,30
13	2,55	2,51	2,46	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,22	2,21
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
15	2,42	2,38	2,33	2,29	2,25	2,20	2,18	2,14	2,12	2,10	2,08	2,07
16	2,37	2,33	2,28	2,24	2,19	2,15	2,12	2,09	2,07	2,04	2,02	2,01
17	2,33	2,29	2,23	2,19	2,15	2,10	2,08	2,04	2,02	1,99	1,97	1,96
18	2,29	2,25	2,19	2,15	2,11	2,06	2,04	2,00	1,98	1,95	1,93	1,92
19	2,26	2,21	2,16	2,11	2,07	2,03	2,00	1,96	1,94	1,91	1,89	1,88
20	2,22	2,18	2,12	2,08	2,04	1,99	1,97	1,93	1,91	1,88	1,86	1,84
21	2,20	2,16	2,10	2,05	2,01	1,96	1,94	1,90	1,88	1,84	1,83	1,81
22	2,17	2,13	2,07	2,03	1,98	1,94	1,91	1,87	1,85	1,82	1,80	1,78
23	2,15	2,11	2,05	2,01	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
24	2,13	2,09	2,03	1,98	1,94	1,89	1,86	1,82	1,80	1,77	1,75	1,73
25	2,11	2,07	2,01	1,96	1,92	1,87	1,84	1,80	1,78	1,75	1,73	1,71
26	2,09	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,73	1,71	1,69
27	2,08	2,04	1,97	1,93	1,88	1,84	1,81	1,76	1,74	1,71	1,69	1,67
28	2,06	2,02	1,96	1,91	1,87	1,82	1,79	1,75	1,73	1,69	1,67	1,65
29	2,05	2,01	1,94	1,90	1,85	1,81	1,77	1,73	1,71	1,67	1,65	1,64
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,70	1,66	1,64	1,62
31	2,03	1,98	1,92	1,88	1,83	1,78	1,75	1,70	1,68	1,65	1,62	1,61
32	2,01	1,97	1,91	1,86	1,82	1,77	1,74	1,69	1,67	1,63	1,61	1,59
33	2,00	1,96	1,90	1,85	1,81	1,76	1,72	1,68	1,66	1,62	1,60	1,58
34	1,99	1,95	1,89	1,84	1,80	1,75	1,71	1,67	1,65	1,61	1,59	1,57
35	1,99	1,94	1,88	1,83	1,79	1,74	1,70	1,66	1,63	1,60	1,57	1,56
36	1,98	1,93	1,87	1,82	1,78	1,73	1,69	1,65	1,62	1,59	1,56	1,55
37	1,97	1,93	1,86	1,82	1,77	1,72	1,68	1,64	1,62	1,58	1,55	1,54
38	1,96	1,92	1,85	1,81	1,76	1,71	1,68	1,63	1,61	1,57	1,54	1,53
39	1,95	1,91	1,85	1,80	1,75	1,70	1,67	1,62	1,60	1,56	1,53	1,52
40	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
41	1,94	1,90	1,83	1,79	1,74	1,69	1,65	1,61	1,58	1,54	1,52	1,50
42	1,94	1,89	1,83	1,78	1,73	1,68	1,65	1,60	1,57	1,53	1,51	1,49
43	1,93	1,89	1,82	1,77	1,72	1,67	1,64	1,59	1,57	1,53	1,50	1,48
44	1,92	1,88	1,81	1,77	1,72	1,67	1,63	1,59	1,56	1,52	1,49	1,48
45	1,92	1,87	1,81	1,76	1,71	1,66	1,63	1,58	1,55	1,51	1,49	1,47
46	1,91	1,87	1,80	1,76	1,71	1,65	1,62	1,57	1,55	1,51	1,48	1,46
47	1,91	1,86	1,80	1,75	1,70	1,65	1,61	1,57	1,54	1,50	1,47	1,46
48	1,90	1,86	1,79	1,75	1,70	1,64	1,61	1,56	1,54	1,49	1,47	1,45
49	1,90	1,85	1,79	1,74	1,69	1,64	1,60	1,56	1,53	1,49	1,46	1,44
50	1,89	1,85	1,78	1,74	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44

Рівень значущості α = 0,05												
k_2 / k_1	14	16	20	24	30	40	50	75	100	200	500	∞
55	1,88	1,83	1,76	1,72	1,67	1,61	1,58	1,53	1,50	1,46	1,43	1,41
60	1,86	1,82	1,75	1,70	1,65	1,59	1,56	1,51	1,48	1,44	1,41	1,39
65	1,85	1,80	1,73	1,69	1,63	1,58	1,54	1,49	1,46	1,42	1,39	1,37
70	1,84	1,79	1,72	1,67	1,62	1,57	1,53	1,48	1,45	1,40	1,37	1,35
75	1,83	1,78	1,71	1,66	1,61	1,55	1,52	1,47	1,44	1,39	1,36	1,34
80	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,45	1,43	1,38	1,35	1,32
85	1,81	1,76	1,70	1,65	1,59	1,54	1,50	1,45	1,42	1,37	1,34	1,31
90	1,80	1,75	1,68	1,64	1,59	1,53	1,49	1,44	1,41	1,36	1,33	1,30
95	1,80	1,75	1,68	1,63	1,58	1,52	1,48	1,43	1,40	1,35	1,32	1,29
100	1,79	1,75	1,68	1,63	1,57	1,52	1,48	1,42	1,39	1,34	1,31	1,28
125	1,77	1,73	1,66	1,60	1,55	1,49	1,45	1,40	1,36	1,31	1,27	1,25
150	1,76	1,71	1,64	1,59	1,54	1,48	1,4	1,38	1,34	1,29	1,25	1,22
175	1,75	1,70	1,63	1,58	1,52	1,46	1,42	1,36	1,33	1,27	1,23	1,20
200	1,74	1,69	1,62	1,57	1,52	1,46	1,41	1,35	1,32	1,26	1,22	1,19
300	1,72	1,68	1,61	1,5	1,50	1,43	1,39	1,33	1,30	1,23	1,19	1,15
400	1,72	1,67	1,60	1,54	1,49	1,42	1,38	1,32	1,28	1,22	1,17	1,13
500	1,71	1,6	1,59	1,54	1,48	1,42	1,38	1,31	1,28	1,21	1,16	1,11
600	1,71	1,6	1,59	1,54	1,48	1,41	1,37	1,31	1,27	1,20	1,15	1,10
700	1,71	1,66	1,59	1,53	1,48	1,41	1,37	1,30	1,27	1,20	1,15	1,09
800	1,70	1,6	1,58	1,53	1,47	1,41	1,37	1,30	1,26	1,20	1,14	1,09
1000	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08
∞	1,69	1,64	1,57	1,52	1,46	1,39	1,35	1,28	1,24	1,17	1,11	1,00

НАВЧАЛЬНЕ ВИДАННЯ

ВОЛОШИН Олександра Романівна,
кандидат фізико-математичних наук,
ГАЛАЙКО Наталія Володимирівна

ЕКОНОМЕТРІЯ

ЧАСТИНА 1

Навчальний посібник

Редактор *Н.Ю. Шайнога*

Комп'ютерна верстка *Н.М. Лесь*

Друк *Н.Я. Ганущак*

Здано до набору 01.06.2012 р. Підписано до друку 20.09.2012 р.
Формат $60\times84/_{16}$. Папір офсетний. Гарнітура Times New Romans.
Віддруковано на різографі. Умов. друк. арк. 11,16.
Наклад 100 прим. Зам. № 89-12.

Львівський державний університет внутрішніх справ
Україна, 79007, м. Львів, вул. Городоцька, 26.

Свідоцтво про внесення суб'єкта видавничої справи до державного реєстру
видавців, виготівників і розповсюджувачів видавничої продукції.
№ 2541 від 26 червня 2006 р.