МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Донецький національний університет економіки і торгівлі
імені Михайла Туган-Барановського

Кафедра вищої математики і інформаційних систем

В.М. Серебреников, Т.В. Квітка, О.К. Копайгора

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОДЕЛІ: ЕКОНОМЕТРИКА

Навчально-методичний посібник для вивчення дисципліни

розділ «Кореляційно-регресійний аналіз»

Кривий Ріг

2019р.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Донецький національний університет економіки і торгівлі
імені Михайла Туган-Барановського

Кафедра вищої математики і інформаційних систем

В.М. Серебреников, Т.В. Квітка, О.К. Копайгора

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОДЕЛІ: ЕКОНОМЕТРИКА

Навчально-методичний посібник для вивчення дисципліни

розділ «Кореляційно-регресійний аналіз»

Затверджено на засіданні
кафедри вищої математики і
інформаційних систем
Протокол № 15
від “08” квітня 2019 р.

Схвалено навчально-методичною радою
ДонНУЕТ ім. М. Туган-Барановського
Протокол № 5
від «13» червня 2019 р.

Кривий Ріг

2019р

УДК 591.21 (075.8)

Рекомендовано до видання науково-методичною радою Донецького національного університету економіки і торгівлі ім. Михайла Туган-Барановського (протокол № 5 від 13 червня 2019р.)

Рецензенти:

Губін Г.В., д.т.н., професор, завідувач кафедри ливарного виробництва і металургії чорних металів Криворізького національного університету.

Тернов С.О., к.т.н., с.н.с., в.о. звідувача кафедри вищої математики та інформаційних систем Донецького національного університету економіки і торгівлі ім. Михайла Туган-Барановського.

Серебреников В.М., Квітка Т.В., Копайгора О.К.
Економіко-математичні методи та моделі: економетрика. Розділ «Кореляційно-регресійний аналіз» [Текст]: навч.-метод. посібник та завдання для виконання самостійної роботи студентів денної та заочної форм навчання / В.М.Серебреников, Т.В. Квітка, О.К.Копайгора; М-во освіти і науки України, Донец. нац. ун-т економіки і торгівлі ім. М.Туган-Барановського, каф. вищої мат. та інформ. систем. – Кривий Ріг: ДонНУЕТ, 2019. – 103 с.

Навчально-методичний посібник містить короткі теоретичні відомості з кожної теми, питання, що актуалізують знання, отримані під час лекцій, приклади розв’язування задач, 30 варіантів комплексних індивідуальних завдань. Рекомендується для використання у процесі навчання математики для економістів: економетрика, при вивченні змістового модуля «Кореляційно-регресійний аналіз». Наведено список рекомендованої літератури.

УДК 591.21 (075.8)

З М І С Т

ВСТУП

Навчально-методичний посібник пропонується для самостійної підготовки студентів до практичних занять з навчальної дисципліни «Економіко-математичні методи та моделі: економетрика».

Передбачається, що самостійна робота з використанням розробленого навчально-методичного посібника допоможе якісно підготуватись до практичних занять та модульних контрольних робіт, розвиватиме вміння та навички самостійної діяльності (планування організації самостійної роботи та самоконтролю навчальних досягнень), формуватиме культуру розумової праці, стимулювати до плідної реалізації стратегії Європейського союзу – «навчання протягом усього життя», що сьогоденні постає необхідним підґрунтям творчої професійній діяльності, самореалізації особистості.

Наразі надається для самостійного опрацювання матеріал до практичних занять першого модуля («Кореляційно-регресійний аналіз») відповідно робочої програми навчальної дисципліни «Економіко-математичні методи та моделі: економетрика».

В результаті вивчення дисципліни студент повинен

знати:

- означення економетрики як розділу економічної науки,
- означення економетричної моделі,
- етапи економетричного дослідження,
- поняття кореляційної залежності,
- поняття регресійного рівняння,
- основні математичні поняття, які використовуються в кореляційно-регресійному аналізі,
- лінійний і нелінійний парний кореляційно-регресійний аналіз,
- лінійний і нелінійний багатофакторний кореляційно-регресійний аналіз,
- мультиколінеарність,
- гетероскадестичність,
- автокореляцію,

вміти:

- проводити парний і багатофакторний кореляційно-регресійний аналіз,
- використовувати табличний процесор MS Excel,
- використовувати пакет «АНАЛИЗ ДАННЫХ»,

мати навички:

- – самостійної роботи при розширенні своїх математичних знань та освоєння довідкових систем.

Порядок поточного оцінювання знань студентів з дисципліни

Успішне виконання студентом завдань поточного контролю є обов’язковою умовою участі його в складанні екзамену. Об’єктом поточного контролю знань студента є :

- контроль систематичності та активності роботи протягом семестру протягом семестру над вивченням програмного матеріалу дисципліни,
- виконання завдань для самостійного опрацювання,
- контроль за виконанням типових розрахунків.

Результати виконаної роботи подаються студентами в окремому зошиті. Завдання з типових розрахунків виконуються студентом в позааудиторний час протягом вивчення дисципліни і здаються в указаний викладачем термін. До захисту типового розрахунку студент допускається при умові правильного виконання всіх завдань.

Навчально-методичний посібник буде корисним студентам усіх форм навчання при підготовці до аудиторних занять, у процесі самоосвіти та при підготовці до контрольних робіт та іспиту з навчальної дисципліни.

ЧАСТИНА 1.

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ З ВИВЧЕННЯ ДИСЦИПЛІНИ

1. Опис дисципліни

Найменування показників	Характеристика дисципліни
Обов'язкова / вибіркова дисципліна	Обов’язкова
Семестр	4
Кількість кредитів	5
Загальна кількість годин	150
Кількість модулів	1
Лекції, годин	30
Практичні/ семінарські, годин	45
Лабораторні, годин	
Самостійна робота, годин	75
Тижневих годин для денної форми навчання:	
аудиторних	5
самостійної роботи студента	5
Вид контролю	екзамен

2. Мета та завдання дисципліни

Мета – формування у майбутніх спеціалістів базових математичних знань для розв’язування задач у професійній діяльності, вмінь аналітичного мислення та математичного формулювання виробничих задач.

Завдання – надання студентам знань із основних розділів економіко-математичного моделювання: означень, теорем, правил; доведення основних теорем; формування початкових умінь самостійно поглиблювати свої знання, розвивати логічне мислення; виробити вміння формулювати свої знання, розв’язувати прикладні задачі і будувати економіко-математичні моделі.

3. Структура дисципліни

Назви змістових модулів і тем	Кількість годин (денна форма навчання)				
	усього	у тому числі			
		лекц.	пр./сем.	лаб.	СРС
1	2	3	4	5	6
Змістовий модуль 1. Класичні моделі регресії. Деякі аспекти багатофакторної регресії.					
Тема 1. Лінійна та нелінійні моделі регресії.	10	2	3		5
Тема 2. Мультиколінеарність в багатофакторних моделях. Тест Фаррара-Глобера.	10	2	3		5
Тема 3. Багатофакторна лінійна регресія. Покроковий регресійний аналіз.	10	2	3		5
Тема 4. Конкретні застосування багато - факторної регресії.	10	2	3		5
Тема 5. Фіктивні змінні в регресійних моделях.	10	2	3		5
Тема 6. Автокореляція залишків. Критерій Дарбіна - Уотсона.	10	2	3		5
Тема 7. Гетероскедастичність. Тест Гольдфельда-Квандта. Тест Спірмена.	10	2	3		5
Разом за змістовим модулем 1	70	14	21	-	35
Змістовий модуль 2. Узагальнена лінійна модель регресії. Аналіз часових рядів і прогнозування					
Тема 8. Узагальнений метод найменших квадратів.	10	2	3		5
Тема 9. Лінійна регресійна модель в умовах гетероскедастичності.	10	2	3		5
Тема 10. Лінійна регресійна модель в умовах автокореляції.	10	2	3		5

Тема 11. Поняття часових рядів. ковзні середні. Лаг.	10	2	3		5
Тема 12. Аналіз часових рядів за допомогою автокореляційної функції.	10	2	3		5
Тема 13. Аналітичне вирівнювання. Криві росту.	10	2	3		5
Тема 14 Адаптивні моделі прогнозування.	10	2	3		5
Тема 15. Статистичний аналіз і прогнозування сезонних коливань.	10	2	3		5
Разом за змістовим модулем 2	80	16	24	-	40
Усього годин	150	30	45	-	75

4. Теми семінарських/практичних/лабораторних занять

№ з/п	Вид та тема семінарського заняття	Кількість годин
1	Семінар з виконанням практичних задач Побудова лінійної параболічної та гіперболічної залежності. Перевірка на адекватність, прогнозування.	3
2	Семінар з виконанням практичних задач Перевірка на мультиколінеарність за тестами Фаррара-Глобера і Спірмена.	3
3	Семінар з виконанням практичних задач Побудова математичної моделі. Перевірка її на адекватність. Практичне застосування моделей.	3
4	Семінар з виконанням практичних задач Побудова багатофакторної моделі, перевірка її на адекватність.	3
5	Семінар з виконанням практичних задач Практичне застосування багатофакторних моделей	3
6	Семінар з виконанням практичних задач Перевірка автокореляції за критерієм Дарбіна-Уотсона.	3
7	Семінар з виконанням розрахункових задач Перевірка гетероскедастичності за тестом Гольдфельда – Квандта та критерієм Спірмена.	3
8	Семінар з виконанням розрахункових задач Узагальнений метод найменших квадратів.	3
9	Семінар з виконанням розрахункових задач Побудова математичної моделі в умовах	3

	гетероскедастичності.	
10	Семінар з виконанням розрахункових задач Побудова математичної моделі в умовах в умовах автокореляції.	3
11	Семінар запитань і відповідей. Вирівнювання часових рядів.	3
12	Семінар з виконанням практичних задач. Аналіз часових рядів з допомогою автокореляційної функції.	3
13	Семінар з виконанням практичних задач. Побудова кривих росту, оцінка моделей, прогноз за моделями.	3
14	Семінар з виконанням практичних задач. Побудова і прогноз за адаптивними моделями.	3
15	Семінар з виконанням практичних задач. Моделювання економічних процесів з сезонними коливаннями.	3
Всього		45

5. Індивідуальні завдання

1. Огляд періодичної і монографічної наукової літератури.
2. Підготовка рефератів, доповідей за обраною темою.
3. Підготовка тез доповідей з метою виступу на університетських, всеукраїнських та міжнародних семінарах та конференціях.

6. Обсяги, зміст та засоби діагностики самостійної роботи

Вид та тема семінарських занять	Кільк. годин	Зміст самостійної роботи	Засоби діагностики
1	2	3	4
Змістовий модуль 1. Класичні моделі регресії. Деякі аспекти багатофакторної регресії			
Семінар з виконанням практичних задач Тема 1.Побудова лінійної, гіперболічної залежності. Перевірка на адекватність, прогнозування.	5	Складання конспекту з використанням навчального посібника та джерел Internet, робота з пошуковими системами Інтернет.	Рукопис
Семінар з виконанням	5	Підготовка додаткового матеріалу до	

практичних задач. Тема 2. Мультиколінеарність. Тест Фаррара-Глобера.		лекції відповідно до заданого плану; аналіз фактів, викладених у друкованих джерелах інформації, з метою підготовки відповідей на запитання, які було поставлено під час аудиторних навчальних занять	Електронний звіт
Семінар з виконанням практичних задач. Тема 3. Багатофакторна лінійна регресія. Покроковий регресійний аналіз.	5	1. Підготовка додаткового матеріалу до лекції відповідно до заданого плану; аналіз фактів, викладених у друкованих джерелах інформації, з метою підготовки відповідей на запитання, які було поставлено під час аудиторних навчальних занять. 2.Виконання домашніх робіт; виконання вправ; підготовка до захисту ндивідуальних робіт.	1. Електронний звіт 2. ІЗС
Семінар з виконанням практичних задач. Тема 4. Конкретні застосування багатофакторної регресії.	5	1. Складання конспекту з використанням навчального посібника та джерел Internet, робота з пошуковими системами Інтернет. 2. Виконання домашніх робіт; виконання вправ;підготовка до захисту індивідуальних робіт.	1. Рукопис 2. ІЗС
Семінар з виконанням практичних задач. Тема 5. Фіктивні змінні в регресійних моделях.	5	Виконання домашніх робіт; виконання вправ;підготовка до захисту індивідуальних робіт.	ІЗС
Семінар з виконанням практичних задач. Тема 6. Автокореляція залишків.Критерій Дарбіна -Уотсона.	5	Виконання домашніх робіт; виконання вправ;підготовка до захисту індивідуальних робіт.	ІЗС
Семінар з виконанням практичних задач. Тема 7. Гетероскедастичність. Тест Гольдфельда – Квандта. Тест Спірмена.	5	Виконання домашніх робіт; виконання вправ; підготовка до захисту індивідуальних робіт.	ІЗС
Разом змістовий модуль 1	35		

1	2	3	4
Змістовий модуль 2. Узагальнена лінійна модель регресії. Аналіз часових рядів і прогнозування			
Семінар з виконанням практичних задач Тема 8. Узагальнений метод найменших квадратів.	5	Складання конспекту з використанням навчального посібника та джерел Internet, робота з пошуковими системами Інтернет.	Рукопис
Семінар з виконанням практичних задач Тема 9. Лінійна регресійна модель в умовах гетероскедастичності.	5	Підготовка додаткового матеріалу до лекції відповідно до заданого плану; аналіз фактів, викладених у друкованих джерелах інформації, з метою підготовки відповідей на запитання, які було поставлено під час аудиторних навчальних занять	Електронний звіт
Семінар з виконанням практичних задач Тема 10. Лінійна регресійна модель в умовах автокореляції.	5	Виконання домашніх робіт; виконання вправ; підготовка до захисту індивідуальних робіт.	ІЗС
Семінар з виконанням розрахункових задач Тема 11. Поняття часових рядів. Ковзні середні. Лаг.	5	Підготовка додаткового матеріалу до лекції відповідно до заданого плану; аналіз фактів, викладених у джерелах інформації, з метою підготовки відповідей на запитання, які було поставлено під час навчальних занять	Електронний звіт
Семінар з виконанням розрахункових задач Тема 12. Аналіз часових рядів за допомогою автокореляційної функції.	5	Виконання домашніх робіт; виконання вправ; підготовка до захисту індивідуальних робіт.	ІЗС
Семінар з виконанням розрахункових задач Тема 13. Аналітичне вирівнювання. Криві росту.	5	Виконання домашніх робіт; виконання вправ;підготовка до захисту індивідуальних робіт.	ІЗС
Семінар з виконанням розрахункових задач Тема 14. Адаптивні моделі прогнозування.	5	Виконання домашніх робіт; виконання вправ;підготовка до захисту індивідуальних робіт.	ІЗС

Семінар з виконанням розрахункових задач Тема 15. Статистичний аналіз і прогнозування Сезонних коливань.	5	Виконання домашніх робіт; виконання вправ;підготовка до захисту індивідуальних робіт.	ІЗС
Разом змістовий модуль 2	40		
Разом	75		

7. Результати навчання

1	Вміти аналізувати та формулювати постановки економетричних задач з використанням математичних та статистичних методів.
2	Використовувати у практичній діяльності набуті знання щодо застосування економетричних методів для дослідження професійних задач
3	Самостійно працювати з навчально-методичною літературою і використовувати необхідні програмні продукти для аналізу і розв'язування професійних задач.
4	Аналізувати, виділяти головне, проводити оцінки, робити висновки, обґрунтовувати висновки.
5	Виробляти алгоритми.

8. Форми навчання

Лекції, практичні заняття, самостійна робота (підготовка додаткового матеріалу до лекції, робота з пошуковими системами Інтернет, складання конспекту з використанням навчального посібника, виконання вправ, самостійно опрацювання додаткових питань за наведеним переліком літератури).

9. Методи оцінювання

Екзамен.

10. Розподіл балів, які отримують студенти

Відповідно до системи оцінювання знань студентів ДонНУЕТ рівень сформованності компетентностей студента оцінюються у випадку проведення екзамену: на протязі семестру (50 балів) та при проведенні підсумкового контролю – екзамену (50 балів).

Оцінювання студентів протягом семестру

№ теми практичного заняття	Вид роботи/бали					
	Тестові завдання	Ситуаційні завдання, задачі	Обговорення теоретичних питань теми	Індиві-дуальне завдання	ПМК	Сума балів
Змістовий модуль 1						
Тема 1	1				1	2
Тема 2	1				1	2
Тема 3	1		1	2	1	5
Тема 4	1		1	2	1	5
Тема 5	1			2	1	4
Тема 6	1			2	1	4
Тема 7				2	1	3
Разом змістовий модуль 1	6		2	10	7	25
Змістовий модуль 2						
Тема 8	1			1	1	3
Тема 9			1	1	1	3
Тема 10	1			1	1	3
Тема 11			1	1	1	3
Тема 12	1			1	1	3
Тема 13			1	1	1	3
Тема 14	1			1	1	3
Тема 15	1		1	1	1	4
Разом змістовий модуль 2	5		4	8	8	25
Разом	11		6	18	15	50

Оцінювання студентів при проведенні екзамену з використанням комп’ютерної програми “TestXPro”

Оцінка на підсумковому контролі складається з двох елементів:
0-40 балів - теоретична частина (тестування);
0-10 балів – практична частина (розрахункове завдання).

Набрані бали за виконання теоретичної та практичної частин додаються. Сума складає загальну кількість балів, отриманих за екзамен.

Теоретична частина екзамену включає тестові завдання з закритою та відкритою відповідю.

Оцінювання *тестових завдань* проводиться на основі інформації, яку надає комп'ютер за результатами тестування (кількість балів). Правильна відповідь на одне тестове завдання оцінюється в балах відповідно до типу завдання.

Загальне оцінювання *теоретичної частини* екзамену відбувається шляхом підбиття підсумку або сумування балів, які набрали студенти під час тестування.

Оцінювання результатів виконання **практичної частини** (розрахункове завдання на 10 балів) здійснюється відповідно до шкали оцінювання практичної частини.

Шкала оцінювання практичної частини

Сума балів	Критерії оцінювання
10	Завдання виконано у повному обсязі, відповідь обґрунтована, висновки і пропозиції аргументовані, розрахунки правильні, оформлення відповідає вимогам
7	Завдання виконано у повному обсязі, але допущено незначні неточності в розрахунках або оформленні, прийняті рішення недостатньо аргументовані
4	Завдання виконано не менше ніж на 70% при правильному оформлені або не менше ніж на 80%, якщо допущені незначні помилки в розрахунках чи оформленні
0-3	Завдання виконано менше ніж на 70%, допущені помилки в розрахунках чи оформленні, прийняте рішення не аргументовано

Загальне оцінювання результатів вивчення дисципліни

Для виставлення підсумкової оцінки визначається сума балів, отриманих за результатами екзамену та за результатами складання змістових модулів. Оцінювання здійснюється за допомогою шкали оцінювання загальних результатів вивчення дисципліни (модулю).

Оцінка		
100-бальна шкала	Шкала ECTS	Національна шкала
90-100	A	5, «відмінно»
80-89	B	4, «добре»
75-79	C	
70-74	D	3, «задовільно»
60-69	E	
59-30	FX	2, «незадовільно»
0-29	F	

11. Методичне забезпечення

1. Навчальний посібник.
2. Електронний конспект лекцій.
3. Методичні вказівки з вивчення дисципліни.
4. Комплекти індивідуальних завдань.
5. Навчальна та наукова література, нормативні документи.

12. Література

Базова

1. Васильченко І. П. Вища математика для економістів підручник / І. П. Васильченко – К. : Знання –Прес, 2002 .– 454с.
2. Высшая математика для экономистов: учебник для студ. вузов, обучающихся по экон. спец. / Н. Ш. Кремер и др., под ред. проф. Н. Ш. Кремера.– М.:ЮНИТИ, 2006. – 479 с.
3. Исследование операций в экономике: уч. пособ. для вузов Под ред. проф. Н. М. Кремера.– М.: ЮНИТИ,2006. – 407с.
4. Ковалев В. Г. Математическое программирование (линейные задачи) учеб. пособие / В. Г. Ковалев, А. Р. Наринян, В. А. Поздев – К.: Европ. ун– т, 2004. – 170 с.
5. Красс М. С. Математика в экономике. Математические методы и модели: учебник/ М. С. Красс, Б. П. Чупринов – М.: Финансы и статистика, 2007.– 544с.
6. Лавріненко Н. М. Основи економіко – математичного моделювання: навч. посіб./ Н. М. Лавріненко, С. М. Латинін, В. В. Фортуна, О. І. Бескровний – Л.: Магнолія, 2006, 2010.–540 с.
7. Орлова И.В.Экономико –математические методы и модели: компьютерное моделирование: Учеб пособие / И. В. Орлова, В. А. Половников – М.: Вузовский учебник, 2007.–365с.
8. Практикум по эконометрике: учеб пособие / И. И. Елисеева и др.; под ред И. И. Елисеевой.– 2-е изд. перераб. и доп. – М.: Финансы и статистика, 2006.– 192 с.
9. Таха Х. Введение в исследование операций / Х. Таха; пер. с англ. и редакция А. А. Минько. – М.: Вильямс, 2007. – 912 с.
10. Эконометрика: учебник / И. И. Елисеева и др. под ред И. И. Елисеевой.– 2- е изд. перераб. и доп. М.: Финансы и статистика, 2006. – 576с.

Допоміжна

11. Вітлінський В. В. Математичне програмування: навч.- метод. посібник для самост. вивч. дисц. /В. В. Вітлінський, С. І. Наконечний, Т. О. Терещенко.–К.:

КНЕУ, 2001. 248 с.
12. Конюховский П. В. Математические методы исследования операций / П. В. Конюховский. – СПб Питер: 2001.–192с.
13. Корольов О. А. Економетрія : навч. посіб. / О. А. Корольов.– К.: КНЕУ, 2000. – 660с.
14. Кулинич Е. И. Эконометрия / Е. И. Кулинич.– М.: Финансы и статистика, 2001.– 304с.
15. Наконечний С. І. Економетрія: підручник /С. І. Наконечний ,Т. О. Терещенко, Т. П. Романюк. – Вид. 4-те доп. та перероб. К.: КНЕУ, 2006.– 528 с.

Інформаційні ресурси

16. Вища освіта України і Болонський процес / Навчальна програма. – Київ Тернопіль: ТДПУ ім. В. Гнатюка 2004.– 18с.
17. ІСУЯ 7.5.1– 03.01/УН «Загальні вимоги до організації процесу проведення навчальних занять».
18. ІСУЯ 7.5.1 – 03.02 / УН «Загальні вимоги до організації методичного забезпечення виконання індивідуальних завдань з дисциплін».
19. ІСУЯ 7.5.1 – 03.03 / УН «Загальні вимоги до організації виконання індивідуальних завдань».
20. ІСУЯ 7.5.1 – 03.04 / УН «Загальні вимоги до організації СРС».
21. ІСУЯ 7.5.1 – 03.05 / УН «Загальні вимоги до організації НДРС».
22. ІСУЯ 7.5.1 – 03.06 / УН «Загальні вимоги до організації поточного контролю».
23. ІСУЯ 7.5.1 – 03.07 / УН «Загальні вимоги до організації підсумкового контролю».
24. ІСУЯ 7.5.1 – 03.08 / УН «Критерії забезпеченості дисциплін навчально-методичною літературою».
25. ІСУЯ 7.5.1 – 03.09 / УН «Загальні вимоги до видання навчально-методичної літератури».

ЧАСТИНА 2.
НАВЧАЛЬНО-МЕТОДИЧНІ МАТЕРІАЛИ З ПІДГОТОВКИ ДО ПРАКТИЧНИХ ЗАНЯТЬ

1. Основні положення кореляційно-регресійного аналізу як розділу економетрики

Економетрика – це наука, що вивчає кількісні закономірності і взаємозалежності економічних процесів і об’єктів за допомогою математико-статистичних методів і моделей.

Означення. Економетрична модель – це рівняння (або система рівнянь), яке наближенно представляє основні кількісні залежності між економічними змінними, що вивчаються.

В найпростіших випадках економетричні моделі є структурними, тобто такими, що відображають сутність, структуру, внутрішні взаємозв’язки досліджуваного економічного об’єкта чи явища. Така економетрична модель кількісно описує взаємозв’язок між вхідною змінною X економічної системи та вихідною змінною Y. У загальному випадку економетричну модель можна записати у вигляді

$$Y = f(X, u), \tag{1}$$

де u – стохастична складова економетричної моделі.

Змінна X може бути як детермінованою, так і стохастичною, але, враховуючи, що u – стохастична складова економетричної моделі, залежна змінна Y завжди буде випадковою незалежно від того, якою є вхідна змінна X. Отже, **економетрична модель завжди є стохастичною**.

Встановлення виду функції (1) є задачею **регресійного аналізу**.

Курс економетрики покликаний навчити різним способам виявлення зв’язків і закономірностей через економетричні моделі, що базуються на даних статистичних спостережень. Економетричний підхід вимагає аналізу відповідності вибраної економетричної моделі і об’єкта, вивчення причин, які приводять до необхідності зміни моделі.

Економетрика є тим інструментом, який дозволяє перейти від якісного рівня аналізу економічного явища до аналізу кількісного аналізу на основі вибіркових даних і отримання представлення про властивості генеральної сукупності.

Кореляційною залежністю між двома змінними називається функціональна залежність між значеннями однієї з них і середнім значенням іншої. Саме така залежність вивчається в економетриці.

На практиці кореляційна залежність вивчається на основі вибіркових даних, в результаті чого отримують **рівняння регресії**

$$y = f(x, a_0, ..., a_m), \qquad (2)$$

де y – середня вихідної змінної Y при фіксованому значенні вхідної змінної $X = x$,

$a_0,...,a_m$ – оцінки параметрів рівняння регресії.

Метою побудови регресійної моделі є вивчення відповідної генеральної сукупності об’єктів чи процесів. Модель будується на основі вибірки із n спостережень. Вибірка формується випадковим чином за певними правилами. Тому на основі вибіркових даних можна отримати тільки **оцінки** потрібних параметрів, а не значення самих цих параметрів. Оцінки значень вихідної змінної $\hat{y}$ чи параметрів $\hat{a}_0,...,\hat{a}_m$ позначаються y, $a_0,...,a_m$.

Залежну змінну Y називають також функцією відгуку, пояснюваною, результативною, **вихідною**, ендогенною змінною, результативною ознакою, регресантом, а незалежну змінну X – пояснюючою, **вхідною**, екзогенною змінною, фактором, факторною ознакою, регресором.

При кожному i-тому спостереженні ($i = 1,...,n$) величина $\hat{y}_i$ складається із двох доданків

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

де u_i – випадкова величина, яка називається **збуренням**.

Основні математичні поняття в регресійному аналізі.

Ранжована $x_1 < x_2 < ... < x_m$ послідовність значень випадкової величини X називається **варіаційним рядом**.

Числа x_i називаються **варіантами** варіаційного ряду.

Числа n_i що показують скільки разів зустрічається варіанта x_i в ряді спостереження, називаються **частотами**.

Перелік варіант x_i і відповідних їм частот n_i називається **статистичним рядом**.

Для дискретного статистичного ряду **середнє арифметичне** випадкової величини X обчислюється за формулою

$$\overline{x} = \frac{1}{m}\sum_{i=1}^{m} n_i \cdot x_i ,$$

де m – число варіант, $\sum_{i=1}^{m} n_i = n$ – обсяг вибірки.

Незміщена оцінка **дисперсії** випадкової величини X обчислюється за формулою

$$s_x^2 = \frac{n}{n-1}\left(\overline{x^2} - (\overline{x})^2 \right),$$

де $\overline{x^2} = \frac{1}{m}\sum_{i=1}^{m} n_i x_i^2$.

Стандартне відхилення випадкової величини X дорівнює кореню квадратному із її дисперсії

$$s_x = \sqrt{s_x^2} .$$

Стандартне відхилення s_x є мірою розсіяння випадкової величини X.

Вибіркова коваріація $Cov(X,Y)$ виступає як мірою взаємозв’язку між двома випадковими величинами X і Y, так і мірою розсіяння значень цих двох змінних відносно їх арифметичних середніх. Вибіркова коваріація розраховується за формулою

$$Cov(X,Y) = \frac{n}{n-1}\left(\overline{XY} - \overline{X} \cdot \overline{Y}\right),$$

де $\overline{XY} = \frac{1}{m}\sum_{i=1}^{m} n_i x_i y_i$.

Вибіркова коваріація дорівнює нулю, якщо величини X і Y незалежні. Такі величини ще називають некорельованими.

Коваріація однієї змінної дорівнює її дисперсії, тобто

$$Cov(X) = s_x^2 .$$

Коваріація величина розмірна. Її розмірність залежить від розмірності і чисельного значення величин, для яких вона була обчислена, це створює незручності в використанні характеристики взаємозв’язку величин.

Більш точною мірою зв’язку між величинами X і Y порівняно з коваріацією є коефіцієнт кореляції, який є безрозмірним.

Вибірковий лінійний коефіцієнт кореляції визначається формулою

$$r = \frac{Cov(X,Y)}{s_x \cdot s_y}$$

Величина r є показником тісноти **лінійного** зв'язку між величинами X і Y.

Властивості коефіцієнта кореляції

1. Коефіцієнт r приймає значення на відрізку $[-1;1]$
2. Якщо $r > 0$, то зв'язок прямий, якщо $r < 0$ – зв'язок обернений.
   При прямому (оберненому) зв'язку збільшення вхідної змінної веде до збільшення (зменшення) вихідної змінної.
3. При $r = \pm 1$ зв'язок між x і y функціональний.
4. При $r = 0$ лінійний зв'язок між x і y відсутній.

Властивості оцінок параметрів рівняння регресії

Оцінки параметрів моделі повинні задовольняти умовам:

- **обґрунтованості**,
- **незміщеності,**
- **ефективності.**

Оцінка a називається **обґрунтованою**, якщо для довільного, як завгодно малого числа $\varepsilon > 0$ виконується умова

$$\lim_{n \to \infty} P(|a - \hat{a}| < \varepsilon) = 1.$$

Тобто, якщо обсяг вибірки прямує до нескінченності ($n \to \infty$), то ймовірність відхилення оцінки параметра a від значення параметра $\hat{a}$ стає все меншою.

Незміщеність означає, що математичне сподівання оцінки параметра $M[a]$ дорівнює значенню параметра

$$M[a] = \hat{a}.$$

Зміщення q оцінки визначається так

$$q = M[a] - \hat{a}$$

Наявність зміщення або його відсутність може бути перевірена з допомогою відношення середнього квадратичного відхилення параметра s_a до абсолютної величини $|a|$.

Якщо

$$\frac{s_a}{|a|} > 0{,}1,$$

то роблять висновок про зміщеність оцінки.

Незміщена оцінка a називається **ефективною**, якщо вона має найменшу дисперсію, серед всіх можливих незміщених оцінок параметра $\hat{a}$, обчислених на основі вибірок однакового обсягу n.

Поряд з коефіцієнтом кореляції, одним із показників, що характеризують кількісний зв'язок між залежною і незалежною змінними є **коефіцієнт еластичності.**

Для лінійної парної регресійної моделі даний коефіцієнт розраховується за формулою

$$E_{y/x} = \frac{\partial y}{\partial x} \cdot \frac{\bar{x}}{\bar{y}} = b\frac{\bar{x}}{\bar{y}}.$$

Він показує, на скільки відсотків в середньому зміниться залежна змінна при зміні незалежної змінної в середньому на один відсоток при інших незмінних умовах.

Однією з найбільш ефективних оцінок адекватності регресійної моделі, характеристикою прогностичної сили моделі є **коефіцієнт детермінації** R^2.

Розглянемо суму квадратів відхилень фактичних значень ознаки $\hat{y}_i$ від середнього значення $\bar{y}$

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

яку можна записати у вигляді

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

або

$$SST = SSR + SSE,$$

де $SST = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ – загальна сума квадратів відхилень залежної змінної від середнього значення,

$SSR = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$ – сума квадратів відхилень, обумовлених регресією,

$SSE = \sum_{i=1}^{n} (y_i - \bar{y})^2$ – залишкова сума квадратів, яка не пояснюється регресією і характеризує вплив на пояснювану змінну всіх неврахованих чинників.

Поділимо ліву і праву частину рівності $SST = SSR + SSE$ на SST, що дасть

$$1 = \frac{SSR}{SST} + \frac{SSE}{SST}.$$

Перший доданок називається **коефіцієнтом детермінації** і позначається

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},$$

або

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

Для лінійної моделі коефіцієнт детермінації може бути знайденим також і за формулою:

$$R^2 = r^2.$$

Величина R^2 показує, яка частка варіації змінної Y пояснюється за допомогою регресії. Наприклад, якщо R^2=0,90, то це означає, що на 90% зміна Y обумовлена зміною X, тобто пояснюється моделлю, а на 10% іншими причинами.

Коефіцієнт R^2 змінюється в межах $0 \le R^2 \le 1$. Чим ближче R^2 до 1 тим краще рівняння регресії апроксимує емпіричні дані.

Ще одним показником якості моделі є **середня похибка апроксимації** A, яка обчислюється за формулою

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

Прийнятна межа значень A не більше $8-10\%$. Якщо $A > 10\%$, то модель вважається неадекватною.

2. Парний лінійний кореляційно-регресійний аналіз

Найпростішою задачею регресійного аналізу є **парний** регресійний аналіз, де вивчається залежність між двома змінними.

Найбільш простим і одночасно таким, що найчастіше зустрічається є парний **лінійний** зв'язок між ознаками

$$y = a + b \cdot x,$$

де y – оцінка вихідної змінної $\hat{y}$,

a, b – оцінки параметрів $\hat{a}_0$ і $\hat{a}_1$.

Нехай маємо вибіркові дані про сукупні доходи населення в містах X (ум. од.) і роздрібний товарооборот Y (ум. од.), які представлені в табл. 1.

Таблиця 1

Доходи населення, X (ум. од.)	10	12	14	15	16	18	20
Роздрібний товарооборот, Y(ум. од.)	8	11	13	14	15	15	18

Найпростіший спосіб вибору форми кривої на основі візуального аналізу. Представимо вибіркові дані точками на координатній площині. Отримане зображення статистичної залежності називається **кореляційним полем** (рис.1).

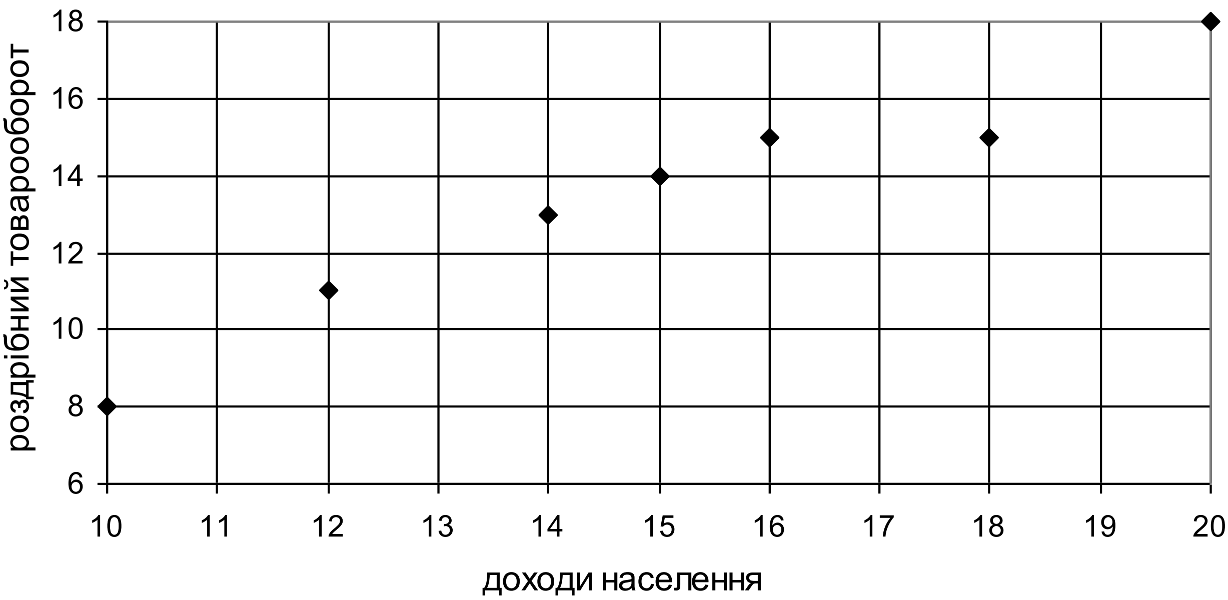

Рис. 1. Кореляційне поле

З виду кореляційного поля можна зробити припущення, що залежність між величинами X і Y лінійна.

Тобто вид кореляційного поля дозволяє зробити попереднє припущення про вид рівняння.

Оцінки параметрів a, b знаходяться **методом найменших квадратів** (МНК). Тобто a, b вибираються такими, щоб сума квадратів відхилень фактичних значень $\hat{y}$ від теоретичних y була мінімальною

$$S(a,b) = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} (\hat{y}_i - a - b \cdot x_i)^2 \to \min_{a,b},$$

де n – обсяг вибірки.

Згідно необхідної умови існування екстремуму функції двох змінних $S(a,b)$, прирівнюючи до нуля її перші частинні похідні, запишемо систему рівнянь

$$\begin{cases} \dfrac{\partial S(a,b)}{\partial a} = -2\sum\limits_{i=1}^{n} (\hat{y}_i - a - b \cdot x_i) = 0 \\ \dfrac{\partial S(a,b)}{\partial b} = -2\sum\limits_{i=1}^{n} (\hat{y}_i - a - b \cdot x_i)x_i = 0 \end{cases}$$

Звідки після елементарних перетворень отримуємо систему нормальних рівнянь для знаходження оцінок параметрів a, b

$$\begin{cases} a + b \cdot \bar{x} = \bar{y} \\ a \cdot \bar{x} + b \cdot \overline{x^2} = \overline{xy} \end{cases},$$

де $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} \hat{y}_i$, $\overline{x^2} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^2$, $\overline{xy} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i \hat{y}_i$

Розв’язуючи систему лінійних рівнянь отримуємо

$$b = r \cdot \frac{s_y}{s_x},$$

$$a = \bar{y} - b \cdot \bar{x}.$$

Коефіцієнт b називається **коефіцієнтом регресії** Y на X.
Він показує на скільки одиниць зміниться Y при зміні X на одну одиницю.
Коефіцієнт a називають вільним членом.

Таким чином, обчисливши за формулами оцінки параметрів a, b, отримаємо рівняння регресії

$$y = a + b \cdot x, \quad \text{або} \quad y - \bar{y} = a + b \cdot (x - \bar{x}),$$

$$\frac{y - \bar{y}}{s_y} = r\frac{x - \bar{x}}{s_x}.$$

Це є рівняння регресії в нормалізованому вигляді.

Як видно з останнього рівняння лінія регресії проходить через точку з координатами ($\bar{x}, \bar{y}$) на кореляційному полі.

В рівнянні регресії в нормалізованому вигляді застосовані усі статистичні дані:

– арифметичні середні $\bar{y}$, $\bar{x}$, які визначають початок відліку,

– стандартні відхилення s_y, s_x які визначають одиниці вимірювання Y та X,

– вибірковий коефіцієнт кореляції r, який визначає тісноту лінійного зв'язку між Y та X.

Лінійний коефіцієнт кореляції може обчислюватися за формулою

$$r = b\frac{s_x}{s_y}.$$

Для лінійної моделі **коефіцієнт детермінації** може бути знайденим за формулою:

$$R^2 = r^2.$$

Оцінка значущості рівняння регресії

Число ступенів свободи. При аналізі регресійної моделі і оцінці значущості моделі і її параметрів важливим поняттям є число ступенів свободи.

Число ступенів свободи (DF – Degree of Freedom) деякої характеристики – це **число незалежних змінних**, які необхідні для визначення даної характеристики.

Розглянемо, як знаходиться число ступенів свободи для кожного із складових виразу

$$SST = SSR + SSE.$$

- ✓ Число ступенів свободи для SST:

$$DF_{SST} = n - 1,$$

- ✓ Число ступенів свободи для SSR:

$$DF_{SSR} = m - 1$$

(m – число незалежних параметрів),

- ✓ Число ступенів свободи для SSE:

$$DF_{SSE} = n - m.$$

Маємо

$$DF_{SST} = n-1,$$
$$DF_{SSR} = m-1,$$
$$DF_{SSE} = n-m,$$

$$\boxed{DF_{SST} = DF_{SSR} + DF_{SSE},}$$

тому що

$$n-1 = (m-1)+(n-m).$$

Середні квадрати. В дисперсійному аналізі і при перевірці адекватності моделей часто використовується поняття середнього квадрата.

Середній квадрат – це сума квадратів поділена на відповідне число ступенів свободи.

Тоді середній квадрат похибок дорівнює

$$MSE = \frac{SSE}{n-m} = \frac{SSE}{n-2},$$

тому, що $m = 2$.

Середній квадрат похибок MSE представляє собою незміщену оцінку дисперсії залишків. Дисперсія залишків характеризує ту частину загальної дисперсії, що спричинена неврахованими факторами, помилками формування вибіркової сукупності, похибками вимірювання

$$MSE = s_u^2.$$

Середній квадрат, який пояснюється регресію,

$$MSR = \frac{SSR}{m-1} = SSR.$$

Середній квадрат MSR представляє собою незміщену оцінку дисперсії залежної змінної Y, яка спричинена змінною X

$$MSR = s_R^2.$$

Для загальної суми поняття середнього квадрата не вводиться.
Оцінкою загальної дисперсії є величина

$$s^2 = \frac{SST}{n-1}.$$

Перевірка значущості рівняння регресії означає встановлення відповідності між математичною моделлю і експериментальними даними. Така необхідність виникає у зв'язку з тим, що рівняння регресії було побудовано на основі вибіркових (тобто випадкових) даних, а отже і параметри його є випадковими величинами.

Оцінка значущості рівняння регресії проводиться на основі дисперсійного аналізу з використанням статистичних критеріїв:

F - критерія Фішера (Фішера –Снедекера) і t - критерія Стьюдента.

F- критерій Фішера служить для перевірки нульової гіпотези $H_0 : b = 0$, що нахил прямої дорівнює нулю ($b = 0$).

Якщо дана гіпотеза підтверджується, то дані краще апроксимувати з допомогою середньої величини $y_i = \bar{y}$, а не використовувати для цього рівняння регресії.

Для цього обчислюють **фактичне значення** F – критерія

$$F = \frac{MSR}{MSE},$$

Або

$$F = \frac{SSR}{SSE} \frac{n-m}{m-1},$$

$$F = \frac{R^2}{1-R^2} \frac{n-m}{m-1}.$$

Рівняння регресії значуще на рівні α , якщо

$$F > F(\alpha, k_1, k_2),$$

де $F(\alpha, k_1, k_2)$– табличне значення F - критерія Фішера, визначене на рівні значущості α при $k_1 = m - 1$ степенях свободи чисельника і $k_2 = n - m$ степенях свободи знаменника.

Рівнем значущості α називається ймовірність допустити помилку першого роду, тобто відкинути правильну гіпотезу $H_0 : b = 0$.

Значущість рівняння парної лінійної регресії може бути перевірена також і через перевірку значущості коефіцієнтів регресії a, b і коефіцієнта лінійної кореляції r за допомогою t - критерія Стьюдента.

Критерій Стьюдента служить для перевірки нульової гіпотези $H_0 : b = 0$ ($a = 0,\ r = 0$) тобто, що вказані показники мають випадкову природу і незначно відрізняються від нуля.

Для цього розраховуються фактичні значення $t-$ критерія Стьюдента за формулами

$$t_b = \frac{|b|}{s_b}, \qquad t_a = \frac{|a|}{s_a}, \qquad t_r = \frac{|r|}{s_r},$$

де s_b, s_a, s_r –стандартні похибки (середні квадратичні відхилення) відповідних показників.

Для парної регресії вказані стандартні похибки розраховуються за формулами

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

$$s_a = s_u \sqrt{\frac{\overline{x^2}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}},$$

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}}.$$

Рівняння парної лінійної регресії, або коефіцієнт регресії b, значущий на рівні α (тобто гіпотеза $H_0 :\ b = 0$ відкидається), якщо

$$t_b > t(1 - \alpha, n - 2),$$

де $t(1 - \alpha, n - 2)$ –табличне значення t - критерія Стьюдента на рівні значущості α , при числі ступенів свободи $n - 2$.

Тут число ступенів свободи дорівнює $n - 2$, тому, що для парної регресії число параметрів дорівнює двом, тому $n - m = n - 2$.

Аналогічно перевіряється значущість параметра b. Якщо

$$t_a > t(1 - \alpha, n - 2),$$

то параметр a значущий на рівні α.

Якщо встановлено, що рівняння регресії значуще, то це означатиме, що і коефіцієнт кореляції r також значущий.

Оцінка значущості лінійного коефіцієнта кореляції r може бути виконана незалежно від оцінки значущості рівняння і параметрів регресії.

Лінійний коефіцієнт кореляції r значущий на рівні α, якщо

$$t_r > t(1-\alpha, n-2).$$

Величину α називають рівень значущості, а $(1-\alpha)$ – рівень довіри. Рівні значущості найчастіше приймають рівними 0,10; 0,05; 0,01.

Прогноз залежної змінної.
Інтервальна оцінка функції регресії і її параметрів.

Рівняння регресії будується на основі вибіркових даних. Вибірка формується випадковим чином, тому і оцінки параметрів рівняння регресії a і b є випадковими величинами. Отже, обчислені значення є насправді **точковими** оцінками істинних значень параметрів $\hat{a}$ і $\hat{b}$.

Поряд з точковими оцінками параметрів рівняння регресії можуть бути обчислені й **інтервальні оцінки**. Інтервальні оцінки параметрів a і b знаходяться за формулами:

$$a - s_a \cdot t(1-\alpha; n-2) < \hat{a} < a + s_a \cdot t(1-\alpha; n-2),$$
$$b - s_b \cdot t(1-\alpha; n-2) < \hat{b} < b + s_b \cdot t(1-\alpha; n-2).$$

Одна з головних задач і один з важливих етапів регресійного аналізу це **здійснення прогнозу**.

Якщо в модель (рівняння регресії) підставимо очікуване значення пояснюючої змінної $x_{пр}$, то отримаємо **точковий** прогноз значення $y_{пр}$

$$y_{пр} = a + b \cdot x_{пр}.$$

Щоб знайти інтервальну оцінку для умовного середнього значення $\bar{Y}(x)$, тобто довірчий інтервал для $\bar{Y}(x)$, необхідно знайти середнє квадратичне відхилення $\bar{Y}(x)$: s_y.

Рівняння регресії може бути записане у вигляді

$$y_{пр} = \bar{y} + b \cdot (x_{пр} - \bar{x})$$

В даному рівнянні $\bar{y}$ і b випадкові незалежні змінні, множник $(x_{пр} - \bar{x})$ не є випадковою величиною.

Тоді дисперсія $s^2_{y_{пр}}$ буде мати дві складові

$$s^2_{y_{пр}} = s^2_{\bar{y}} + s^2_b \cdot (x_{пр} - \bar{x})^2 .$$

Знаходимо

$$s_{\bar y}^2 = s^2(\bar y) = s^2\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} s_y^2 = \frac{s_y^2}{n},$$

де $s_y^2 = s_u^2$.

Для знаходження s_b^2 представимо оцінку цього параметра у вигляді:

$$b = \frac{\sum_{i=1}^{n} (x_i - \bar x)(y_i - \bar y)}{\sum_{i=1}^{n} (x_i - \bar x)^2}.$$

Тоді враховуючи, що x_i і $\bar x$ не є випадковими величинами отримуємо

$$s_b^2 = \frac{\sum_{i=1}^{n} (x_i - \bar x)^2}{(\sum_{i=1}^{n} (x_i - \bar x)^2)^2} s_y^2,$$

$$s_b^2 = \frac{s_u^2}{\sum_{i=1}^{n} (x_i - \bar x)^2}.$$

Таким чином дисперсія $s_{y_{np}}^2$ знаходиться за формулою

$$s_{y_{np}}^2 = \frac{s_u^2}{n} + \frac{s_u^2}{\sum_{i=1}^{n} (x_i - \bar x)^2}(x_{np} - \bar x)^2,$$

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

Тоді довірчий інтервал для умовного середнього $\bar Y(x)$ матиме вигляд

$$y_{np} - s_{y_{np}} \cdot t\,(1-\alpha; n-2) < \bar Y(x) < y_{np} + s_{y_{np}} \cdot t\,(1-\alpha; n-2).$$

Встановлений за формулою інтервал означає, що з ймовірністю $(1-\alpha)$, середнє значення можливих значень Y буде знаходитися в межах

$$(y_{np} - s_{y_{np}} \cdot t\,(1-\alpha;\; n-2);\quad y_{np} + s_{y_{np}} \cdot t\,(1-\alpha;\; n-2)).$$

Ясно, що чим ширший інтервал, тим менш цінним буде прогноз.

З формул видно, що:

- ✓ інтервал буде тим вужчим, чим меншою є залишкова дисперсія, тобто, чим точніше рівняння регресії апроксимує фактичні дані;
- ✓ найвужчим (найточнішим) буде прогнозний інтервал для $x = \bar{x}$, чим більше відхиляється x від $\bar{x}$ тим ширшим буде інтервал;
- ✓ інтервал можна звузити зменшивши надійність $1-\alpha$, так як при цьому зменшується $t(1-\alpha; n-2)$.

Крім встановлення довірчого інтервалу для умовного середнього значення $\bar{Y}(x)$ може бути обчисленим також довірчий інтервал для індивідуального значення. Ясно, що довірчий інтервал для індивідуальних значень буде значно ширшим ніж довірчий інтервал для умовного середнього значення.

При визначенні довірчих інтервалів для індивідуальних значень y_{np} необхідно враховувати ще одне джерело варіації – розсіяння навколо лінії регресії, тобто в оцінку дисперсії сумарної регресії треба включити ще й s_u^2.

Тоді дисперсія індивідуальних значень y_{np} при $x = x_{np}$ дорівнює

$$s^2_{y_{np}(i)} = s_u^2 (1 + \frac{1}{n} + \frac{(x_{np} - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}).$$

Відповідний довірчий інтервал для індивідуальних значень y_{np} знаходиться за формулою

$$y_{np} - s_{y_{np}(i)} \cdot t\ (1-\alpha; n-2) < Y_i(x) < y_{np} + s_{y_{np}(i)} \cdot t\ (1-\alpha; n-2)$$

Приклад 1. За даними вибіркового дослідження відомі значення двох показників: X – середньоденна заробітна плата одного працюючого, Y – витрати на придбання продовольчих товарів в відсотковому відношенні до загальних витрат.

Таблиця 2

X, грн.	45	60	57	62	59	47	55
Y , %	69	61	60	57	55	67	65

Побудувати лінійну модель залежності Y від X. Розрахувати коефіцієнти кореляції, детермінації.

Оцінити модель через середню похибку апроксимації A та F - критерій Фішера, зробити прогноз для індивідуальних значень y_{np} при 75 грн.

Розв'язання.

Будуємо кореляційне поле.

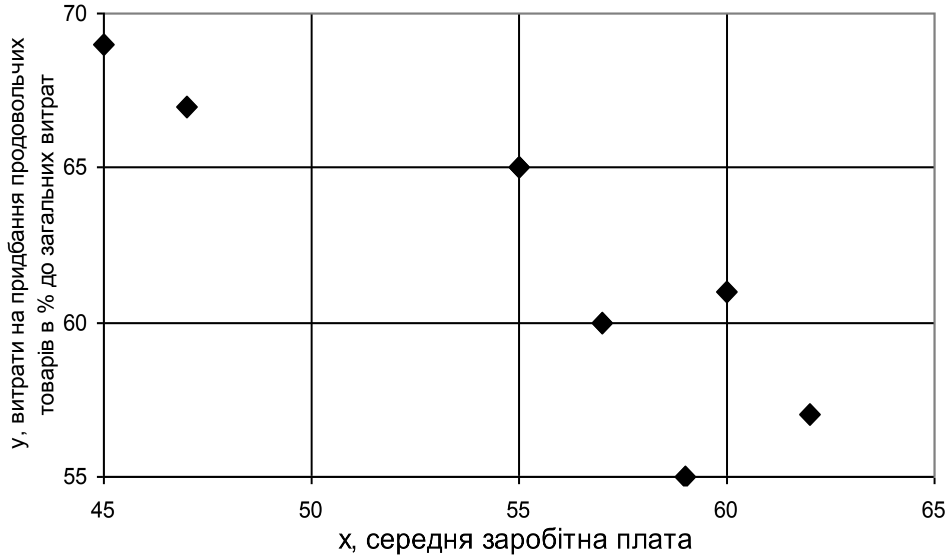

Рис. 2. Кореляційне поле.

По розташуванні точок на кореляційному полі робимо висновок про лінійну залежність, тобто модель парної регресії має вигляд

$$y = a + b \cdot x.$$

Спочатку знайдемо рівняння прямої аналітично, тобто по двом точкам. Вибираємо точки $(45;69)$, $(62;57)$. Записуємо рівняння прямої за двома точками:

$$\frac{x-45}{62-45} = \frac{y-69}{57-69},$$

$$\frac{x-45}{17} = \frac{y-69}{-12},$$

$$y - 69 = -\frac{12}{17}(x-45),$$

$$y = -\frac{12}{17}(x-45) + 69,$$

$$y = -\frac{12}{17}x + \frac{540}{17} + 69,$$

$$y = -\frac{12}{17}x + \frac{1713}{17},$$

$$y = 100{,}8 - 0{,}706 \cdot x.$$

Для розрахунку параметрів a і b лінійної регресії методом МНК записуємо систему нормальних рівнянь

$$\begin{cases} a + b \cdot \bar{x} = \bar{y} \\ a \cdot \bar{x} + b \cdot \overline{x^2} = \overline{xy} \end{cases},$$

$\bar{x}$, $\bar{y}$, $\overline{x^2}$, $\overline{xy}$ – середні значення.

Розрахунки середніх значень можна проводити використовуючи розрахункову табл. 3.

Таблиця 3

№	$\hat{y}_i$	x_i	$\hat{y}_i x_i$	x_i^2	$\hat{y}_i^2$	y_i	$\hat{y}_i - y_i$	A_i	$(\hat{y}_i - y_i)^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2$
1	69	45	3105	2025	4761	68,04	0,95	1,67	0,909	36,56	100
2	61	60	3660	3600	3721	58,97	2,02	1,89	4,093	9,14	25
3	60	57	3420	3249	3600	60,79	-0,79	1,92	0,625	1,461	4
4	57	62	3534	3844	3249	57,76	-0,77	2,02	0,589	17,91	49
5	55	59	3245	3481	3025	59,58	-4,58	2,09	20,98	5,85	16
6	67	47	3149	2209	4489	66,83	0,16	1,72	0,026	23,40	64
7	65	55	3575	3025	4225	62,00	3	1,77	8,99	0	0
$\sum$	434	385	23688	21433	27070		0	13,06	36,23	94,32	258
Середні	62	55	3384	3061,9	3867,1			1,87			
s^2	27	43	–	–	–						
s	5,2	6,6									

За формулами знаходимо:

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x^2} = \frac{3384 - 62 \cdot 55}{43} \approx -0{,}61$$

$$a = \bar{y} - b \cdot \bar{x} = 62 - (-0{,}61) \cdot 55 = 95{,}26$$

Рівняння регресії має вигляд:

$$y = 95{,}26 - 0{,}61 \cdot x .$$

Якщо порівняти рівняння, знайдене аналітично

$$y = 100{,}8 - 0{,}706 \cdot x$$

і рівняння регресії, то є відповідні різниці, але невеликі.

На рис. 3 представлені графіки відповідних рівнянь.
Тобто аналітичний метод теж можна застосовувати, як перший крок в моделюванні.

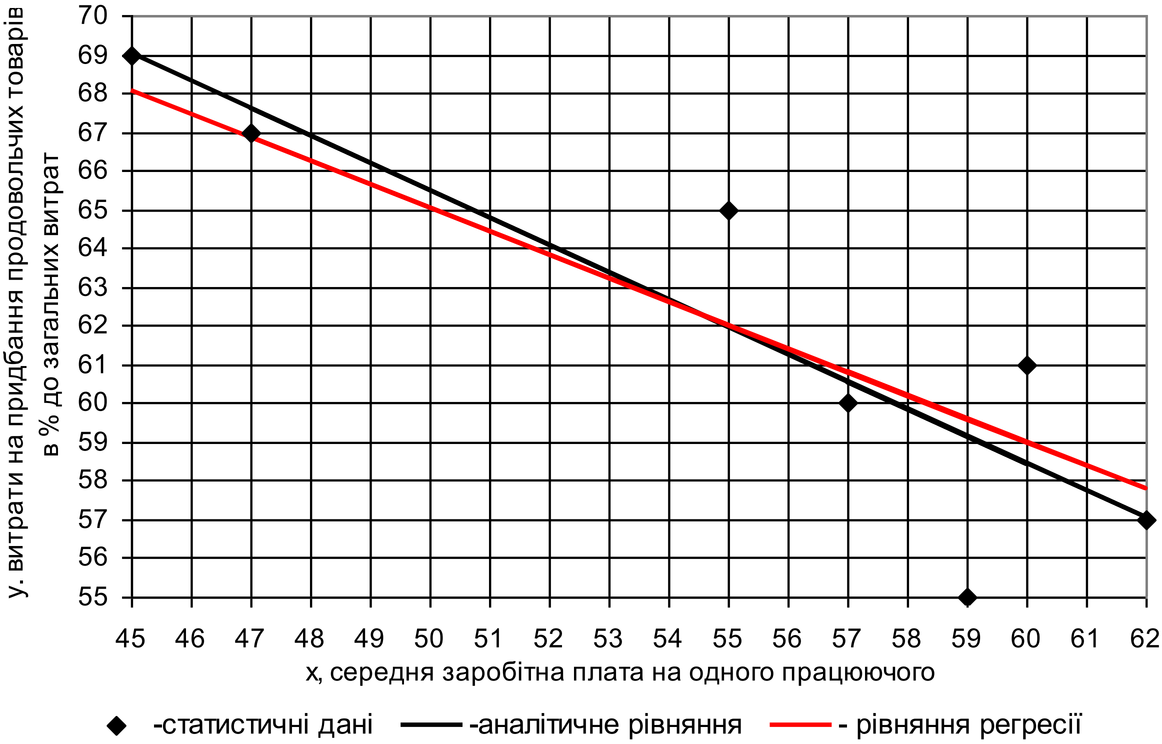

Рис. 3. Статистичні дані, графік рівняння, знайденого аналітично, і рівняння регресії

Із рівняння регресії витікає, що при збільшенні середньої заробітної плати на 1 грн. частка витрат на придбання продовольчих товарів знижується в середньому на 0,605%.

Розрахуємо лінійний коефіцієнт парної кореляції за формулою

$$r = b\frac{s_x}{s_y} = -0{,}61\frac{6{,}6}{5{,}2} = \text{- } 0{,}774,$$

тобто лінійний зв'язок – обернений, тісний.

Визначаємо коефіцієнт детермінації за формулою

$$R^2 = r^2 = (-0{,}774)^2 = 0{,}60$$

Варіація результативної ознаки на 60 % пояснюється варіацією змінної x.

Підставляючи в рівняння регресії фактичні значення x_i, визначаємо теоретичні (розрахункові) значення y_i.

В останньому стовпчику таблиці записано значення величини

$$A_i = \left|\frac{\hat{y}_i - y_i}{y_i}\right| \cdot 100\% \ .$$

Знаходимо величину середньої похибки апроксимації $\overline{A}$

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n} A_i = 1{,}87\% .$$

В середньому розрахункові значення відхиляються від фактичних на 1,87%, що знаходиться в допустимих межах.

Оцінимо значущість рівняння регресії за допомогою F – критерія Фішера.

Знайдемо фактичне значення F – критерія. Для цього в розрахунковій таблиці можна додатково внести колонки. Тоді

$$F = \frac{s_R^2}{s_u^2},$$

$$F = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}\frac{n-m}{m-1} = \frac{94{,}33}{36{,}23}\cdot\frac{7-2}{2-1} = 13{,}02 \ .$$

Знаходимо табличне значення критерія Фішера $F(0{,}05;1;5) = 6{,}61$.
Оскільки $F = 13{,}02 > F(0{,}05;1;5) = 6{,}61$, отримане рівняння є статистично значущим.

Знайдемо прогнозні значення для індивідуальних значень y_{np} при $x_{np} = 75$ грн.

Знаходимо точкове значення прогнозу: $y = 95{,}26 - 0{,}61 \cdot x$

$$y_{np} = 95{,}26 - 0{,}61 \cdot x_{np}$$

$$y_{np} = 95{,}26 - 0{,}61 \cdot 75 = 49{,}51 \ \%.$$

Для лінійної моделі число параметрів $m = 2$, тому незміщена оцінка дисперсії залишків

$$s_u^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-m} = \frac{36{,}23}{7-2} = 7{,}246$$

$$s_u = 2{,}69$$

Знаходимо дисперсію індивідуальних значень $y_{пр}$ при $x = x_{пр}$

$$s^2_{y_{пр}(i)} = s_u^2\left(1 + \frac{1}{n} + \frac{(x_{пр} - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) = 7{,}246\left(1 + \frac{1}{7} + \frac{(75-55)^2}{258}\right) = 19{,}52,$$

$$s_{y_{пр}(i)} = 4{,}42.$$

За таблицями знаходимо

$$t(\alpha, n-m) = t(0{,}05;5) = 2{,}57.$$

За формулою

$$y_{пр} - s_{y_{пр}(i)} \cdot t\ (1-\alpha; n-2) < Y_i(x_{пр}) < y_{пр} + s_{y_{пр}(i)} \cdot t\ (1-\alpha; n-2)$$

отримуємо:

$$49{,}51 - 2{,}57 \cdot 4{,}42 < Y_i(x_{пр}) < 49{,}51 + 2{,}57 \cdot 4{,}42.$$

Отже з ймовірністю 0,95 витрати на придбання продовольчих товарі будуть знаходитися в інтервалі

$$38{,}15 < Y_i(x_{пр}) < 60{,}87.$$

3. Парний нелінійний регресійний аналіз

Якщо між економічними показниками існують нелінійні співвідношення, то вони виражаються за допомогою відповідних нелінійних функцій.

Серед нелінійних регресійних рівнянь розрізняють два класи:

- ✓ регресійні рівняння, нелінійні за вхідною змінною, але лінійні за оцінюваними параметрами,
- ✓ регресійні рівняння, нелінійні за оцінюваними параметрами.

Прикладами регресійних рівнянь, нелінійних за вхідною змінною, можуть служити наступні:

- поліноміальні:

  $y = a + b \cdot x + c \cdot x^2$

  $y = a + b \cdot x + c \cdot x^2 + d \cdot x^3$
- гіперболічна

  $y = a + b \cdot \frac{1}{x},$
- логістична

$$y = \frac{1}{a + b \cdot x}$$

До регресійних рівнянь, нелінійних за оцінюваними параметрами, відносяться залежності:

- степенева
  $y = a \cdot x^b$,
- показникова
  $y = a \cdot b^x$,
- експоненційна
  $y = a \cdot e^{b \cdot x}$.

Регресійні рівняння, нелінійні за вхідною змінною, легко зводяться до лінійних регресійних рівнянь заміною змінних.

Так, наприклад,

$$y = a + b \cdot x_1 + c \cdot x_2 \ , \qquad y = a + b \cdot x_1 + c \cdot x_2 + d \cdot x_3 ,$$

де $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$.

Регресійні рівняння нелінійні за оцінюваними параметрами: степенева, показникова, експоненційна зводяться до лінійних шляхом логарифмування. Ці рівняння можна записати у вигляді:

- $\ln y = \ln a + b \cdot \ln x$
- $\ln y = \ln a + x \cdot \ln b$
- $\ln y = \ln a + b \cdot x$

Отримані рівняння після відповідної заміни змінної зводяться до лінійних рівнянь:

$$\ln y = y_1, \ \ln a = a_0, \ \ln b = a_1, \ \ln x = x_1$$

$$y_1 = a_0 + b \cdot x_1, \qquad y_1 = a_0 + a_1 \cdot x_1, \quad y_1 = a_0 + b \cdot x .$$

До отриманих рівнянь можна застосовувати метод найменших квадратів.

Дослідження нелінійних рівнянь парної регресії

Для нелінійного рівняння регресії тісноту зв'язку вхідної і вихідної змінних оцінює **індекс кореляції** ρ_{xy}, який знаходиться за формулою

$$\rho_{xy} = \sqrt{1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} ,$$

Індекс детермінації визначається за формулою $R^2 = \rho_{xy}^2$.
Величина даного показника знаходиться в межах: $0 \le \rho_{xy} \le 1$.
Чим ближче ρ_{xy} до одиниці, тим тісніший зв'язок розглянутих змінних, тим більш надійним буде знайдене рівняння регресії.

Зауваження. Якщо **нелінійне щодо вхідної змінної** рівняння при лінеаризації приймає форму лінійного рівняння парної регресії, то для оцінки тісноти зв'язку може бути використаний лінійний коефіцієнт кореляції. Величина якого в цьому випадку збігається з індексом кореляції.
Так, наприклад, для гіперболічного рівняння регресії:

$$\rho_{xy} = r_{y\frac{1}{x}}.$$

Зауваження. Для регресійних рівнянь, **нелінійних за оцінюваними параметрами**, лінійний коефіцієнт кореляції для перетворених значень ознак дає лише наближену оцінку тісноти зв'язку й чисельно не збігається з індексом кореляції.
Так, наприклад, для степеневої функції $\rho_{xy} \ne r_{\ln x \ln y}$, незважаючи на близькість їхніх значень.

Для оцінки значущості індексу детермінації використовується ***F*-критерій Фішера.**

Фактичне значення F - критерія Фішера знаходиться за формулою

$$F = \frac{R^2}{1-R^2}\frac{n-m}{m-1}.$$

Для оцінки значущості параметрів гіперболічної, показникової, степеневої, експоненціальної функцій використовується t***–критерій Стьюдента***.

Фактичні значення t – критерія Стьюдента знаходяться за формулами

$$t_b = \frac{|b|}{s_b}, \qquad t_a = \frac{|a|}{s_a},$$

де

$$s_b = \frac{s_u}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}, \qquad s_a = s_u \sqrt{\frac{\overline{x^2}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}.$$

Індекс детермінації ρ_{xy}^2 можна порівнювати з коефіцієнтом детермінації r_{xy}^2.

Для обґрунтування можливості застосування лінійної функції, чим більша кривизна лінії регресії, тим величина r_{xy}^2 менше ρ_{xy}^2. Близькість цих показників означає, що немає необхідності ускладнювати форму рівняння регресії і можна використати лінійну функцію. Практично, якщо величина

$$\rho_{xy}^2 - r_{xy}^2$$

не перевищує 0,1, то припущення про лінійну форму зв'язку вважається виправданим.

Приклад 2. Знаходження нелінійного рівняння парної регресії.
Для восьми різних територіальних одиниць України відомі значення двох ознак.

Таблиця 4.

Частка витрат на продовольчі товари в загальних витратах, y,%	53	55	56	57	58	59	60	65
Середньоденна заробітна плата, x, *грн*	64	59	58	55	55	56	53	53

Необхідно:

а) розрахувати параметри рівнянь регресій, заданих у вигляді функцій: рівносторонньої гіперболи, степеневої, показникової.

б) розрахувати індекси кореляції і оцінити якість рівнянь регресій.

Розв'язання.

Будуємо поле кореляції.

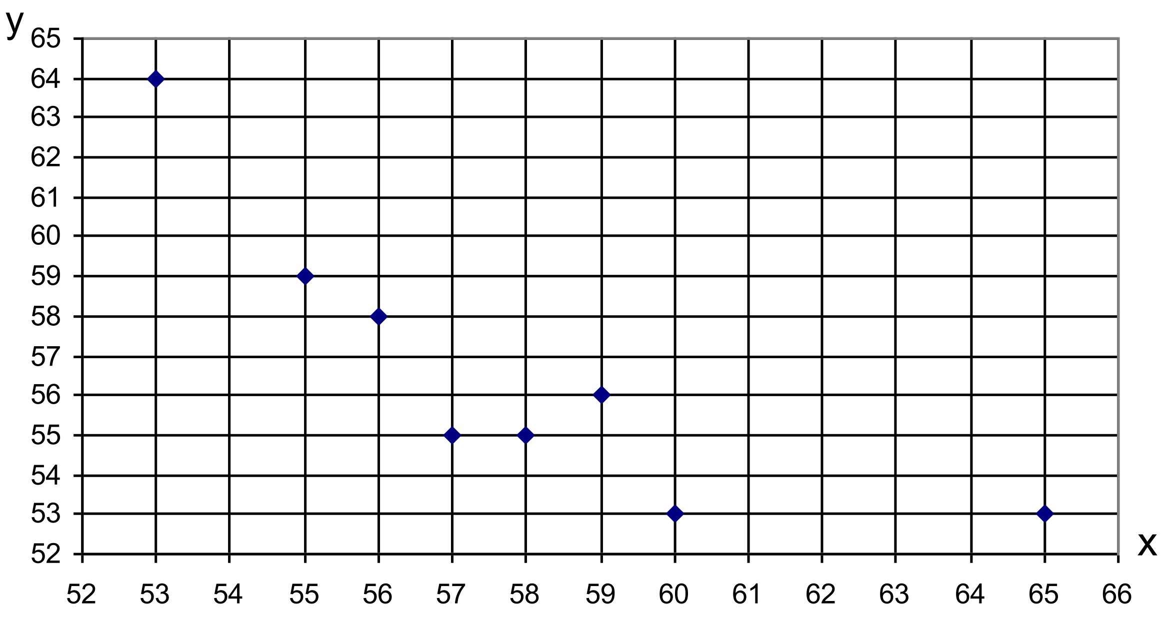

Рис. 5. Кореляційне поле.

Аналіз розташування точок на полі кореляцій показує, що має місце нелінійна парна регресія.

Згідно умови задачі розглядаємо структури функцій трьох видів

- рівностороння гіпербола $y = a + b\frac{1}{x}$,
- степенева $y = a \cdot x^b$,
- показникова $y = a \cdot b^x$.

Спочатку скористаємось **аналітичними методами** розв’язання задачі. Кожна з функцій має два невідомих параметра a і b, для знаходження яких треба два рівняння. Розглянемо кожну з функцій.

Рівностороння гіпербола.

$$y = a + b\frac{1}{x}$$

З таблиці вибираємо дві пари значень $(53;64)$ і $(60;53)$.
Підставляємо ці значення в рівняння гіперболи.

$$a + b\frac{1}{53} = 64,$$

$$a + b\frac{1}{60} = 53.$$

Маємо систему двох лінійних алгебраїчних рівнянь. Для розв’язання запишемо систему у вигляді

$$\begin{cases} 53a + b = 3392 \\ 60a + b = 3180 \end{cases}$$

Скористаємось формулами Крамера

$$\Delta = \begin{vmatrix} 53 & 1 \\ 60 & 1 \end{vmatrix} = 53 - 60 = -7 \neq 0 \qquad \Delta_a = \begin{vmatrix} 3392 & 1 \\ 3180 & 1 \end{vmatrix} = 3392 - 3180 = 212$$

$$\Delta_b = \begin{vmatrix} 53 & 3392 \\ 60 & 3180 \end{vmatrix} = 168540 - 203520 = -34980$$

$$a = \frac{\Delta_a}{\Delta} = \frac{212}{-7} = -30{,}29 \qquad b = \frac{\Delta_b}{\Delta} = \frac{-34980}{-7} = 4997{,}1$$

Таким чином, рівняння має вигляд

$$y = -30{,}29 + 4997{,}1 \cdot \frac{1}{x}.$$

Степенева функція

$$y = a \cdot x^b$$

З таблиці вибираємо дві пари значень $(53;64)$ і $(60;53)$. Підставляємо ці значення в рівняння степеневої функції.

$$a \cdot 53^b = 64$$
$$a \cdot 60^b = 53$$

Розв'язуємо цю систему. Ділимо одне рівняння на друге.

$$\frac{a \cdot 60^b}{a \cdot 53^b} = \frac{53}{64}, \qquad \left(\frac{60}{53}\right)^b = \frac{53}{64},$$

Логарифмуємо

$$b \cdot \ln\left(\frac{60}{53}\right) = \ln\left(\frac{53}{64}\right), \qquad b = \frac{\ln\left(\frac{53}{64}\right)}{\ln\left(\frac{60}{53}\right)}, \qquad b = -1{,}52,$$

$$a \cdot 53^{-1{,}52} = 64, \qquad a = 64 \cdot 53^{1{,}52}, \qquad a = 26761{,}61,$$

$$y = 26761{,}61 \cdot x^{-1{,}52}, \text{ або } \qquad y = \frac{26761{,}61}{x^{1{,}52}}$$

Показникова функція

$$y = a \cdot b^x$$

З таблиці вибираємо дві пари значень $(53;64)$ і $(60;53)$. Підставляємо ці значення в рівняння показникової функції.

$$64 = a \cdot b^{53}$$
$$53 = a \cdot b^{60}$$

Розв'язуємо цю систему. Ділимо одне рівняння на друге.

$$\frac{a \cdot b^{60}}{a \cdot b^{53}} = \frac{53}{64}, \qquad b^7 = \frac{53}{64}, \quad b = \sqrt[7]{\frac{53}{64}}, \quad b = 0{,}973$$

.

$$a = \frac{64}{0{,}973^{53}}, \qquad a = 273{,}0,$$

$$y = 273 \cdot 0{,}973^x.$$

В таблиці 5 приведені результати розрахунків по знайденим вище формулам.

Таблиця 5

№	x	y	$y_{гіп.}$	$y_{степ.}$	$y_{показ}$
1	53	64	63,99	64,06	64,00
2	55	59	60,57	60,56	60,64
3	56	58	58,94	58,92	59,03
4	57	55	57,38	57,36	57,46
5	58	55	55,87	55,86	55,93
6	59	56	54,41	54,43	54,45
7	60	53	53,00	53,05	53,00
8	65	53	46,59	46,98	46,32

На рис. 6 приведені відповідні графіки згідно даних табл. 5.

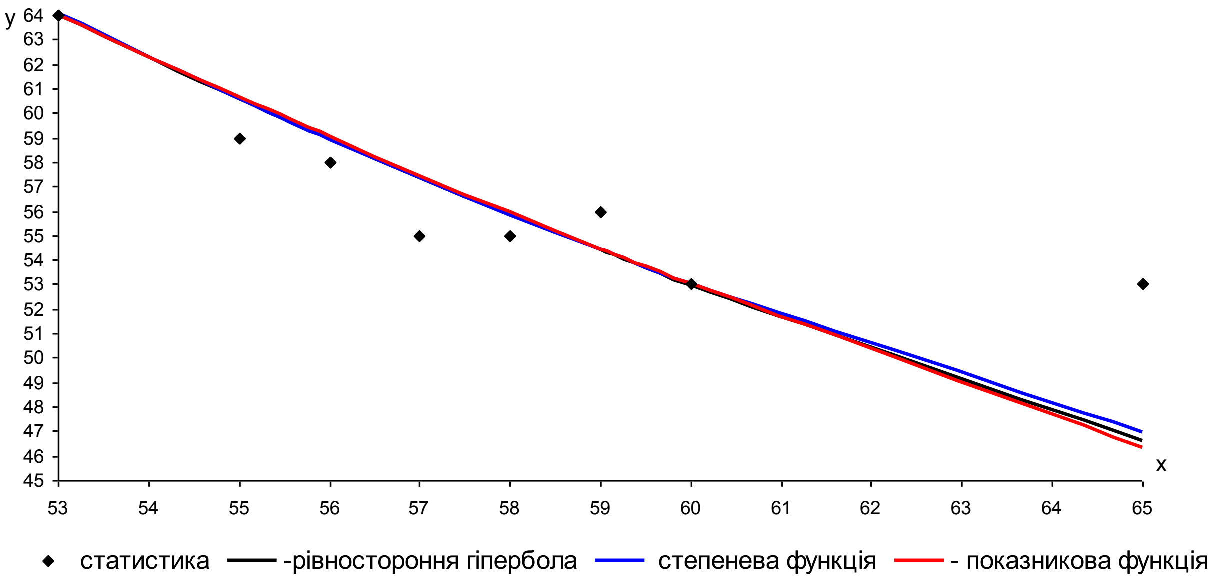

Рис. 6. Графіки функцій, знайдених аналітично, які наближають статистичні дані.

Аналіз графіків, приведених на рис. 6, показує, що вони практично співпадають. Таким чином, можна зробити висновок, що має місце ідентичність вибраних структур функцій.

Далі застосуємо МНК для знаходження параметрів заданих структур функцій.

Рівностороння гіпербола.

$$y = a + b\frac{1}{x}$$

Зробимо заміну змінної $\frac{1}{x} = t$.

Тоді рівняння приймає вигляд $y = a + b \cdot t$.

Записуємо систему нормальних рівнянь

$$\begin{cases} a + b \cdot \bar{t} = \bar{y} \\ a \cdot \bar{t} + b \cdot \overline{t^2} = \overline{y \cdot t} \end{cases}.$$

Знаходження середніх зводимо до таблиці 6.

Таблиця 6

№	x_i	y_i	$t_i = 1/x_i$	t_i^2	$y_i \cdot t_i$	$y_{i\,расч.}$	$(y_i - y_{i\,расч})^2$	$(y_i - \bar{y})^2$
1	53	64	0,018868	0,000356	1,207547	61,260	7,509	21,480
2	55	59	0,018182	0,000331	1,072727	59,183	0,034	6,544
3	56	58	0,017857	0,000319	1,035714	58,201	0,040	2,482
4	57	55	0,017544	0,000308	0,964912	57,252	5,073	0,394
5	58	55	0,017241	0,000297	0,948276	56,337	1,787	0,083
6	59	56	0,016949	0,000287	0,949153	55,452	0,300	1,375
7	60	53	0,016667	0,000278	0,883333	54,598	2,552	4,111
8	65	53	0,015385	0,000237	0,815385	50,717	5,210	34,900
середнє		56,625	0,017337	0,000302	0,984631			
сума							22,506	71,369

Тоді система нормальних рівнянь запишеться у вигляді.

$$\begin{cases} a + 0{,}017337 \cdot b = 56{,}625 \\ 0{,}017337a + 0{,}000302 \cdot b = 0{,}984631 \end{cases}$$

Розв'язуємо систему по формулам Крамера.

$$\Delta = \begin{vmatrix} 1 & 0{,}017337 \\ 0{,}017337 & 0{,}000302 \end{vmatrix} = 9{,}73938\text{E-}07 \qquad \Delta_a = \begin{vmatrix} 56{,}625 & 0{,}017337 \\ 0{,}984631 & 0{,}000302 \end{vmatrix} = 4{,}04712\text{E-}06$$

$$\Delta_b = \begin{vmatrix} 1 & 56{,}625 \\ 0{,}017337 & 0{,}984631 \end{vmatrix} = 0{,}00294765 \quad a = \frac{\Delta_a}{\Delta} = 4{,}155414 \qquad b = \frac{\Delta_b}{\Delta} = 3026{,}526$$

Таким чином, рівняння регресії має вигляд

$$y = 4{,}16 + \frac{3026{,}53}{x}.$$

Знайдемо індекс кореляції

$$\rho_{xy} = \sqrt{1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$$\rho_{xy} = \sqrt{1 - \frac{22{,}506}{71{,}369}} = 0{,}827$$

Тоді індекс детермінації знаходиться по формулі $R^2 = \rho_{xy}^2 = 0{,}685$.

Обчислюємо критерій Фішера

$$F = \frac{R^2}{1-R^2}\frac{n-m}{m-1} \qquad F = \frac{0{,}685}{1-0{,}685}\frac{8-2}{2-1} = 13{,}05$$

Знаходимо табличне значення

$$F(0{,}05;1;6) = 5{,}99\,.$$

Оскільки $F = 13{,}05 > F(0{,}05;1;6) = 5{,}99$, то нульова гіпотеза відхиляється і рівняння регресії статистично значуще.

<u>Степенева функція</u>

$$y = a \cdot x^b$$

Логарифмуємо рівняння

$$\ln y = \ln a + b \cdot \ln x\,.$$

Вводимо нові позначення

$$\ln y = z\,, \quad \ln a = \alpha\,, \qquad \ln x = t\,.$$

Тоді рівняння приймає вигляд

$$z = \alpha + b \cdot t$$

Для застосування МНК виписуємо систему нормальних рівнянь

$$\begin{cases} \alpha + b \cdot \bar{t} = \bar{z} \\ \alpha \cdot \bar{t} + b \cdot \overline{t^2} = \overline{zt} \end{cases}$$

Знаходження середніх зводимо до таблиці 7.

Таблиця 7.

№	x_i	y_i	$z_i = \ln y_i$	$t_i = \ln x_i$	t_i^2	$z_i \cdot t_i$	$y_{i\,расч.}$	$(y_i - y_{i\,расч})^2$	$(y_i - \bar{y})^2$
1	53	64	4,159	3,970	15,763	16,512	60,986	9,082	19,022
2	55	59	4,078	4,007	16,059	16,340	59,030	0,001	5,784
3	56	58	4,060	4,025	16,203	16,345	58,101	0,010	2,179
4	57	55	4,007	4,043	16,346	16,202	57,203	4,852	0,334
5	58	55	4,007	4,060	16,487	16,272	56,334	1,779	0,085
6	59	56	4,025	4,078	16,626	16,414	55,492	0,258	1,283
7	60	53	3,970	4,094	16,764	16,256	54,677	2,814	3,793
8	65	53	3,970	4,174	17,426	16,574	50,958	4,172	32,120
середнє			4,035	4,057	16,459	16,364			
сума								22,967	64,599

Тоді система нормальних рівнянь запишеться у вигляді.

$$\begin{cases} \alpha + 4{,}057 \cdot b = 4{,}035 \\ 4{,}057\alpha + 16{,}459 \cdot b = 16{,}364 \end{cases}$$

Розв'язуємо систему по формулам Крамера.

$$\Delta = \begin{vmatrix} 1 & 4{,}057 \\ 4{,}057 & 16{,}459 \end{vmatrix} = 0{,}003 \qquad \Delta_\alpha = \begin{vmatrix} 4{,}035 & 4{,}057 \\ 16{,}364 & 16{,}459 \end{vmatrix} = 0{,}025 \qquad \Delta_b = \begin{vmatrix} 1 & 4{,}035 \\ 4{,}057 & 16{,}364 \end{vmatrix} = \text{-}0{,}003$$

$$\alpha = \frac{\Delta_\alpha}{\Delta} = 7{,}606 \qquad b = \frac{\Delta_b}{\Delta} = -0{,}880 \qquad \ln a = \alpha \qquad a = e^\alpha = e^{7{,}606} = 2009$$

Рівняння регресії має вигляд

$$y = 2009 \cdot x^{-0{,}880}.$$

Знайдемо індекс детермінації

$$\rho_{xy} = \sqrt{1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}},$$

$$\rho_{xy} = \sqrt{1 - \frac{22{,}967}{64{,}599}} = 0{,}803$$

Тоді індекс детермінації $R^2 = \rho_{xy}^2 = 0{,}644$.

Обчислюємо критерій Фішера

$$F = \frac{R^2}{1-R^2}\frac{n-m}{m-1}$$

$$F = \frac{0{,}644}{1-0{,}644}\frac{8-2}{2-1} = 10{,}88$$

Знаходимо табличне значення критерія Фішера

$$F(0{,}05;1;6) = 5{,}99.$$

Оскільки $F = 10{,}88 > F(0{,}05;1;6) = 5{,}99$, то нульова гіпотеза відхиляється і рівняння статистично значуще.

<u>Показникова функція</u>

$$y = a \cdot b^x$$

Логарифмуємо рівняння $\ln y = \ln a + x \cdot \ln b$.

Вводимо нові позначення $\ln y = z$, $\ln a = \alpha$, $\ln b = \beta$

Тоді рівняння приймає вигляд

$$z = \alpha + \beta \cdot x$$

Для застосування МНК виписуємо систему нормальних рівнянь

$$\begin{cases} \alpha + \beta \cdot \bar{x} = \bar{z} \\ \alpha \cdot \bar{x} + \beta \cdot \overline{x^2} = \overline{zx} \end{cases}$$

Знаходження середніх зводимо до таблиці 8.

Таблиця 8

№	x_i	y_i	$z_i = \ln y_i$	x_i^2	$z_i \cdot x_i$	$y_{i\,расч.}$	$(y_i - y_{i\,расч})^2$	$(y_i - \bar{y})^2$
1	53	64	4,159	2809	220,421	61,204	7,819	20,964
2	55	59	4,078	3025	224,265	59,247	0,061	6,874
3	56	58	4,060	3136	227,385	58,292	0,085	2,779
4	57	55	4,007	3249	228,418	57,353	5,534	0,529
5	58	55	4,007	3364	232,425	56,428	2,040	0,039
6	59	56	4,025	3481	237,496	55,519	0,232	1,224
7	60	53	3,970	3600	238,218	54,624	2,637	4,004
8	65	53	3,970	4225	258,069	50,362	6,959	39,226
середнє	57,875		4,035	3361,125	233,337	56,543		
сума							25,368	75,639

Тоді система нормальних рівнянь запишеться у вигляді.

$$\begin{cases} \alpha + 57{,}875 \cdot \beta = 4{,}035 \\ 57{,}875 \cdot \alpha + 3361{,}125 \cdot \beta = 233{,}337 \end{cases}$$

Розв'язуємо систему по формулам Крамера.

$$\Delta = \begin{vmatrix} 1 & 57{,}875 \\ 57{,}875 & 3361{,}125 \end{vmatrix} = 11{,}609; \ \Delta_\alpha = \begin{vmatrix} 4{,}035 & 57{,}875 \\ 233{,}337 & 3361{,}125 \end{vmatrix} = 57{,}761; \ \Delta_\beta = \begin{vmatrix} 1 & 4{,}035 \\ 57{,}875 & 233{,}337 \end{vmatrix} = -0{,}189.$$

$$\alpha = \frac{\Delta_\alpha}{\Delta} = 4{,}975 \qquad \beta = \frac{\Delta_\beta}{\Delta} = -0{,}016$$

$$\ln a = \alpha \quad a = e^{\alpha} = 144{,}80 \qquad \ln b = \beta \quad b = e^{\beta} = 0{,}98$$

Рівняння регресії має вигляд $y = 144{,}8 \cdot 0{,}98^x$

Знайдемо індекс детермінації

$$\rho_{xy} = \sqrt{1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}},$$

$$\rho_{xy} = \sqrt{1 - \frac{25{,}368}{75{,}639}} = 0{,}815$$

Тоді індекс детермінації $R^2 = \rho_{xy}^2 = 0{,}665$

Обчислюємо критерій Фішера

$$F = \frac{R^2}{1 - R^2} \frac{n - m}{m - 1}$$

$$F = \frac{0{,}665}{1 - 0{,}665} \frac{8 - 2}{2 - 1} = 11{,}89$$

Знаходимо табличне значення

$$F(0{,}05;1;6) = 5{,}99.$$

Оскільки $F = 11{,}89 > F(0{,}05;1;6) = 5{,}99$, то нульова гіпотеза відхиляється і рівняння статистично значуще.

На рис. 7 приведені рівнянь регресій, знайдених МНК. Аналіз графіків, приведених на рис. 7, показує, що вони практично співпадають. Таким чином, можна зробити висновок, що має місце ідентичність вибраних структур функцій.

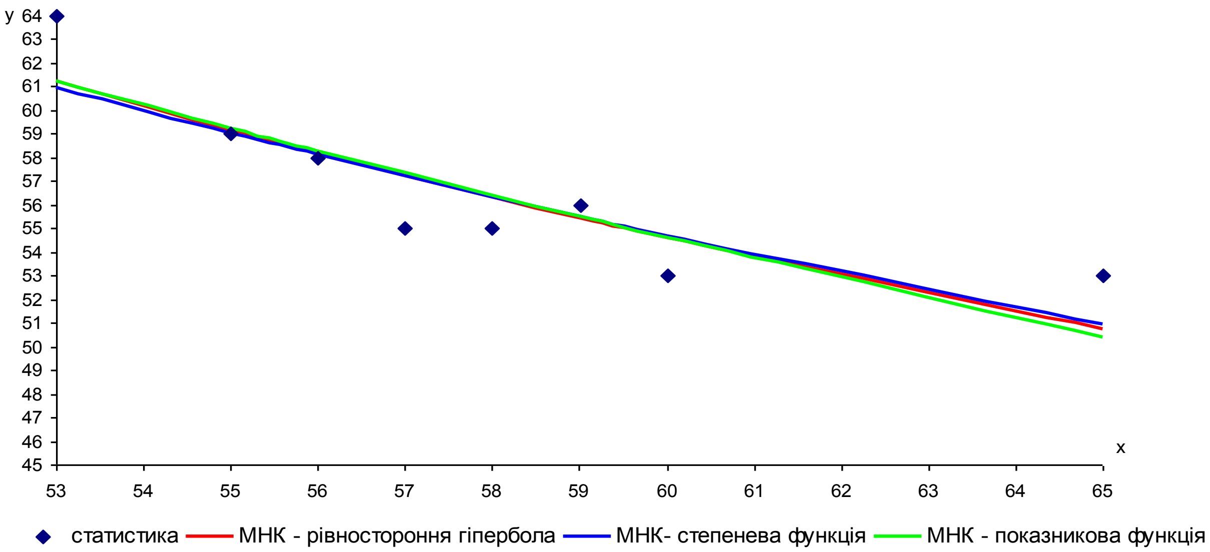

Рис. 7.
Графіки функцій, знайдених за допомогою МНК, які наближаютьстатистичні дані.

4. Багатофакторний лінійний кореляційно-регресійний аналіз

Багатофакторний кореляційно-регресійний аналіз є узагальненням парного кореляційно-регресійного аналізу на випадок, коли вихідна змінна зв'язана більш ніж з однією вхідною змінною.

У багатофакторному кореляційно-регресійному аналізі в порівнянні з парним кореляційно-регресійним аналізом виникають, принаймні, **дві** проблеми.

По-перше, необхідно вирішити проблему **специфікації**, тому що вихідна змінна Y залежить від багатьох вхідних змінних X_1, X_2, …, X_p і треба встановити, які з них є домінуючими і які необхідно включати в модель.

По-друге, при аналізі впливу конкретної вхідної змінної X_i на Y необхідно **відокремити** вплив на Y інших вхідних змінних.

Спочатку будемо вважати, що проблема специфікації вирішена і багатофакторна лінійна регресійна модель в загальному вигляді має вигляд

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + ... + a_p \cdot X_p,$$

де Y – вихідна змінна,

X_1, $X_2, \dots, X_p$ – вхідні змінні,

$a_0, a_1, a_2, ..., a_p$ – параметри,

p – кількість вхідних змінних.

Позначимо i -е спостереження вихідної змінної y_i , а відповідні значеннявхідних змінних x_{i1}, $x_{i2}, \ldots,$ x_{ip}. Тоді лінійна модель матиме вигляд

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + ... + a_p x_{ip}.$$

Модель, в якій вихідна змінна Y, вхідні змінні X_1, X_2, $\ldots, X_p$, а також залишки $u_j = \hat{y}_j - y_j$ задовольняють умовам Гаусса-Маркова і, крім того, вхідні змінні X_1, X_2, $\ldots, X_p$ є незалежними, називається **класичною нормальною лінійною моделлю** багатофакторної регресії (Classic Normal Linear Multiple Regression Model).

В випадку багатофакторної лінійної регресії зручно всі викладки проводити в матричному вигляді. Для цього введемо позначення

$$X = \begin{pmatrix} 1 & x_{11} & . & x_{1p} \\ 1 & x_{21} & . & x_{2p} \\ . & . & . & . \\ 1 & x_{n1} & . & x_{np} \end{pmatrix}$$ – матриця вхідних змінних розміром $n \times (p+1)$,

де $n-$ кількість спостережень, $p-$кількість вхідних змінних, (у матриці X введено додатковий стовпчик із одиниць, чим враховується наявність в моделі вільного члена a_0),

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ y_n \end{pmatrix}$$ – вектор-стовбець вихідної змінної розмірності ($n \times 1$),

$$A = \begin{pmatrix} a_0 \\ a_1 \\ . \\ a_p \end{pmatrix}$$ – вектор-стовбець параметрів моделі розмірності ($(p+1) \times 1$).

Тоді модель в матричному вигляді запишеться так

$$Y = X \cdot A.$$

Розв’яжемо це рівняння, вважаючи, що матриця X є прямокутною, але число строк не більше числа стовбців ($n \geq p+1$).
Домножимо матричне рівняння зліва на транспоновану матрицю X^T і скористаємось асоціативним законом

$$(X^T X) \cdot A = X^T Y.$$

Матриця в дужках є квадратною, тому для неї може існувати обернена $\left(X^T X\right)^{-1}$, якщо вона є неособливою.

Домножаємо останне рівняння зліва на обернену матрицю

$$\left(X^T X\right)^{-1}(X^T X)\cdot A = \left(X^T X\right)^{-1} X^T Y .$$

Враховуючи, що

$$\left(X^T X\right)^{-1}(X^T X) = E ,$$

маємо

$$A = \left(X^T X\right)^{-1} X^T Y$$

Підкреслимо, що визначник $\left|X^T X\right| \neq 0$, тобто $(X^T \cdot X)^{-1}$ існує.

При цьому A називається **оператор оцінювання**.

Останнє рівняння (оператор оцінювання) представляє собою матричну форму запису розв'язку системи нормальних рівнянь для знаходження параметрів моделі.

Коефіцієнти детермінації і кореляції.

Як і у випадку парного рівняння регресії, в багатофакторному рівнянні регресії загальна варіація SST (сума квадратів відхилень вихідної змінної від середнього значення) може бути розкладена на дві складові

$$SST = SSR + SSE .$$

Значення SST, SSR, SSE можуть бути обчислені за формулами

$$SST = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \hat{Y}^T \hat{Y} - \bar{Y}^T \bar{Y} ,$$

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \hat{Y}^T \hat{Y} - A^T X^T \hat{Y} ,$$

$$SSR = \sum_{i=1}^{n} (y_i - \bar{y})^2 = A^T X^T \hat{Y} - \bar{Y}^T \bar{Y} ,$$

де $\bar{Y} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \cdot \\ \bar{y} \end{pmatrix}$, $\hat{Y} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \cdot \\ \hat{y}_n \end{pmatrix}$ – вектор-стовбець спостережень вихідної змінної,

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i$.

Коефіцієнт детермінації R^2 дає одну з найбільш ефективних оцінок адекватності регресійної моделі. Він є мірою якості моделі і її прогностичної сили. Даний коефіцієнт показує, яка частина дисперсії вихідної змінної пояснюється регресією. Коефіцієнт детермінації знаходиться за формулою

$$R^2 = \frac{SSR}{SST}.$$

Коефіцієнт детермінації змінюється в межах $0 < R^2 < 1$. Чим ближче R^2 до одиниці, тим краще регресія описує залежність між вхідними і вихідними змінними. Іноді обчислюють скоректований (адаптований, нормований, виправлений) коефіцієнт детермінації $\hat{R}^2$, який враховує число ступенів свободи і обчислюється за формулою

$$\hat{R}^2 = 1 - \frac{n-1}{n-m}(1 - R^2).$$

Коефіцієнт детермінації R^2, обчислений без врахування числа ступенів свободи буде збільшуватися при введені в модель нових пояснюючих змінних, хоча це і не завжди буде означати покращення моделі. Скоректований коефіцієнт детермінації $\hat{R}^2$ не має такого недоліку. Він може навіть зменшуватися при введені в модель нових пояснюючих змінних, якщо вони не дають суттєвого покращення якості моделі.

Зауважимо, що збільшення коефіцієнта детермінації не завжди означає покращення якості моделі. Велике значення коефіцієнта детермінації може бути наслідком того, що досліджувані показники мають часовий тренд.

Коефіцієнт багатофакторної кореляції (або теоретичне кореляційне співвідношення, **індекс кореляції** Y по X) дорівнює квадратному кореню із коефіцієнту детермінації, тобто

$$R = \sqrt{R^2}.$$

Для оцінки тісноти зв'язку між показниками, що входять до моделі розраховуються **коефіцієнти парної кореляції**. Розрахунок коефіцієнтів парної кореляції може бути проведено за формулами

$$r_{y/x_j} = \frac{\overline{yx_j} - \bar{y} \cdot \bar{x}_j}{s_y \cdot s_{x_j}}, \quad r_{x_k/x_j} = \frac{\overline{x_k x_j} - \bar{x}_k \cdot \bar{x}_j}{s_{x_k} \cdot s_{x_j}}.$$

Для характеристики зв'язку між ознаками розраховують також коефіцієнти еластичності. Для лінійної багатофакторної моделі вони обчислюються за формулою аналогічною як для парної лінійної регресії

$$E_{y/x_j} = \frac{\partial y}{\partial x_j} \frac{\bar{x}_j}{\bar{\hat{y}}} = a_j \frac{\bar{x}_j}{\bar{\hat{y}}}.$$

Коефіцієнти еластичності в даному випадку мають такий самий зміст, як і для парної лінійної регресії: E_{y/x_j} показує, на скільки відсотків в середньому зміниться вихідна змінна Y при зміні вхідної змінної X_j в середньому на один відсоток при інших незмінних умовах.

Перевірка значущості моделі здійснюється за допомогою F **-критерія Фішера**. Якщо коефіцієнт детермінації відомий, то значення F - статистики критерія Фішера розраховується за формулою

$$F = \frac{R^2}{1-R^2}\frac{n-m}{m-1}.$$

Отримане значення F - критерія порівнюється з табличним для рівня значущості α і числа ступенів свободи чисельника $k_1 = m-1$ і $k_2 = n-m$ – числа ступенів свободи знаменника, n – число спостережень, $m = p+1$– число параметрів моделі, p– число вхідних змінних.

Якщо

$$F > F(\alpha; k_1; k_2),$$

то рівняння вважається статистично значущим.

Перевірка значущості параметрів рівняння регресії та коефіцієнта кореляції

У окремих випадках багатофакторної регресії можливі ситуації, коли рівняння регресії в цілому є статистично значущим за F - критерієм, але параметри його не є статистично значущими. Тому виникає необхідність окремо перевірити статистичну значущість параметрів рівняння.

Перевірка значущості параметрів рівняння регресії здійснюється за допомогою t–критерія Стьюдента.

Для оцінки значущості параметрів регресійної моделі і обчислення довірчихм інтервалів обчислюється дисперсійно-коваріаційна матриця параметрів моделі

$$\operatorname{cov}(A) = s_u^2 \left(X^T X\right)^{-1},$$

де s_u^2 –незміщена оцінка дисперсії залишків.

$$s_u^2 = \frac{\hat{Y}\hat{Y}^T - A^T X^T X}{n-m}.$$

Якщо позначити елементи матриці

$$\left(X^T X\right)^{-1} = (c_{ij}), \quad i = 0,1,..., p, \; j = 0,1,..., p,$$

то дисперсійно-коваріаційна матриця параметрів моделі матиме вигляд

$$\operatorname{cov}(A)=s_u^2\begin{pmatrix} c_{00} & c_{01} & . & c_{0p} \\ c_{10} & c_{11} & . & c_{1p} \\ . & . & . & . \\ c_{p0} & c_{p1} & . & c_{pp} \end{pmatrix}.$$

Обчислення значення t - критерію Стьюдента для кожного параметра a_j здійснюється за формулою

$$t_j=\frac{|a_j|}{\sqrt{s_u^2\cdot c_{jj}}}.$$

Знаменник цього відношення

$$\sqrt{s_u^2\cdot c_{jj}}$$

називається стандартною похибкою оцінки параметра a_j моделі.

Обчислене значення t - критерію порівнюється з табличним при обраному рівні значущості α і $n-m$ ступенях свободи.

Якщо

$$t>t_{кр}(\alpha,n-m),$$

то відповідна оцінка параметра економетричної моделі є статистично значущою.

На основі t- критерію і стандартної похибки будуються довірчі інтервали для параметрів $\hat{a}_j$

$$(a_j\pm t_{кр}(\alpha,n-m)\cdot\sqrt{s_u^2\cdot c_{jj}}).$$

Приведена формула аналогічна відповідній формулі для парного рівняння регресії.

Прогноз вихідної змінної

Економетричне моделювання зв'язку між економічними показниками складається, як зазначалося, з кількох етапів, одним з яких є прогнозування на основі моделі.

Точкове значення прогнозу отримаємо, якщо в модель

$$Y=X\cdot A,$$

підставимо очікувані значення вхідних змінних $\tilde{x}_1$, $\tilde{x}_2$, …, $\tilde{x}_p$.

В результаті отримуємо

$$\tilde{Y}=\tilde{X}\cdot A,$$

де $\tilde{X}^T=(1,\tilde{x}_1,\tilde{x}_2,...,\tilde{x}_p)$, $A^T=(a_0,a_1,...,a_p)$.

Тоді

$$\tilde{y}=a_0+a_1\cdot\tilde{x}_1+a_2\cdot\tilde{x}_2+...+a_p\cdot\tilde{x}_p.$$

Щоб знайти інтервальний прогноз необхідно знайти середню квадратичну похибку прогнозу, або, що те саме, середню квадратичну похибку $\tilde{Y}$.

У матричному вигляді дисперсія похибки прогнозу групової середньої має вигляд

$$s_{пр}^2 = s_u^2 \tilde{X}^T \left(X^T X\right)^{-1} \tilde{X} .$$

Довірчий інтервал для математичного сподівання $M[Y]$ прогнозних значень знаходиться по формулі

$$\tilde{y} \pm t_{кр}(\alpha, n-m) \cdot s_{пр} .$$

Для визначення інтервального прогнозу індивідуального значення $\tilde{y}$ необхідно спочатку знайти відповідну стандартну похибку (середнє квадратичне відхилення прогнозу). Дисперсія індивідуального значення обчислюється за формулою

$$s_{пр(i)}^2 = s_u^2 + s_u^2 \tilde{X}^T \left(X^T X\right)^{-1} \tilde{X} ,$$

або

$$s_{пр(i)}^2 = s_u^2 (1 + \tilde{X}^T \left(X^T X\right)^{-1} \tilde{X}) .$$

Отже, інтервальний прогноз **індивідуального** значення визначається так

$$\left(\tilde{y} \pm s_u \sqrt{1 + \tilde{X}^T (X^T X)^{-1} \tilde{X}}\right).$$

Приклад. Знаходження двофакторної моделі лінійної регресії.

Побудувати двофакторну модель лінійної регресії залежності заробітної плати Y від рівня рентабельності X_1 та затрат капіталу X_2 для деякого підприємства. Необхідні дані наведено в таблиці.

Таблиця

Y, *ум. од.*	63	70	80	84	69	72	70
X_1,%	10,7	11	12,2	12,4	10,9	11,3	11,1
X_2, *ум. од*	38	29	30	25	24	31	30

За допомогою статистичних критеріїв перевірити модель на адекватність. Зробити прогноз значення Y для X_1 =15 %, X_2 =40 ум. од.

Розв'язання.

1. Лінійна двофакторна модель має вигляд

$$Y = a_0 + a_1 X_1 + a_2 X_2$$

Система нормальних рівнянь має вигляд

$$\begin{cases} a_0 + a_1 \bar{x}_1 + a_2 \bar{x}_2 = \bar{y} \\ a_0 \bar{x}_1 + a_1 \overline{x_1^2} + a_2 \overline{x_1 x_2} = \overline{x_1 y} \\ a_0 \bar{x}_2 + a_1 \overline{x_1 x_2} + a_2 \overline{x_2^2} = \overline{x_2 y} \end{cases}$$

Подальші обчислення зводимо в таблицю.

№	x_1	x_2	y	x_1^2	x_2^2	$x_1 \cdot x_2$	$x_1 \cdot y$	$x_2 \cdot y$	y^2	y_r	$(y-\bar{y})^2$	$(y_r-\bar{y})^2$
1	10,7	38	63	114,49	1444	406,6	674,1	2394	3969	63,49	91,58	82,54
2	11	29	70	121	841	319	770	2030	4900	69,17	6,60	11,56
3	12,2	30	80	148,84	900	366	976	2400	6400	80,44	55,20	61,94
4	12,4	25	84	153,76	625	310	1041,6	2100	7056	83,92	130,64	128,82
5	10,9	24	69	118,81	576	261,6	752,1	1656	4761	69,76	12,74	7,92
6	11,3	31	72	127,69	961	350,3	813,6	2232	5184	71,45	0,32	1,27
7	11,1	30	70	123,21	900	333	777	2100	4900	69,83	6,60	7,54
Середні	11,37	29,57	72,57	129,69	892,43	335,21	829,20	2130,29	5310,00			
Сума											303,71	301,58

Тоді система нормальних рівнянь приймає вигляд.

$$\begin{cases} a_0 + 11{,}37a_1 + 29{,}57a_2 = 72{,}57 \\ 11{,}37a_0 + 129{,}69a_1 + 335{,}21a_2 = 829{,}2 \\ 29{,}57a_0 + 335{,}21a_1 + 892{,}43a_2 = 2130{,}29 \end{cases}$$

Для її розв’язання скористаємось формулами Крамера.

$$\Delta = \begin{vmatrix} 1 & 11{,}37 & 29{,}57 \\ 11{,}37 & 129{,}69 & 335{,}21 \\ 29{,}57 & 335{,}21 & 892{,}43 \end{vmatrix} = 5{,}65 \qquad \Delta_0 = \begin{vmatrix} 72{,}57 & 11{,}37 & 29{,}57 \\ 829{,}2 & 129{,}69 & 335{,}21 \\ 2130{,}29 & 335{,}21 & 892{,}43 \end{vmatrix} = -158{,}03$$

$$\Delta_1 = \begin{vmatrix} 1 & 72{,}57 & 29{,}57 \\ 11{,}37 & 829{,}2 & 335{,}21 \\ 29{,}57 & 2130{,}29 & 892{,}43 \end{vmatrix} = 54{,}48 \qquad \Delta_2 = \begin{vmatrix} 1 & 11{,}37 & 72{,}57 \\ 11{,}37 & 129{,}69 & 829{,}2 \\ 29{,}57 & 335{,}21 & 2130{,}29 \end{vmatrix} = -1{,}75$$

$$a_0 = \frac{-158{,}03}{5{,}65} = -27{,}99 \qquad a_1 = \frac{54{,}48}{5{,}65} = 9{,}65 \qquad a_2 = -\frac{1{,}75}{6{,}45} = -0{,}31$$

Отже, економетрична модель багатофакторної регресії для заробітної плати запишеться так

$$Y = -27{,}99 + 9{,}65 \cdot X_1 - 0{,}31 \cdot X_2,$$

або для кожного значення

$$y_i = -27{,}99 + 9{,}65 \cdot x_{1i} - 0{,}31 \cdot x_{2i}, \ \ (i = 1,...,n).$$

Виходячи з побудованої моделі, можна зробити такі попередні висновки: якщо незалежна змінна X_1 (рівень рентабельності) збільшується на одну одиницю, то Y – заробітна плата збільшується на 9,65 одиниць, якщо змінна X_2 – затрати капіталу зростають на одну одиницю, то заробітна плата зменшується на 0,31 одиниць.

Обчислимо коефіцієнт детермінації R^2, для цього скористаємось формулою

$$R^2 = \frac{SSR}{SST},$$

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

Згідно даних таблиці, маємо

$$R^2 = \frac{301{,}58}{303{,}71} = 0{,}993$$

Коефіцієнт детермінації показує, що на 99,3 % значення заробітної плати визначається факторами затратами капіталу та рівнем рентабельності і на 0,7 % визначається іншими факторами.

Коефіцієнт багатофакторної кореляції 0,993 показує, що зв’язок між змінними X та Y достатньо тісний.

Визначимо значущість рівняння регресії за допомогою F - критерію Фішера

$$F = \frac{R^2}{1-R^2}\frac{n-m}{m-1} \qquad F = \frac{0{,}993}{1-0{,}993}\frac{7-3}{3-1} = 265{,}71$$

Обчислене фактичне значення критерію Фішера порівнюємо з табличним $F(\alpha; k_1; k_2)$ при ступенях свободи чисельника $k_1 = n - m = 7 - 3 = 4$ і знаменника $k_2 = m - 1 = 3 - 1 = 2$ і прийнятому рівні значущості $\alpha = 0{,}05$ – $F(0{,}05;4;2) = 19{,}2$.
Оскільки $F = 265{,}71 > F(0{,}05;4;2) = 19{,}2$, то отримане рівняння регресії є статистично значущим. Це означає, що отримана модель може використовуватися для прогнозів, якщо оцінки параметрів також виявляться статистично значущими.

Для обчислення значущості оцінок параметрів моделі a_0, a_1 і a_2 скористаємось t – критерієм Стьюдента.

Дисперсія залишків обчислюється за формулою

$$s_u^2 = \frac{SSE}{n-m} \qquad SSE = SST - SSR = 303{,}71 - 301{,}58 = 2{,}13$$

$$s_u^2 = \frac{2{,}13}{7-3} = 0{,}53.$$

Для знаходження величини t – критерія Стьюдента необхідно обчислити величину

$$s_u \sqrt{c_{jj}} \,,$$

де c_{jj} – діагональні елементи матриці $\left(X^T X\right)^{-1} = (c_{ij})$.

Враховуючи, що

$$X = \begin{pmatrix} 1 & 10{,}7 & 38 \\ 1 & 11 & 29 \\ 1 & 12{,}2 & 30 \\ 1 & 12{,}4 & 25 \\ 1 & 10{,}9 & 24 \\ 1 & 11{,}3 & 31 \\ 1 & 11{,}1 & 30 \end{pmatrix},$$

послідовно знаходимо

$$X^T X = \begin{pmatrix} 7 & 79{,}6 & 207 \\ 79{,}6 & 907{,}8 & 2346{,}5 \\ 207 & 2346{,}5 & 6247 \end{pmatrix} \qquad (X^T X)^{-1} = \begin{pmatrix} 85{,}19 & -5{,}96 & -0{,}59 \\ -5{,}96 & 0{,}45 & 0{,}03 \\ -{,}0{,}59 & 0{,}03 & 0{,}01 \end{pmatrix}.$$

Таким чином,

$$c_{00} = 85{,}19 \qquad c_{11} = 0{,}45 \qquad c_{22} = 0{,}01.$$

Тоді величини t – критерія Стьюдента мають відповідні значення

$$t_0 = \frac{|a_0|}{\sqrt{s_u^2 \cdot c_{00}}} = \frac{27{,}99}{\sqrt{0{,}53 \cdot 85{,}19}} = 4{,}17$$

$$t_1 = \frac{|a_1|}{\sqrt{s_u^2 \cdot c_{11}}} = \frac{9{,}65}{\sqrt{0{,}53 \cdot 0{,}45}} = 19{,}76$$

$$t_2 = \frac{|a_2|}{\sqrt{s_u^2 \cdot c_{22}}} = \frac{0{,}31}{\sqrt{0{,}53 \cdot 0{,}01}} = 4{,}26$$

Знаходимо табличне значення t -критерію для рівня значущості $\alpha = 0{,}05$ і числа ступенів свободи $k = n - m = 7 - 3 = 4$, що дає

$$t(0{,}05;4) = 2{,}13.$$

Оскільки $t_0 = 4{,}17 > t(0{,}05;4) = 2{,}13$, $t_1 = 19{,}76 > t(0{,}05;4) = 2{,}13$, $t_2 = 4{,}6 > t(0{,}05;4) = 2{,}13$, то отримані значення $a_0 = -27{,}99$, $a_1 = 9{,}65$, $a_2 = -0{,}31$. Іншими словами, гіпотеза H_0 про те, що $a_0 = 0$, $a_1 = 0$, $a_2 = 0$ відкидається. Такий результат перевірки значущості параметрів рівняння означає, що побудована модель регресії може використовуватися для аналізу і прогнозу економічного процесу.

Побудуємо точковий та інтервальний прогнози для значень вихідної змінної при $x_{10} = 15\%$ і $x_{20} = 40$ ум. од.

$$y_{пр} = -27{,}99 + 9{,}65 \cdot 15 - 0{,}31 \cdot 40 = 104{,}36$$

Знайдемо дисперсію похибки прогнозу групової середньої, скориставшись формулою

$$s_{пр}^2 = s_u^2 X_0^T (X^T X)^{-1} X_0,$$

де $X_0^T = \begin{pmatrix} 1 & 15 & 40 \end{pmatrix}$.

Для цього спочатку обчислимо

$$X_0^T (X^T X)^{-1} = \begin{pmatrix} 1 & 15 & 40 \end{pmatrix} \begin{pmatrix} 85{,}19 & -5{,}96 & -0{,}59 \\ -5{,}96 & 0{,}45 & 0{,}03 \\ -0{,}59 & 0{,}03 & 0{,}01 \end{pmatrix} = \begin{pmatrix} -27{,}58 & 1{,}93 & 0{,}20 \end{pmatrix}$$

$$X_0^T (X^T X)^{-1} X_0 = \begin{pmatrix} -27{,}58 & 1{,}93 & 0{,}20 \end{pmatrix} \begin{pmatrix} 1 \\ 15 \\ 40 \end{pmatrix} = 9{,}18.$$

Тоді

$$s_{пр}^2 = 0{,}53 \cdot 9{,}18 = 4{,}87, \qquad s_{пр} = 2{,}21.$$

Знаходимо табличне значення t -критерію для рівня значущості $\alpha = 0{,}05$ і числа ступенів свободи $k = n - m = 7 - 3 = 4$, що дає

$$t(0{,}05;4) = 2{,}13.$$

Тоді довірчий інтервал для математичного сподівання знаходимо по формулі

$$\tilde{y}_{пр} \pm t_{кр}(\alpha, n-m) \cdot s_{пр},$$

$$104{,}36 \pm 2{,}13 \cdot 2{,}21, \text{ тобто } 104{,}36 \pm 4{,}70,$$

$$(99{,}66; 109{,}06)$$

Останнє означає, що з надійністю 0,95 середня заробітна плата при рентабельності 15% і затратах капіталу 40 ум. од. буде знаходитися в межах від 99,66 до 109,06.

Знайдемо довірчий інтервал для індивідуальних значень $y_{пр}$. Спочатку обчислюємо

$$s_{пр(i)}^2 = s_u^2(1 + 9{,}18) = 0{,}53 \cdot 10{,}18 = 5{,}40. \quad s_{пр(i)} = 2{,}32$$

Тоді

$$\tilde{y}_{пр(i)} \pm t_{кр}(\alpha, n-m) \cdot s_{пр(i)},$$

$$104{,}36 \pm 2{,}13 \cdot 2{,}32, \text{ тобто } 104{,}36 \pm 4{,}95,$$

$$(99{,}41; 109{,}31).$$

З надійністю 0,95 індивідуальне значення заробітної плати при затратах капіталу 40 ум. од. і рентабельності 15% буде знаходитися в межах від 99,41 до 109,31 ум. од. при незмінності умов проведення спостереження.

5. Багатофакторний нелінійний кореляційно-регресійний аналіз

Як і в парній залежності, використовуються **нелінійні** види рівнянь множинної регресії. На першому етапі доцільно виділити лінеаризуємі рівняння множинної регресії, тобто такі рівняння, які шляхом алгебраїчних перетворень або відповідних замін можливо звести до лінійних рівнянь.

Зважаючи на чітку інтерпретацію параметрів, найширше серед нелінійних рівнянь множинної регресії, які лінеарізуються, використовується степенева функція. Цей вид рівняння регресії отримав найбільше поширення у виробничих функціях, в дослідженнях попиту і споживання.

У степеневій функції, яка має вигляд

$$y = a_0 x_1^{a_1} \cdot \ldots \cdot x_p^{a_p},$$

коефіцієнти a_k є коефіцієнтами еластичності. Вони показують, на скільки відсотків в середньому змінюється результат зі зміною відповідного фактора на 1 % при незмінності дії інших факторів.

Дійсно, згідно означенню коефіцієнта еластичності маємо

$$E_k = \frac{\partial y}{\partial x_k}\frac{x_k}{y} .$$

Для спрощення обчислень логарифмуємо степеневу функцію

$$\ln y = \ln a_0 + a_1 \ln x_1 + \ldots + a_k \ln x_k + \ldots + a_p \ln x_p .$$

Тоді

$$\frac{\partial \ln y}{\partial x_k} = a_k \frac{\partial \ln x_k}{\partial x_k},$$

або, обчислюючи частинні похідні, маємо

$$\frac{1}{y}\frac{\partial y}{\partial x_k} = a_k \frac{1}{x_k},$$

тобто

$$\frac{\partial y}{\partial x_k}\frac{x_k}{y} = a_k .$$

Таким чином, маємо

$$E_k = a_k$$

Враховуючи, що загальна еластичність знаходиться по формулі

$$E = E_1 + \ldots + E_k + \ldots + E_p ,$$

маємо для нашого випадку

$$E = a_1 + \ldots + a_k + \ldots + a_p$$

Ця величина фіксує узагальнену характеристику еластичності виробництва.

Можливі і інші лінеаризуємі функції для побудови рівняння множинної регресії:

- експонента $y = e^{a_0 + a_1 x_1 + \ldots + a_p x_p + u}$
- гіпербола $y = \dfrac{1}{a_0 + a_1 x_1 + \ldots + a_p x_p + u}$,

яка використовується при зворотних зв'язках ознак.

Так, степенева функція лінеарізується шляхом її логарифмування

$$\ln y = \ln a_0 + a_1 \ln x_1 + \ldots + a_k \ln x_k + \ldots + a_p \ln x_p + \ln u$$

з послідуючою заміною змінних

$$z = b_0 + a_1 u_1 + \ldots + a_k u_k + \ldots + a_p u_p + w,$$

де $z = \ln y$, $u_1 = \ln x_1, \ldots, u_k = \ln x_k, \ldots,$ $u_p = \ln x_p$, $w = \ln u$, $b_0 = \ln a_0$.

Експонента лінеарізується також шляхом логарифмування

$$z = a_0 + a_1 x_1 + \ldots + a_k x_k + \ldots + a_p x_p + u,$$

де $z = \ln y$.

Гіпербола лінеарізується шляхом заміни вихідної змінної

$$z = a_0 + a_1 x_1 + \ldots + a_k x_k + \ldots + a_p x_p + u,$$

де $z = \dfrac{1}{y}$.

Якщо дослідника не влаштовує пропонований набір функцій регресії, то можна використовувати будь-які інші функції, що наводяться шляхом відповідних перетворень до лінійного вигляду, наприклад,

$$y = a_0 + a_1 x_1 + a_2 \frac{1}{x_2} + a_3 \sqrt{x_3} + a_4 \ln x_4 .$$

Позначивши

$$x_1 = z_1, \; \frac{1}{x_2} = z_2, \; \sqrt{x_3} = z_3, \; \ln x_4 = z_4,$$

отримаємо лінійне рівняння множинної регресії

$$y = a_0 + a_1 z_1 + a_2 z_2 + a_3 z_3 + a_4 z_4 .$$

Стандартні комп'ютерні програми обробки регресійного аналізу дозволяють перебирати різні функції і вибрати ту з них, для якої залишкова дисперсія і помилка апроксимації мінімальні, а коефіцієнт детермінації максимальний.

При нелінійної залежності ознак, що приводиться до лінійного вигляду, параметри множинної регресії також визначаються МНК з тією лише різницею, що він використовується не до вихідної інформації, а до перетворених даних.

Так, розглядаючи ступеневу функцію

$$y = a_0 x_1^{a_1} \cdot \ldots \cdot x_p^{a_p} \cdot u$$

ми перетворюємо її, як показано вище, в лінійний вид

$$z = b_0 + a_1 u_1 + \ldots + a_k u_k + \ldots + a_p u_p + w.$$

Далі обробка МНК та ж, що і описана вище: будується система нормальних рівнянь і визначаються параметри b_0, $a_1, \ldots, a_k, \ldots, a_p$. Потенціюючи значення b_0, знайдемо параметр a_0 й відповідно загальний вигляд рівняння степеневої функції.

При інших нелінійних функціях методика оцінки параметрів МНК проводиться так само. На відміну від попередніх функцій параметри більш складних моделей не мають чіткої економічної інтерпретації: вони не є показниками сили зв'язку і її еластичності. Це не виключає можливості їх застосування, але робить їх менш привабливими в практичних розрахунках.

6. Мультиколінеарність

Мультиколінеарність – це існування лінійної залежності, або кореляції між двома чи більше вхідними змінними.

Ясно, що мультиколінеарність може проявлятися, коли вивчається залежність вихідної змінної Y від кількох вхідних змінних X.

Якщо користуватися МНК при наявності мультиколінеарності, то це веде до зниження практичної цінності моделі.

Основні наслідки мультиколінеарності такі.

1. Стандартні похибки параметрів регресії стають великими, в результаті чого оцінка їх значущості за t - критерієм Стьюдента не має змісту. Хоча в цілому за F - критерієм Фішера модель може бути визнана статистично значущою.
2. Оцінки параметрів стають дуже чутливими до обсягу вибірки.
3. Оцінки параметрів стають також дуже чутливими до незначної зміни результатів спостереження.
4. Рівняння регресії не мають реального змісту, тому що деякі з його коефіцієнтів мають неправильні з економічних міркувань значення.
5. Проводячи економетричні дослідження необхідно знати, чи має місце мультиколінеарність між пояснюючими змінними. Виявити мультиколінеарність можна за багатьма ознаками. Найчастіше в цих цілях використовують алгоритм Фаррара – Глобера.

Алгоритм Фаррара – Глобера

За допомогою цього алгоритму виявляється наявність мультиколінеарності

- всього масиву вхідних змінних,
- кожної вхідної змінної з рештою вхідних змінних,
- кожної пари вхідних змінних.

Для цього застосовується три види критеріїв.

Перший критерій χ^2 (Пірсона) використовується для виявлення мультиколінеарності всього масиву вхідних змінних.

Другий, F – критерій Фішера використовується для виявлення зв'язку конкретної вхідної змінної з рештою вхідних змінних.

Третій, t – критерій Стьюдента, для виявлення кореляції між двома вхідними змінними.

Алгоритм Фаррара - Глобера розкладається на вісім кроків.

Крок 1-й. Стандартизація пояснюючих змінних.

Для кожного спостереження всіх пояснюючих змінних обчислюють стандартизовані значення ознак за формулами

$$x^*_{jk} = \frac{x_{jk} - \bar{x}_k}{s_k},$$

де $j = 1,...,n$ – число спостережень, $k = 1,..., p$ – число вхідних змінних,

$\bar{x}_k$ – середнє арифметичне вхідної змінної x_k,

s_k – стандартне відхилення вхідної змінної x_k.

В результаті отримують вектори стандартизованих вхідних змінних, які утворюють матрицю

$$X^* = \begin{pmatrix} x^*_{11} & x^*_{21} & \dots & x^*_{p1} \\ x^*_{12} & x^*_{22} & \dots & x^*_{p2} \\ . & . & \dots & . \\ x^*_{1p} & x^*_{2p} & \dots & x^*_{np} \end{pmatrix}$$

Крок 2-й. Знаходять кореляційну матрицю для вхідних змінних за формулою.

$$\mathbf{r} = \frac{1}{n} X^{*T} \cdot X^*.$$

де X^{*T} – матриця транспонована по відношенню до матриці X^*.

Матриця $\mathbf{r}$ буде симетричною, і на головній діагоналі будуть стоять одиниці. Елементами матриці $\mathbf{r}$ є коефіцієнти парної кореляції відповідних вхідних змінних. Елементи матриці $\mathbf{r}$ за модулем не перевищують одиниці. Матриця $\mathbf{r}$ є квадратною розміру p.

Крок 3-й. Обчислюють визначник матриці $\mathbf{r}$. За отриманим значенням можна зробити попередній висновок про наявність мультиколінеарності. Якщо визначник дорівнює **нулю**, то існує **повна** мультиколінеарність.

Мультиколінеарність **присутня**, якщо визначник **близький до нуля**.

Крок 4-й. Обчислюють χ^2 – статистику за формулою

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2p + 5)\right]\ln(|r|),$$

де $|\mathbf{r}|$ – визначник кореляційної матриці r.
χ^2 – статистика служить для перевірки **нульової гіпотези** H_0 **:**
мультиколінеарність присутня.

Значення цього критерію порівнюється з табличним $\chi^2(\alpha; k)$ при $k = \frac{1}{2}p(p-1)$ – ступенях свободи і рівні значущості α.

Якщо $\chi^2 > \chi^2(\alpha; k)$, то нульова гіпотеза **підтверджується**, тобто в масиві пояснюючих змінних **існує мультиколінеарність** і перевірку необхідно продовжити.

Якщо $\chi^2 < \chi^2(\alpha; k)$, то серед пояснюючих змінних мультиколінеарність **відсутня** або говорять, що мультиколінеарність знаходиться в допустимих межах, і на цьому перевірку на мультиколінеарність можна завершити.

Крок 5-й. Обчислюється матриця $C = \mathbf{r}^{-1}$ обернена до матриці $\mathbf{r}$.

$$C = (X^{*T} \cdot X^*)^{-1} = \frac{1}{|r|}\begin{pmatrix} A_{11} & A_{21} & \ldots & A_{p1} \\ A_{12} & A_{22} & \ldots & A_{p2} \\ . & . & \ldots & . \\ A_{1p} & A_{2p} & \ldots & A_{pp} \end{pmatrix},$$

де A_{ij} – алгебраїчні доповнення до елементу r_{ij} матриці r.

Крок 6-й. Обчислюємо F – критерій за формулою

$$F_k = \left(\frac{A_{kk}}{|r|} - 1\right)\frac{n - p}{p - 1}$$

Таким чином, отримаємо стільки фактичних значень F_k, скільки є вхідних змінних. Отримані фактичні значення F_k порівнюються з табличними $F(\alpha; k_1; k_2)$ при $k_1 = n - p$ ступенях свободи чисельника, $k_2 = p - 1$ ступенях свободи знаменника і рівні значущості α. . Якщо $F_k > F(\alpha; k_1; k_2)$, то відповідна вхідна змінна X_k корелює з іншими вхідними змінними.
Якщо $F_k < F(\alpha; k_1; k_2)$, то вхідна змінна X_k не корелює з іншими вхідними змінними.

Крок 7-й. Обчислюються частинні коефіцієнти кореляції

$$r_{k/j} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}.$$

Частинні коефіцієнти кореляції характеризують тісноту зв'язку між двома вхідними змінними X_k і X_j при умові, що вся решта вхідних змінних не впливає на ці дві вхідні змінні.

Крок 8-й. Розраховується $t-$критерій за формулою

$$t_{ij} = \frac{r_{ij}\sqrt{n-p}}{\sqrt{1-r_{ij}^2}}$$

для кожної пари вхідних змінних X_i і X_j.

Отримані фактичні значення $t-$статистики порівнюють з табличним $t(\alpha,k)$ при $k=n-p$ числі ступенів свободи і рівні значущості α.

Якщо $|t_{ij}| > t(\alpha,k)$, то це означає, що між вхідними змінними X_i і X_j існує залежність. Якщо $|t_{ij}| < t(\alpha,k)$, то це означає, що між вхідними змінними X_i і X_j немає залежності.

Для зменшення негативного впливу мультиколінеарності можуть застосовуватися багато різних методів. Один із них полягає у виключенні однієї із пояснюючих змінних із розгляду в моделі. Із двох вхідних змінних X_i і X_j, які мають високий коефіцієнт кореляції і для яких $t_{ij} > t(\alpha;k)$ одну виключать із розгляду. При цьому, яку змінну залишити в моделі, а яку виключити, вирішують, виходячи в першу чергу із економічних міркувань. Якщо із економічних міркувань ні одній із них не можна надати перевагу, то залишають ту із змінних, яка має більший коефіцієнт кореляції із вихідною змінною Y, або ту, яка має слабшу кореляцію з масивом інших пояснюючих змінних, тобто для якої F_k менше. Проте слід пам'ятати, що при цьому можуть виникнути інші труднощі. По-перше, не завжди можна правильно визначити “лишню вхідну змінну ”. По-друге, в багатьох випадках виключення деяких вхідних змінних із моделі може негативно відобразитися на економічній суті моделі. По-третє, виключення із моделі вхідних змінних які істотно впливають на вихідну змінну, приводить до зміщення МНК оцінок.

Розглянемо реалізацію алгоритму Фаррара-Глобера на прикладі.

Приклад. Відомо, що на середньомісячну заробітну плату Y, впливають три фактори: рівень рентабельності – X_1, витрати капіталу – X_2, фондовіддача – X_3. Необхідні дані наведено в таблиці.

X_1,%	10,7	11,0	12,2	12,4	10,9	11,3	11,1
X_2, *ум.од.*	38	29	30	25	24	31	30
X_3, *ум.од.*	39	33	38	31	29	37	36
Y, *ум.од.*	63	70	80	84	69	72	70

Перевірити фактори X_1, X_2, X_3 на наявність мультиколінеарності за допомогою алгоритму Фаррара – Глобера.

Розв'язання.

Кроки 1-2-й. Для зручності числових обчислень об'єднаємо кроки перший і другий в один і розраховуємо кореляційну матрицю $\mathbf{r}$. Ця матриця симетрична. В нашому випадку розмір матриці 3×3. Матриця $\mathbf{r}$ має вигляд.

$$\mathbf{r}=\begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix},$$

де елементи матриці $\mathbf{r}$ розраховуються за формулами

$$r_{ij}=\frac{\overline{x_i x_j}-\bar{x}_i\cdot\bar{x}_j}{s_i\cdot s_j},\quad \bar{x}_i=\frac{1}{n}\sum_{k=1}^{n}x_{ik},\quad \overline{x_i^2}=\frac{1}{n}\sum_{k=1}^{n}x_{ik}^2,\quad \overline{x_i x_j}=\frac{1}{n}\sum_{k=1}^{n}x_{ik}x_{jk},\qquad s_i=\sqrt{\overline{x_i^2}-\bar{x}_i^{\,2}}\,.$$

Для розрахунку $\mathbf{r}$ сформуємо допоміжну розрахункову таблицю. Тоді

$$r_{12}=\frac{\overline{x_1 x_2}-\bar{x}_1\cdot\bar{x}_2}{s_1\cdot s_2}=\frac{335{,}21-11{,}37\cdot 29{,}57}{0{,}61\cdot 4{,}24}=\text{-}0{,}41,$$

$$r_{13}=\frac{\overline{x_1 x_3}-\bar{x}_1\cdot\bar{x}_3}{s_1\cdot s_3}=\text{-}0{,}08\qquad r_{23}=\frac{\overline{x_2 x_3}-\bar{x}_2\cdot\bar{x}_3}{s_2\cdot s_3}=0{,}88$$

№	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	$x_1\cdot x_2$	$x_1\cdot x_3$	$x_2\cdot x_3$
1	10,7	38	39	114,49	1444	1521	406,6	417,3	1482
2	11	29	33	121	841	1089	319	363	957
3	12,2	30	38	148,84	900	1444	366	463,6	1140
4	12,4	25	31	153,76	625	961	310	384,4	775
5	10,9	24	29	118,81	576	841	261,6	316,1	696
6	11,3	31	37	127,69	961	1369	350,3	418,1	1147
7	11,1	30	36	123,21	900	1296	333	399,6	1080
середні	11,37	29,57	34,71	129,69	892,43	1217,29	335,21	394,59	1039,57
s^2	0,38	17,96	12,20						
s	0,61	4,24	3,49						

Кореляційна матриця має вигляд

$$\mathbf{r}=\begin{pmatrix} 1 & -0{,}41 & -0{,}08 \\ -0{,}41 & 1 & 0{,}88 \\ -0{,}08 & 0{,}88 & 1 \end{pmatrix}$$

Елементами матриці $\mathbf{r}$ є коефіцієнти парної кореляції вхідних змінних. Між парами вхідних змінних існує зв'язок різної тісноти.

Крок 3-й. Знаходимо визначник матриці

$$|\mathrm{r}|=0{,}11$$

Крок 4-й. Обчислюємо χ^2 – статистику за формулою

$$\chi^2 = -\left[n-1-\frac{1}{6}(2p+5)\right]\ln(|r|).$$

Число спостережень $n=7$, число вхідних змінних $p=3$, тому

$$\chi^2 = -\left[7-1-\frac{1}{6}(2\cdot 3+5)\right]\ln(0{,}11) = 9{,}20.$$

При рівні значущості $\alpha = 0{,}05$ і числі ступенів свободи $p(p-1)/2 = 3$ за таблицею знаходимо критичне значення

$$\chi^2(p(p-1)/2;\alpha) = \chi^2(3;0{,}05) = 7{,}81.$$

Враховуючи, що

$$\chi^2 = 9{,}2 > \chi^2(3;0{,}05) = 7{,}81,$$

то мультиколінеарність в масиві вхідних змінних присутня.

Якби виявилось, що $\chi^2 < \chi^2(3;0{,}05)$, то це б означало, що мультиколінеарність відсутня, тобто нею можна знехтувати, і дальші дослідження за алгоритмом можна не проводити.

Крок 5-й. Знайдемо обернену матрицю r $^{-1}$ до матриці r за допомогою формули

$$\mathrm{r}^{-1} = \frac{1}{|r|}\begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} \qquad \mathrm{r}^{-1} = \begin{pmatrix} 2{,}07 & 3{,}12 & -2{,}58 \\ 3{,}12 & 9{,}13 & -7{,}78 \\ -2{,}58 & -7{,}78 & 7{,}64 \end{pmatrix}.$$

Крок 6-й. Розраховуємо $F-$ критерій за формулою

$$F_k = (c_{kk}-1)\frac{n-p}{p-1},$$

де c_{kk} – діагональні елементи матриці r^{-1}.

Таким чином отримуємо стільки фактичних значень F_k, скільки є вхідних змінних, тобто три

$$F_1 = (2{,}07-1)\frac{7-3}{3-1} = 2{,}15 \qquad F_2 = (9{,}13-1)\frac{7-3}{3-1} = 16{,}26 \qquad F_3 = (7{,}64-1)\frac{7-3}{3-1} = 13{,}29.$$

Отримані фактичні значення статистики порівнюються з табличним $F(\alpha;k_1;k_2)$ для $k_1 = p-1 = 3-1 = 2$ ступенів свободи чисельника, для $k_2 = n-p = 7-3 = 4$ ступенів свободи знаменника і рівня значущості $\alpha = 0{,}05$

$$F(0{,}05;2;4) = 6{,}94.$$

Оскільки

$$F_1 = 2{,}15 < F(0{,}05;2;4) = 6{,}94,$$

то X_1 не корелює з X_2 і X_3.

Оскільки

$$F_2 = 16{,}26 > F(0{,}05;2;4) = 6{,}94, \qquad F_3 = 13{,}29 > F(0{,}05;2;4) = 6{,}94,$$

то X_2 корелює з X_1 і X_3, а X_3 корелює з X_1 і X_2.

Крок 7-й. Обчислюємо частинні коефіцієнти кореляції

$$r_{k/j} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}.$$

Маємо

$r_{1/2} = \frac{-c_{12}}{\sqrt{c_{11} \cdot c_{22}}} = \frac{-3{,}12}{\sqrt{2{,}07 \cdot 9{,}13}} = -0{,}72$; $r_{1/3} = \frac{-c_{13}}{\sqrt{c_{11} \cdot c_{33}}} = \frac{2{,}58}{\sqrt{2{,}07 \cdot 7{,}64}} = 0{,}65$; $r_{2/3} = \frac{-c_{23}}{\sqrt{c_{22} \cdot c_{33}}} = \frac{7{,}78}{\sqrt{9{,}13 \cdot 7{,}64}} = \underline{0{,}93}$.

Як бачимо між вхідними змінними X_2 і X_3 зв'язок найтісніший.

Крок 8-й. Розраховуємо t – статистику

$$t_{ij} = \frac{r_{ij}\sqrt{n-p}}{\sqrt{1-r_{ij}^2}}.$$

$t_{12} = \frac{r_{12}\sqrt{7-3}}{\sqrt{1-r_{12}^2}} = \frac{-0{,}72 \cdot 2}{\sqrt{1-0{,}72^2}} = -2{,}08$; $t_{13} = \frac{r_{13}\sqrt{7-3}}{\sqrt{1-r_{13}^2}} = \frac{0{,}65 \cdot 2}{\sqrt{1-0{,}65^2}} = 1{,}71$; $t_{23} = \frac{r_{23}\sqrt{7-3}}{\sqrt{1-r_{23}^2}} = \frac{0{,}93 \cdot 2}{\sqrt{1-0{,}93^2}} = 5{,}06$.

Модулі отриманих фактичних значень $t-$ статистики порівнюємо з табличним $t(\alpha, n-p)$, яке знаходиться для $n-p$ числа ступенів свободи і рівня значущості $\alpha = 0{,}05$

$$t(0{,}05;4) = 2{,}776.$$

Оскільки

$$t_{23} = 5{,}06 > t(0{,}05;4) = 2{,}776,$$

то між вхідними змінними X_2 і X_3 існує взаємозалежність.

Взаємозв'язку між змінними X_1 і X_2, та X_1 і X_3 не існує, тому що $|t_{12}| = 2{,}08 < t(0{,}05;4) = 2{,}776$ і $t_{13} = 1{,}71 < t(0{,}05;4) = 2{,}776$.

Один із можливих методів усунення мультиколінеарності – це виключення однієї із вхідних змінних з моделі. Із економічних міркувань, в нашому випадку, при дослідженні залежності рівня заробітної плати Y від вхідних змінних X_1, X_2 і X_3 краще виключити із розгляду вхідну зміну X_3– фондовіддачу і залишити вхідну змінну X_2 – затрати капіталу. Після виключення цієї змінної повторимо перевірку на мультиколінеарність серед масиву змінних, що залишилися.

Кроки 1-2-й. Так як залишилися змінні X_1 і X_2, то кореляційна матриця матиме вигляд

$$\mathrm{r} = \begin{pmatrix} 1 & -0{,}41 \\ -0{,}41 & 1 \end{pmatrix}.$$

Крок 3-й. Знаходимо визначник матриці

$$|r| = 0{,}83$$

Крок 4-й. Обчислюємо χ^2 статистику за формулою

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2p+5)\right]\ln(|r|), \qquad \chi^2 = -\left[7 - 1 - \frac{1}{6}(2\cdot 2+5)\right]\ln(0{,}83) = 0{,}84,$$

Оскільки,

$$\chi^2\left(\frac{p(p-1)}{2}; \alpha\right) = \chi^2(1; 0{,}05) = 3{,}84,$$

то

$$\chi^2 = 0{,}84 < \chi^2(1;0{,}05) = 3{,}84,$$

тобто мультиколінеарність в масиві вхідних змінних X_1 і X_2 відсутня або, ще кажуть, знаходиться в допустимих межах.

Методи усунення мультиколінеарності

Різні методи, які можуть бути використані для зменшення негативних наслідків мультиколінеарності, умовно класифікують на дві категорії: **прямі** і **непрямі способи**.

Прямі способи базуються на наступних міркуваннях і відповідних їм діях.

По-перше, дисперсії оцінок параметрів пропорційні залишковій дисперсії s_u^2. Похибка U відображає вплив на вихідну змінну Y всіх вхідних змінних, які **не** включені в модель. Отже, якщо можна знайти важливу, з економічних міркувань, вхідну змінну X і включити її в модель, то таким чином ми зменшимо s_u^2 і покращимо надійність оцінок. Разом з тим, нова чи нові вхідні змінні можуть бути зв'язані лінійною залежністю із вхідними змінними, які вже включені в модель, і, таким чином ми можемо, навіть, підсилити проблему мультиколінеарності.

По-друге, дисперсії оцінок параметрів обернено-пропорційні об'єму вибірки n. Значить, збільшивши кількість спостережень, ми автоматично зменшимо дисперсії оцінок параметрів і покращимо їх надійність.

По-третє, оскільки дисперсії оцінок параметрів $s^2(a_i)$ обернено пропорційні дисперсії пояснюючих змінних $s^2(X_i)$, то дані необхідно підбирати таким чином, щоб $s^2(X_i)$ були найбільшими. Наприклад, якщо ми вивчаємо залежність товарообігу магазинів від величини торгової площі, то необхідно так сформувати вибірку, щоб там були і невеликі магазини з малою торговою площею і великі магазини.

По-четверте, при плануванні експерименту необхідно підбирати такі вхідні змінні, які були б якомога менше залежні один від одного.

Непрямі способи базуються на наступних міркуваннях і відповідних їм діях, направлених на зменшення негативних наслідків мультиколінеарності.

По-перше, самий простий спосіб зменшення мультиколінеарності полягає в тому, щоб із двох пояснюючих змінних X_i і X_j, які мають високий коефіцієнт кореляції і для яких $t_{ij} > t_{кр}$ одну виключити із розгляду. При цьому, яку змінну залишити, а яку виключити із аналізу, вирішують, в першу чергу, виходячи із економічних міркувань, тобто залишають в моделі більш важливу з економічної точки зору змінну. Якщо із економічних міркувань ні одній із них не можна надати перевагу, то залишають ту із змінних, яка має більший коефіцієнт кореляції із вихідною змінною.

По-друге, якщо корельовані змінні зв'язані між собою концептуально, то можливо, краще об'єднати їх в один сукупний індекс.

По-третє, для зменшення мультиколінеарності можливий перехід від вхідних змінних $X_1, \ldots, X_m$, між якими виявлена кореляція до нових змінних, які виражаються через лінійні комбінації вихідних. Нові змінні Z_j, які є лінійними комбінаціями змінних $X_1, \ldots, X_m$ називаються головними компонентами. При цьому нові змінні підбираються таким чином, щоб вони не корелювали між собою, а потім знаходять рівняння зв'язку $Y(Z)$.

6. Гетероскадестичність

Однією із умов застосування методу найменших квадратів є така.

Дисперсія залишків в кожному спостереженні повинна бути сталою
$s^2(u_i) = s^2 = const$

Іншими словами, в цій умові теореми Гаусса-Маркова стверджується, що ймовірність конкретного значення випадкового члена u_i не залежить від конкретних значень змінної X_i, а отже і дисперсія $s^2(u_i)$ не залежить від значень цих змінних, тобто

$$s^2(u_i) \neq f(X_i).$$

Якщо дисперсія залишків постійна $s^2(u_i) = const$ для кожного спостереження, то ця її властивість називається **гомоскедастичністю**. Гомоскедастичність означає “однакове розсіяння ”. В практичних дослідженнях явище гомоскедастичності часто порушується. В багатьох випадках більш реалістичним є припущення, що розподіл випадкового члена u_i в різних спостереженнях буде різним.

Наприклад, якщо вивчається залежність розміру заробітної плати Y_i співробітника від стажу роботи X_i то цілком природно сподіватися, що значення відхилення фактичних заробітних плат від середнього значення для певного стажу буде тим більше, чим більшою є заробітна плата.

Дуже часто так і трапляється. Це означає, що ймовірності різних можливих значень u_i будуть залежними від конкретних значень змінних

$X_j, j = 1,2,..., p$. При малих значеннях X_i малими будуть і u_i , при великих X_i , великими будуть і u_i, і, таким чином,

$$s^2(u_i) \neq const.$$

Якщо дисперсія залишків $s^2(u_i) \neq const$, то це явище називається **гетероскедастичністю**. Гетероскедастичність означає “неоднакове розсіяння ”. Розрізняють **чисту і мішану** (нечисту) гетероскедастичність.

Чиста гетероскедастичність виникає, коли умова про постійність дисперсії залишків $s^2(u_i) = const$ порушується при правильній специфікації моделі, внаслідок об’єктивних причин. Такими об’єктивними причинами можуть бути неточності вимірюваних факторів, вплив на пояснювану змінну другорядних факторів, які не включені в модель.

Мішана (нечиста) гетероскедастичність виникає при неправильній специфікації моделі, внаслідок не включення в рівняння регресії важливих пояснюваних змінних. Отже, величина залишку u_i акумулює в собі неточності вимірювань вхідних змінних, вплив на вихідну змінну другорядних факторів, які не включені в модель, неточності в специфікації моделі і т. д.

Наслідки гетероскедастичності.

За наявності гетероскедастичності оцінки параметрів моделі методом МНК залишаються незміщеними, обґрунтованими, але неефективними. Незміщенність і обґрунтованість оцінок означає, що модель, яка побудована методом МНК може бути використана для знаходження прогнозних значень залежної вихідної змінної Y. Однак інтервальні оцінки для параметрів моделі і для вихідної змінної будуть непридатними для аналізу, тому що інтервали будуть занадто широкими. Збільшення дисперсії залишків приводить до того, що F і $t-$критерії дають неточні результати. Ці критерії обчислюються в припущенні, що дисперсія похибки u_i гомоскедастична. Якщо не враховувати гетероскедастичність, то ми отримаємо завищенні значення цих критеріїв і тоді можна зробити висновок про значущість моделі і її параметрів, коли насправді вони будуть незначущими.

Отже, якщо знехтувати гетероскедастичністю і прийняти модель, параметри якої оцінені МНК, то це приводить до таких *трьох основних наслідків*:

1. Оцінки параметрів моделі можуть бути незміщеними, але неефективними, тобто, вибіркові дисперсії оцінок можуть бути невиправдано великими.

2. Для виявлення статистичної значущості моделі не можуть бути використані F і $t-$ критерії , оскільки обчисленні значення статистик будуть завищеними.

3. Неефективність оцінок параметрів моделі приводить до неефективності прогнозу, а отже і в цілому модель є непридатною для аналізу явища.

Виявлення гетероскадестичності. Тест Гольдфельда-Квандта

Природа гетероскедастичності, як зазначалося, може приймати різні форми. Для її виявлення застосовують різні методи і тести. Тестів на виявлення гетероскедастичності, по крайній мірі, сім. Як правило, у всіх цих тестах перевіряється основна гіпотеза $H_0 : s^2(u_i) = const$ проти альтернативи $H_1 : s^2(u_i) \neq const$.

Розглянемо параметричний тест Гольдфельда-Квандта. Даний тест перевіряє гіпотезу $H_0 : s^2(u_i) = const$ проти альтернативної $H_1 : s^2(u_i) \neq const$, коли $\overline{u_i \cdot u_i^{/}} = s_u^2 \cdot x_i^2$, тобто дисперсія зростає пропорційно до квадрата однієї із вхідних змінних.

В цьому тесті часто приймається цілком природне для багатьох випадків припущення: розсіяння тим більше, чим більшими є значення фактора.

Тест застосовується до значних за обсягом вибірок, причому ставиться умова, щоб число спостережень було вдвічі більшим за число пояснюючих змінних. Для реалізації тесту необхідно виконати **шість** кроків.

Крок 1-й. Впорядковуємо спостереження в порядку зростання значень незалежної змінної X_i, $i = 1,...,p$. Якщо вхідна змінна не одна, то вибирають одну з них, яка найбільш важлива, найбільш сильно впливає на вихідну змінну Y. Якщо таких важливих змінних декілька, то необхідно кожну з них перевіряти на гетероскедастичність.

Крок 2-й. Відкидаємо k спостережень. Відкидаються ті спостереження, що знаходяться в центрі впорядкованої сукупності. Емпірично встановлено, що число відкинутих k спостережень знаходиться із співвідношення

$$k = \frac{4}{15} n,$$

де $n-$число спостережень.

Решта спостережень утворюють дві рівновеликі вибірки, одну з яких утворюють ті спостереження, де значення змінної X_i менші, друга вибірка – та, де значення X_i більші. Якщо виявиться так, що відкинувши k спостережень залишиться непарне число спостережень, то можна відкинути ще одне значення, або в одній з двох новоутворених вибірок відкинути одне з крайніх спостережень.

Крок 3-й. Будуються дві економетричні моделі, на основі МНК за двома утвореними сукупностями. Ясно, що для побудови цих моделей необхідно, щоб

$$n_1 = \frac{n-k}{2} > m,$$

де m – число параметрів моделі, n_1 – обсяг двох новоутворених рівновеликих вибірок.
Для кожної із побудований моделей обчислюються залишки u_i.

Крок 4-й. Обчислюємо суму квадратів залишків для першої і другої моделі

$$S_1 = \sum_{i=1}^{n_1} u_{1i}^2, \qquad S_2 = \sum_{i=1}^{n_1} u_{2i}^2.$$

Крок 5-й. Обчислюємо значення статистики

$$F^* = \frac{S_2}{S_1}.$$

В чисельнику повинно стояти більше значення.

Крок 6-й. Порівнюємо обчислене значення статистики з табличним. Статистика F має розподіл Фішера. Обчислене значення F^* порівнюється з табличним $F(\alpha, k_1, k_2)$ для числа ступенів свободи $k_1 = k_2 = n_1 - m$ і рівня значущості α.
Якщо $F^* > F(\alpha, k_1, k_2)$, то гіпотеза H_0 відкидається,
а коли $F^* \leq F(\alpha, k_1, k_2)$, немає підстав відкидати гіпотезу H_0.

Якщо виявлено явище гетероскедастичності, то одним із шляхів для побудови економетричної моделі є застосовування узагальненого методу найменших квадратів або, як він називається, методу Ейткена.

Приклад. Дослідження даних на наявність гетероскедастичності за тестом Гольдфельда-Квандта.

Дослідити вхідні змінні – рівень рентабельності X_1 та затрати капіталу X_2 на наявність гетероскедастичності за тестом Гольдфельда-Квандта.
Необхідні дані наведено в таблиці.

Y, *ум. од.*	62	64	66	67	70	78	80	87	89	90	100
X_1,%	10,7	10,9	11,0	11,1	11,3	12,2	12,4	14,1	14,2	14,5	15,1
X_2, *ум. од.*	38	35	30	29	31	28	26	23	21	20	20

Розв'язання

Крок 1-й. Впорядковуємо дані спостереження відповідно до зростання значень змінної X_1, яка може вважатися більш важливою.

Крок 2-й. Відкидаємо k спостережень, що знаходяться в центрі впорядкованої сукупності

$$k = \frac{4}{15} n = \frac{4}{15} \cdot 11 \approx 3.$$

№	y_i	x_{1i}	x_{2i}	$\hat{y}_i$	$u_i = y_i - \hat{y}_i$	u_i^2
1	62	10,7	38	61,96	3,74E-02	1,40E-03
2	64	10,9	35	64,03	-3,40E-02	1,16E-03
3	66	11	30	66,04	-4,20E-02	1,76E-03
4	67	11,1	29	66,94	6,12E-02	3,75E-03
5	70	11,3	31			
6	78	12,2	28			
7	80	12,4	26			
8	87	14,1	23	87,45	-4,50E-01	2,03E-01
9	89	14,2	21	87,67	1,33E+00	1,77E+00
10	90	14,5	20	91,27	-1,27E+00	1,60E+00
11	100	15,1	20	99,63	3,65E-01	1,33E-01

Крок 3-й. На основі МНК будуємо дві економетричні моделі.
Перша модель: $\hat{y}_i = 6{,}285 + 6{,}19 \cdot x_1 - 0{,}2778 \cdot x_2$,
Друга модель: $\hat{y}_i = -122{,}72 + 13{,}947 \cdot x_1 + 0{,}587 \cdot x_2$.

Крок 4-й. Обчислюємо суму квадратів залишків для першої і другої моделей

$$S_1 = 8{,}06\text{E-}03 \qquad S_2 = 3{,}71$$

Крок 5-й. Обчислюємо значення статистики

$$F^* = \frac{S_2}{S_1} = \frac{3{,}71}{0{,}00806} = 459{,}97$$

Крок 6-й. При $k_1 = k_2 = n_1 - m = 4 - 3 = 1$ і $\alpha = 0{,}05$

$$F(\alpha, k_1, k_2) = F(0{,}05;1;1) = 161$$

Оскільки

$$F^* = 459{,}97 > F(0{,}05;1;1) = 161,$$

то гіпотеза про гомоскедастичність залишків відхиляється. В такому випадку можливі два шляхи усунення негативних наслідків гетероскедастичності:

1) змінити специфікацію моделі так, щоб нова специфікація була вільна від гетероскедастичності,

2) провести оцінювання за узагальненим методом найменших квадратів.

Виявлення гетероскадестичності. Тест рангової кореляції Спірмена

Ідея цього методу полягає в наступному: оскільки абсолютні величини залишків u_i можуть бути оцінками залишкової дисперсії $s^2(u_i)$, тому у випадку гетероскедастичності ($s^2(u_i) \neq const$) залишки корелюють із значеннями вхідних змінних.

Встановити наявність такої кореляції можна с допомогою тесту рангової кореляції Спірмена. Для застосування даного тесту об’єкти аналізу спочатку впорядковують відповідно до величини однієї із ознак, які вивчаються. При цьому кожному об'єкту присвоюється свій порядковий номер.

Рангом називається порядковий номер, який присвоюється об'єкту. Отже ранг, характеризує величину ознаки. Даний тест може бути застосований до вибірок довільного розміру.

На закон розподілу залишків не накладаються ніякі обмеження. Алгоритм розрахунку складається із 5 кроків.

Крок 1-й. За допомогою МНК будується рівняння регресії $\hat{y} = f(x)$.

Крок 2-й. На основі побудованої моделі розраховуються величини залишків $u_i = y_i - \hat{y}_i$, $i = 1,...,n$.

Крок 3-й. Ранжуються абсолютні величини залишків $|u_i|$ в порядку зростання або в порядку спадання. Кожному значенню $|u_i|$ присвоюється свій ранг R_{u_i}. Наприклад, найбільшому (найменшому) значенню $|u_i|$ присвоюємо ранг 1, наступному меншому (більшому) ранг 2 і т.д. Кількість рангів буде дорівнювати обсягу вибірки.

Аналогічно операція виконується для всіх пояснюючих змінних X_j, $j = 1,...,m$.

Тобто проводиться ранжування значень x_{j1}, $x_{j2}, \ldots, x_{j2n}$. Відповідні ранги ознак позначимо $R_{x_{ji}}$.

Крок 4-й. Для кожної змінної X_j обчислюється різниця рангів

$$d_{ij} = R_{u_i} - R_{x_{ij}},$$

і визначається **коефіцієнт рангової кореляції Спірмена** за формулою

$$r_j = 1 - 6\frac{\sum_{i=1}^{n} d_{ij}^2}{n(n^2 - 1)},$$

де $n -$ обсяг вибірки.

Коефіцієнт r_j може приймати значення $|r_j| \leq 1$. Якщо ранги співпадають, то $r_j = 1$. В цьому випадку між залишками і величиною значень ознаки X_j є однозначний прямий зв'язок, тобто найбільшому значенню залишку відповідає найбільше значення ознаки X_j. Якщо $r_j = -1$, то ранги ознаки X_j і ранги залишків розміщені в протилежному напрямку. В цьому випадку зв'язок однозначний, але обернений.

Крок 5-й. Перевіряється значущість коефіцієнта рангової кореляції r_j за критерієм Стьюдента. Для цього розрахуємо t - критерій Стьюдента за формулою

$$t_j = \frac{|r_j|\sqrt{n-1}}{\sqrt{1 - r_j^2}}.$$

Обчислене значення t_j t- критерію порівнюється з табличним для числа ступенів свободи $k = n - 2$ і рівня значущості α $-$ $t(k;\alpha)$.

Якщо

$$t_j > t(k;\alpha),$$

то гіпотеза про наявність гетероскедастичності приймається і вона викликана саме ознакою X_j.

Якщо

$$t_j < t(k;\alpha) \;,$$

то вважається, що в моделі має місце гомоскедастичність.

При ранжуванні об’єктів за якісними ознаками можуть бути випадки, коли деякі об’єкти за ознаками не відрізняються один від одного. Такі об’єкти називають **зв’язаними**. Зв’язаним об’єктам присвоюється однаковий середній ранг.

Наприклад, якщо не можна відрізнити 5, 6 і 7 об’єкти, то їм необхідно присвоїти однаковий ранг, що дорівнює 6=(5+6+7)/3. Об’єкти 5, 6, 7 будуть зв’язаними.

Найбільший можливий ранг зажди дорівнює числу спостережень.

Розглянемо застосування тесту рангової кореляції Спірмена на прикладі тих самих даних, що досліджувалися за допомогою тесту Гольдфельда-Квандта.

Приклад.

Дослідити вхідні змінні: рівень рентабельності X_1 та затрати капіталу X_2 на гетероскедастичність за тестом рангової кореляції Спірмена.

Необхідні дані наведено в таблиці.

Y, *ум. од.*	62	64	66	67	70	78	80	87	89	90	100
X_1,%	10,7	10,9	11,0	11,1	11,3	12,2	12,4	14,1	14,2	14,5	15,1
X_2, *ум. од.*	38	35	30	29	31	28	26	23	21	20	20

Розв’язання.

Крок 1-й. За звичайними правилами використовуючи МНК, будуємо лінійну регресійну модель двофакторної регресії

$$\hat{y}_i = 5{,}5 + 6{,}448 \cdot x_{1i} - 0{,}3127 \cdot x_{2i}.$$

Крок 2-й. На основі побудованої моделі розраховуємо величини залишків $|u_i| = |y_i - \hat{y}_i|$. Для зручності всі розрахунки заносимо в таблицю.

№	y_i	x_{1i}	x_{1i}	$\hat{y}_i$	u_i	R_{u_i}	$R_{x_{1i}}$	$R_{x_{2i}}$	d_{i1}^2	d_{i2}^2
1	62	10,7	38	62,61	0,61	9	11	1	4	64
2	64	10,9	35	64,84	0,84	8	10	2	4	36
3	66	11	30	67,05	1,05	10	9	4	1	36
4	67	11,1	29	68,01	1,01	11	8	5	9	36
5	70	11,3	31	68,67	1,33	4	7	3	9	1
6	78	12,2	28	75,41	2,59	7	6	6	1	1
7	80	12,4	26	77,33	2,67	5	5	7	0	4
8	87	14,1	23	89,23	2,23	3	4	8	1	25
9	89	14,2	21	90,50	1,50	6	3	9	9	9
10	90	14,5	20	92,74	2,74	2	2	10,5	0	72,5
11	100	15,1	20	96,61	3,39	1	1	10,5	0	90,5
сума									38	375

Крок 3-й. Ранжуємо абсолютні величини залишків $|u_i|$. Найбільшому значенню $|u_i|$ присвоюємо ранг 1, наступному меншому ранг 2 і т.д. Аналогічну операцію виконуємо для вхідних змінних X_1 і X_2. Тобто проводиться ранжування значень x_{1i}, x_{2i}, $i = 1,...,n$.Найбільшим значенням ознак X_1, X_2 присвоюємо ранги 1, наступним меншим ранги 2 і т.д. Відповідні ранги вхідних змінних позначимо $R_{x_{1i}}$, $R_{x_{2i}}$.Дані занесімо в таблицю.

Крок 4-й. Для кожної вхідної змінної X_1, X_2 обчислюємо різницю рангів $d_{ij} = R_{u_i} - R_{x_{ij}}$, $i = 1,...,n$, $j = 1,2$, і визначаємо коефіцієнти рангової кореляції Спірмена за формулою

$$r_j = 1 - 6\frac{\sum_{i=1}^{n} d_{ij}^2}{n(n^2-1)}.$$

Таким чином, отримаємо

$$r_1 = 1 - 6\frac{\sum_{i=1}^{n} d_{i1}^2}{n(n^2-1)} = 1 - 6\frac{38}{11\cdot(11^2-1)} = 0{,}83,$$

$$r_2 = 1 - 6\frac{\sum_{i=1}^{n} d_{i2}^2}{n(n^2-1)} = 1 - 6\frac{375}{11\cdot(11^2-1)} = -0{,}70.$$

Тобто зв'язок між значеннями X_1 і величиною залишків прямий і тісний. Зв'язок між значеннями X_2 і величиною залишків обернений і тісний.

Крок 5-й. Перевіряємо значущість коефіцієнтів рангової кореляції r_1 і r_2 за критерієм Стьюдента. Для цього розрахуємо t- критерій Стьюдента за формулою

$$t_j = \frac{|r_j|\sqrt{n-1}}{\sqrt{1-r_j^2}}.$$

Знаходимо

$$t_1 = \frac{|r_1|\sqrt{n-1}}{\sqrt{1-r_1^2}} = \frac{0{,}83 \cdot \sqrt{11-1}}{\sqrt{1-0{,}83^2}} = 4{,}71\ ,$$

$$t_2 = \frac{|r_2|\sqrt{n-1}}{\sqrt{1-r_2^2}} = \frac{0{,}7 \cdot \sqrt{11-1}}{\sqrt{1-0{,}7^2}} = 3{,}10.$$

Обчислені значення критеріїв порівнюємо з табличним для числа ступенів свободи $k = n - m = 11 - 2 = 9$ і рівня значущості $\alpha = 0{,}05 - t(9;0{,}05) = 2{,}262$

Оскільки

$$t_1 = 4{,}71 > t(9;0{,}05) = 2{,}262 \text{ і } \quad t_2 = 3{,}1 > t(9;0{,}05) = 2{,}262\,,$$

то гіпотеза про наявність гетероскедастичності приймається, як за ознакою X_1, так і за ознакою X_2.

8. Автокореляція

Автокореляція – це наявність взаємозв'язку між послідовними значеннями часового чи просторового ряду даних.

В економетричних моделях часто виникають такі ситуації, коли дисперсія залишків є сталою, але між ними спостерігається кореляція. Це явище називають автокореляцією залишків.

Однією з передумов застосування МНК є виконання умови теореми Гауса - Маркова: $\overline{u_i \cdot u_j} = 0$. На практиці ця умова часто не виконується. Це означає, що залишки в моделі не є незалежними випадковими величинами, між ними є кореляція. Моделі, в яких умова $\overline{u_i \cdot u_j} = 0$ не виконується, називаються моделями з автокореляцією.

Автокореляція буває **додатною** і **від'ємною**. Якщо, наприклад, вивчається курс цінного паперу, то коли на попередніх торгах він виявився дещо завищеним, то дуже часто і на наступних торгах він також буде завищеним. В таких випадках говорять про **додатну автокореляцію**.

Від'ємна автокореляція зустрічається в тих випадках, коли спостереження діють один на одного за принципом маятника – завищенні значення ознаки в попередньому спостереженні приводить до заниженого значення ознаки в наступному спостереженні.

Як правило, якщо автокореляція має місце, то найбільший вплив на наступні спостереження має результат сусіднього попереднього спостереження. Наприклад, якщо розглядається ряд значень курсу деякого цінного паперу, то саме результат останніх торгів найсильніше впливає на значення курсу на наступних торгах.

Таким чином, відсутність автокореляції між сусідніми членами часто є достатньою умовою для висновку про відсутність автокореляції в цілому.

Якщо ж знехтувати автокореляцією залишків і приймати модель параметри якої оцінені МНК, то це приводить до таких самих трьох основних наслідків як і у випадку гетероскедастичності.

Для виявлення автокореляції є кілька тестів. Найчастіше застосовують критерій Дарбіна – Уотсона (DW – критерій).

Критерій Дарбіна – Уотсона

Даний критерій виявляє наявність автокореляції між сусідніми залишками. Цей тест базується на такій ідеї: якщо автокореляція в залишках u_i є, то вона буде проявлятися і в залишках u_i, які отримані в результаті застосування МНК.

Застосування тесту Дарбіна-Уотсона можна розбити на три кроки.

Крок 1-й. Методом МНК будується регресійна модель.

Крок 2-й. Обчислюються залишки $u_i = y_i - \hat{y}_i$. Обчислюємо статистику Дарбіна-Уотсона за формулою

$$DW = \frac{\sum_{i=2}^{n} (u_i - u_{i-1})^2}{\sum_{i=1}^{n} u_i^2}.$$

Значення даної статистики можуть знаходитися на інтервалі від 0 до 4. Можна показати, що

$$DW = 2\left(1 - \frac{\sum_{i=2}^{n} u_i u_{i-1}}{\sum_{i=1}^{n} u_i^2}\right) - \frac{u_1^2 + u_n^2}{\sum_{i=1}^{n} u_i^2}$$

При великій вибірці, очевидно, що

$$u_1^2 + u_n^2 << \sum_{i=1}^{n} u_i^2,$$

тому

$$DW \approx 2(1-r),$$

де $r = \frac{\sum_{i=2}^{n} u_{i-1} u_i}{\sum_{i=2}^{n} u_i^2}$ – коефіцієнт кореляції між сусідніми членами.

Очевидно, що у випадку відсутності автокореляції $r \approx 0$, тому $DW \approx 2$. Якщо $DW \approx 0$, то це означає додатну автокореляцію, якщо $DW \approx 4$– від'ємну автокореляцію. Отже, $0 \le DW \le 4$.

Приймається гіпотеза про додатну автокореляцію	Зона невизначеності	Приймається гіпотеза про відсутність автокореляції	Зона невизначеності	Приймається гіпотеза про від'ємну автокореляцію

$0 \quad DW_L \quad DW_U \quad 2 \quad 4-DW_U \quad 4-DW_L \quad 4$

Крок 3-й. Порівнюється обчислене значення DW з табличними $DW_L(\alpha, n, p)$ і $DW_U(\alpha, n, p)$ – нижньою і верхньою границями.

DW_L і DW_U знаходяться для заданого рівня значущості α і числа ступенів свободи n – кількість спостережень, p – число пояснюючих змінних в регресійній моделі.

Для проведення порівняння розрахованого значення DW і двох табличних DW_L і DW_U використовується наведена шкала.

Отже, якщо, наприклад, $DW_U < DW < 4 - DW_U$, то робиться висновок, що автокореляція залишків відсутня.

Критичні значення DW_L і DW_U можна знайти, якщо обсяг вибірки не менше 15.

Як видно з шкали інтервалів, тест Дарбіна-Уотсона має той недолік, що він має зони невизначеності при попаданні в які, питання про наявність чи відсутність автокореляції залишається відкритим. Якщо обчислене значення DW попадає в зону невизначеності, в цьому випадку пропонується приєднати область невизначеності до області відхилення, тобто вважати, що автокореляція присутня, або застосувати так звані $h-$ або $m-$ тести Дарбіна, які в даному посібнику не розглядаються, або обчислити циклічний коефіцієнт автокореляції і на основі цих обчислень зробити висновок.

Циклічний коефіцієнт автокореляції r^0 обчислюється за формулою

$$r^0 = \frac{\sum_{i=2}^{n} u_{i-1}u_i}{\sum_{i=2}^{n} u_i^2}.$$

Якщо обсяг вибірки невеликий, то коефіцієнт r^0 обчислюється за формулою

$$r^0 = \frac{\frac{1}{n-1}\sum_{i=2}^{n} u_{i-1}u_i}{\frac{1}{n}\sum_{i=2}^{n} u_i^2}.$$

Обчислене фактичне значення коефіцієнта автокореляції r^0 порівнюється з табличним $r(\alpha, n)$ для вибраного рівня значущості α і обсягу вибірки n.

Якщо $r^0 > r(\alpha, n)$, то автокореляція існує.

На застосування критерію Дарбіна – Уотсона накладаються такі обмеження:

1. Критерій Дарбіна – Уотсона виявляє автокореляцію залишків тільки першого порядку. При перевірці залишків на автокореляцію більш високого порядку необхідно застосовувати інші методи.
2. Критерій Дарбіна –Уотсона застосовується для вибірок обсягом не менше 15.

Зауваження.

Низьке значення DW статистики не завжди означає, що в моделі присутня додатна автокореляція першого порядку. Значення DW статистики може виявитися низьким, якщо в моделі пропущена важлива факторна ознака або якщо специфікація моделі помилкова.

Приклад. Дослідити модель, побудовану згідно даних таблиці, на наявність
автокореляції залишків з допомогою критерію Дарбіна – Уотсона.

X, грн.	45	60	57	62	59	47	55
Y , %	69	61	60	57	55	67	65

Згідно МНК маємо рівняння регресії

$$\hat{y} = 100{,}8 - 0{,}705 \cdot x$$

Будуємо допоміжну таблицю.

№	x_i	y_i	$\hat{y}_i$	$u_i = y_i - \hat{y}_i$	u_i^2	$u_i - u_{i-1}$	$(u_i - u_{i-1})^2$
1	45	69	69,075	-0,075	0,01	X	X
2	47	67	67,665	-0,665	0,44	-0,59	0,35
3	55	65	62,025	2,975	8,85	3,64	13,25
4	57	60	60,615	-0,615	0,38	-3,59	12,89
5	59	55	59,205	-4,205	17,68	-3,59	12,89
6	60	61	58,5	2,5	6,25	6,71	44,96
7	62	57	57,09	-0,09	0,01	-2,59	6,71
$\sum$	X	X	434,18	X	33,61	X	90,69

Обчислюємо значення критерію Дарбіна – Уотсона

$$DW = \frac{\sum_{i=2}^{n} (u_i - u_{i-1})^2}{\sum_{i=1}^{n} u_i^2} = \frac{90{,}69}{33{,}61} = 2{,}70$$

Порівнюємо значення критерію з табличними для $\alpha = 0{,}05$, $n = 7$, $p = 1$.

В даному випадку критичні значення DW такі: $DW_L(0{,}05;7;1) = 0{,}7$, $DW_U(0{,}05;7;1) = 1{,}36$.

Оскільки

$$4 - DW_U = 2{,}64 < DW = 2{,}70 < 4 - DW_L = 3{,}3,$$

то попадаємо в зону невизначеності, висновок зробити неможливо. Однак рекомендується поступати так, як при наявності автокореляції.

9. Комплексні індивідуальні завдання

Завдання 1

Для характеристики залежності Y ***від*** X ***:***

1. Побудувати кореляційне поле залежності Y від X ;
2. Побудувати модель парної регресії використавши лінійну, параболічну і гіперболічну функції. Для побудови моделей застосувати методи аналітичної геометрії.
2. Побудувати модель парної регресії використавши лінійну, параболічну і гіперболічну функції. Для побудови моделей застосувати МНК.
Порівняти моделі згідно найдених параметрів і зробити висновки.
3. Серед побудованих моделей по МНК з допомогою середнього коефіцієнта апроксимації вибрати найбільш адекватну.
4. Знайти коефіцієнти та індекси кореляції та детермінації, зробити висновок.
5. Оцінити значущість вибраної моделі з допомогою F і t -критеріїв.
6. Знайти коефіцієнт еластичності.
7. Знайти точкові і інтервальні оцінки вихідної змінної Y для значень вхідної змінної на 10% більших від найбільш фактичного.

Зауваження. Необхідні дані для кожного варіанту беруть із відповідної таблиці за номером свого варіанту.

Таблиця 1

Чистий дохід, Y, млн. грн.	9	1,8	0,7	1,7	6	3	4	6	7	8
Оборот капіталу, X, млн. гр.	31,2	13,4	4,5	10	20	15	14	18	16,5	25

Таблиця 2

Чистий дохід, Y, млн. грн.	1,6	0,4	1,3	1,9	1,2	1,4	0,4	0,8	1,8	0,9
Оборот капіталу, X, млн. гр.	18	5	7	27	13	10	4	7	24	12

Таблиця 3

Чистий дохід, Y, млн. грн.	1,8	2,6	0,9	1,1	1,6	2,8	1,3	2	0,6	0,7
Оборот капіталу, X, млн. гр.	13,4	20	12	10	13	21	13,5	18	4,2	8

Таблиця 4

Чистий дохід, Y, млн. грн.	2,6	0,8	2	1,1	1,6	2,8	1,3	1,6	0,6	0,7
Оборот капіталу, X, млн. гр.	31,2	13,4	25	10	20	40	13,5	18	4,2	2

Таблиця 5

Чистий дохід, Y, млн. грн.	1,6	1,5	1,3	2,5	1,6	2,8	1,3	1,6	0,6	1
Оборот капіталу, X, млн. гр.	18	20	7	27	13	29	13,5	7	4,2	10

Таблиця 6

Чистий дохід, Y, млн. грн.	1,8	2,6	0,9	2,3	1,3	2,6	1,3	2	0,6	0,7
Оборот капіталу, X, млн. гр.	13,4	20	5	18	8	21	13,5	13,4	4,2	3

Таблиця 7

Ринкова капіталізація , Y, млн. грн.	35	40,5	39	38,5	38	15	25	32	20	28
Оборот капіталу ,X, млн. грн.	55	13,4	4,5	10	40	180	140	80	165	110

Таблиця 8

Ринкова капіталізація ,Y, млн. грн.	20	18	35	10	26	33	15	32	15	29,8
Оборот капіталу ,X, млн. грн.	18	20	7	27	13	10	20	7	25	12

Таблиця 9

Ринкова капіталізація ,Y, млн. грн.	41	37	30	25	29,5	23,2	29,2	29	20	27
Оборот капіталу ,X, млн. грн.	4	7	9	18	13	21	13,5	13,4	21	15,5

Таблиця 10

Ринкова капіталізація ,Y, млн.грн.	41	40	39	38,5	37,5	26,5	37	36,8	36	35
Оборот капіталу ,X ,млн. грн.	0,9	0,8	0,7	1,7	2,6	1,2	4	1,6	7	0,4

Таблиця 11

Ринкова капіталізація , Y млн. грн.	37	35	25	20	15	20	32,5	32	21	29,5
Оборот капіталу ,X, млн. грн.	0,3	0,4	1,3	1,9	1,9	1,4	0,4	0,8	1,8	0,9

Таблиця 12

Ринкова капіталізація ,Y, млн. грн.	40	45	30	35	42	25	29	38	28	27
Оборот капіталу ,X, млн. грн.	1,8	2,6	0,9	1,1	1,9	0,4	1,3	2	0,5	0,7

Таблиця 13

Чистий дохід, Y, млн. грн.	0,9	0,7	3,5	4	6,9	6	3,5	0,4	2	5
Оборот капіталу ,X, млн. грн.	31	4,5	110	137	165	150	75	19	51	125

Таблиця 14

Ринкова капіталізація ,Y, млн. грн.	41	30	28	37	36	30	26	33	28	25
Оборот капіталу ,X, млн. грн.	180	100	60	137	165	68	13,4	110	45	3

Таблиця 15

Чистий дохід, Y, млн. грн.	1,7	1,7	1,3	1,6	0,4	1,9	0,8	0,8	0,9	1,9
Затрати капіталу,X, млн. грн.	23	20	15	18	2	27	9,8	6,8	12,4	25

Таблиця 16

Ринкова капіталізація ,Y, млн. грн.	40,5	38,5	26,5	37	35	35	33	32	30	29
Оборот капіталу ,X, млн. грн.	2	10	28	18	15	5	9,8	20	25	23

Таблиця 17

Ринкова капіталізація , Y, млн. грн.	5	2	6	2,5	7	0,4	1,3	4,5	3,1	1,4
Оборот капіталу ,X, млн. грн.	120	50	137	72	165	2	7	100	90	30

Таблиця 18

Чистий дохід, Y, млн. грн.	37	26,5	17	37	15	39	35	30	26	33
Затрати капіталу,X, млн. грн.	40	125	137	18	165	5	55	90	110	100

Таблиця 19

Чистий дохід, Y, млн. грн.	0,4	0,8	1,8	0,9	1,1	1,9	1,5	2	0,6	1,1
Затрати капіталу,X, млн. грн.	2	6,8	27	12	17,5	25	21,5	28	4	15,5

Таблиця 20

Ринкова капіталізація ,Y ,млн. грн.	25	15	30,5	17	25,5	20	29	20	10	27
Оборот капіталу ,X_1 млн. грн.	19,5	6,8	27	12	17,5	10	21,5	13,5	4	15,5

Таблиця 21

Чистий дохід,Y ,млн. грн.	4	2,5	0,7	1,7	5	1,2	6	3	7	0,4
Затрати капіталу,X, млн. грн.	100	60	4,5	40	110	15	140	90	165	2

Таблиця 22

Чисельність працівників, Y, тис. чол.	85	105	7	130	62	42	105	34	142	58
Оборот капіталу , X ,млн. грн.	18	23	4,5	27	13	10	20	7	27	12

Таблиця 23

Чисельність працівників, Y тис. чол.	100	106	96	140	60	130	71	65	23	81
Оборот капіталу , X , млн. грн.	17	15	12	21	7	20	8	5	2	10

Таблиця 24

Чисельність працівників, Y тис. чол.	63	65	24	50	106	40	34,7	86	74,5	8
Чистий дохід , X ,млн. грн.	21	23	7	17	30	12	4	25	28	2

Таблиця 25

Чисельність працівників,Y тис. чол.	200	180	150	42	62	212	105	125	110	120
Чистий дохід , X ,млн. грн.	1,6	1,5	1,3	0,2	0,4	1,8	0,4	0,8	0,6	0,9

Таблиця 26

Чисельність працівників, Y тис. чол.	65	106	96	140	60	130	40	65	23	30
Чистий дохід , X ,млн. грн.	1,8	2,6	2,3	2,8	1,9	2,5	1,3	2	0,5	0,7

Таблиця 27

Чисельність працівників, Y тис. чол.	120	65	21	50	106	97	110	86	74,5	4
Оборот капіталу, X ,млн. грн.	31,2	13,4	4,5	10	20	15	25	18	16,5	2

Таблиця 28

Чисельність працівників, Y тис. чол.	86	120	18	125	62	21,2	105	33,5	142	36
Оборот капіталу, X ,млн. грн.	18	23	7	25	13	10	20	9	27	12

Таблиця 29

Чисельність працівників, Y тис. чол.	65	106	96	140	60	131	71	65,5	23	81
Оборот капіталу, X ,млн. грн.	10	20	12	23	8	21	13,5	6	4,2	15,5

Таблиця 30

Чисельність працівників, Y тис. чол.	43	110	60	85	80	97	80	60	74,5	20
Оборот капіталу, X ,млн. грн.	31	4,5	20	13,7	16,5	6,8	10	25	15	33

Завдання 2

Для характеристики залежності Y від **двох** факторів

1. Побудувати лінійне рівняння багатофакторної регресії аналітичним способом і МНК та пояснити економічний зміст його параметрів.
2. Зробити висновок про силу зв'язку вихідної змінної і вхідних змінних.
3. Оцінити значущість рівняння регресії на основі F - критерію Фішера і зробити висновок.
4. Оцінити значущість параметрів рівняння регресії на основі □ t – критерію Стьюдента, зробити висновок.
5. Знайти точкові і інтервальні оцінки вихідної змінної Y для очікуваних значень вхідних змінних, які на 10% більші від найбільших фактичних значень.
6. Знайти довірчі інтервали для параметрів регресії.
7. Використати пакет "АНАЛИЗ ДАННЫХ" для розрахунку параметрів регресії.
8. Використовуючи алгоритм покрокового регресійного аналізу отримати рівняння регресії з максимальним числом значимих коефіцієнтів регресії, при цьому попередньо не відкидаючи ні однієї з вхідних змінних.

Таблиця 1

Чистий дохід, Y, млн. грн.	9	1,8	0,7	1,7	6	3	4	6	7	8
Оборот капіталу, X_1 , млн. гр.	31,2	13,4	4,5	10	20	15	14	18	16,5	25
Затрати капіталу, X_2, млн. гр.	19	5	2	8	22	12	10	20	16	25

Таблиця 2

Чистий дохід, Y, млн. грн.	1,6	0,4	1,3	1,9	1,2	1,4	0,4	0,8	1,8	0,9
Оборот капіталу, X_1 , млн. гр.	18	5	7	27	13	10	4	7	24	12
Затрати капіталу, X_2, млн. гр.	20	3	8	19	10	12,6	4	5	13	7

Таблиця 3

Чистий дохід, Y, млн. грн.	1,8	2,6	0,9	1,1	1,6	2,8	1,3	2	0,6	0,7
Оборот капіталу, X_1 , млн. гр.	13,4	20	12	10	13	21	13,5	18	4,2	8
Затрати капіталу, X_2 , млн. гр.	14	22	7	15	12	16	8,6	11,5	1,9	5

Таблиця 4

Чистий дохід, Y, млн. грн.	2,6	0,8	2	1,1	1,6	2,8	1,3	1,6	0,6	0,7
Оборот капіталу, X_1, млн. гр.	31,2	13,4	25	10	20	40	13,5	18	4,2	2
Затрати капіталу, X_2,млн. гр.	40	14	70	35	22	58	8,6	50	1,9	10

Таблиця 5

Чистий дохід, Y, млн. грн.	1,6	1,5	1,3	2,5	1,6	2,8	1,3	1,6	0,6	1
Оборот капіталу, X_1 , млн. гр.	18	20	7	27	13	29	13,5	7	4,2	10
Затрати капіталу, X_2, млн. гр.	60	20	50	80	70	126	40	100	1,9	7

Таблиця 6

Чистий дохід, Y, млн. грн.	1,8	2,6	0,9	2,3	1,3	2,6	1,3	2	0,6	0,7
Оборот капіталу, X_1, млн. гр.	13,4	20	5	18	8	21	13,5	13,4	4,2	3
Затрати капіталу, X_2, млн. гр.	14	18	7	15	12	16	8,6	11,5	1,9	5

Таблиця 7

Ринкова капіталізація , Y млн. грн.	35	40,5	39	38,5	38	15	25	32	20	28
Оборот капіталу ,X_1 млн. грн.	55	13,4	4,5	10	40	180	140	80	165	110
Затрати капіталу, X_2 млн. грн.	42	14	30	4,8	22	105	99	55	61	71

Таблиця 8

Ринкова капіталізація , Y млн. грн.	20	18	35	10	26	33	15	32	15	29,8
Оборот капіталу ,X_1 млн. грн.	18	20	7	27	13	10	20	7	25	12
Затрати капіталу, X_2 млн. грн.	20	23	3	19	13	12,6	16	3,2	18	7

Таблиця 9

Ринкова капіталізація , Y млн. грн.	41	37	30	25	29,5	23,2	29,2	29	20	27
Оборот капіталу ,X_1 млн. грн.	4	7	9	18	13	21	13,5	13,4	21	15,5
Затрати капіталу, X_2 млн. грн.	20	17	7	15	12	1,6	8,6	11,5	1,9	5

Таблиця 10

Ринкова капіталізація , Y млн. грн.	41	40	39	38,5	37,5	26,5	37	36,8	36	35
Оборот капіталу ,X_1 млн. грн.	0,9	0,8	0,7	1,7	2,6	1,2	4	1,6	7	0,4
Затрати капіталу, X_2 млн. грн.	31,2	13,4	4,5	10	20	15	140	18	165	2

Таблиця 11

Ринкова капіталізація , Y млн. грн.	37	35	25	20	15	20	32,5	32	21	29,5
Оборот капіталу ,X_1 млн. грн.	0,3	0,4	1,3	1,9	1,9	1,4	0,4	0,8	1,8	0,9
Затрати капіталу, X_2 млн. грн.	4	3	7	18	25	20	5	7	15	12

Таблиця 12

Ринкова капіталізація , Y млн. грн.	40	45	30	35	42	25	29	38	28	27
Оборот капіталу ,X_1 млн. грн.	1,8	2,6	0,9	1,1	1,9	0,4	1,3	2	0,5	0,7
Затрати капіталу, X_2 млн. грн.	13,4	10	20	18	13	21	23	13,4	25	26

Таблиця 13

Чистий дохід, Y млн. грн.	0,9	0,7	3,5	4	6,9	6	3,5	0,4	2	5
Оборот капіталу ,X_1 млн. грн.	31	4,5	110	137	165	150	75	19	51	125
Затрати капіталу, X_2 млн. грн.	19	18,5	47	60	90	80	38	12	25	70

Таблиця 14

Ринкова капіталізація , Y млн. грн.	41	30	28	37	36	30	26	33	28	25
Оборот капіталу ,X_1 млн. грн.	180	100	60	137	165	68	13,4	110	45	3
Затрати капіталу, X_2 млн. грн.	110	45	22	99	61	30	13,2	80	40	5

Таблиця 15

Чистий дохід, Y млн. грн.	1,7	1,7	1,3	1,6	0,4	1,9	0,8	0,8	0,9	1,9
Затрати капіталу,X_1 млн. грн.	23	20	15	18	2	27	9,8	6,8	12,4	25
Оборот капіталу , X_2 млн. грн.	14	17	10	13	1,4	19	5	3,2	7	24

Таблиця 16

Ринкова капіталізація , Y млн. грн.	40,5	38,5	26,5	37	35	35	33	32	30	29
Оборот капіталу ,X_1 млн. грн.	2	10	28	18	15	5	9,8	20	25	23
Затрати капіталу, X_2 млн. грн.	2	4,8	26	10	18	8	12,6	16	20	23

Таблиця 17

Ринкова капіталізація , Y млн. грн.	5	2	6	2,5	7	0,4	1,3	4,5	3,1	1,4
Оборот капіталу ,X_1 млн. грн.	120	50	137	72	165	2	7	100	90	30
Затрати капіталу, X_2 млн. грн.	90	45	99	60	110	1,4	8	81	65	25

Таблиця 18

Чистий дохід, Y млн. грн.	37	26,5	17	37	15	39	35	30	26	33
Затрати капіталу,X_1 млн. грн.	40	125	137	18	165	5	55	90	110	100
Оборот капіталу , X_2 млн. грн.	30	70	99	20	110	1,4	8	60	80	40

Таблиця 19

Чистий дохід, Y млн. грн.	0,4	0,8	1,8	0,9	1,1	1,9	1,5	2	0,6	1,1
Затрати капіталу,X_1 млн. грн.	2	6,8	27	12	17,5	25	21,5	28	4	15,5
Оборот капіталу , X_2 млн. грн.	1	3,2	11	5	8	12	10	11,5	2	5,8

Таблиця 20

Ринкова капіталізація , Y млн. грн.	25	15	30,5	17	25,5	20	29	20	10	27
Оборот капіталу ,X_1 млн. грн.	19,5	6,8	27	12	17,5	10	21,5	13,5	4	15,5
Затрати капіталу, X_2 млн. грн.	12,2	3,2	13	7	11	9	15	5	2	10

Таблиця 21

Чистий дохід, Y млн. грн.	4	2,5	0,7	1,7	5	1,2	6	3	7	0,4
Затрати капіталу,X_1 млн. грн.	100	60	4,5	40	110	15	140	90	165	2
Оборот капіталу , X_2 млн. грн.	80	50	18,5	40	91	25	99	60	110	1,4

Таблиця 22

Чисельність працівників, Y тис. чол.	85	105	7	130	62	42	105	34	142	58
Оборот капіталу , X_1 млн. грн.	18	23	4,5	27	13	10	20	7	27	12
Затрати капіталу,X_2 млн. грн.	90	60	130	19	100	126	80	120	13	110

Таблиця 23

Чисельність працівників, Y тис. чол.	100	106	96	140	60	130	71	65	23	81
Оборот капіталу , X_1 млн. грн.	17	15	12	21	7	20	8	5	2	10
Затрати капіталу,X_2 млн. грн.	14	22	17	23	4	20	8,6	11,5	1,9	16

Таблиця 24

Чисельність працівників, Y тис. чол.	63	65	24	50	106	40	34,7	86	74,5	8
Чистий дохід , X_1 млн. грн.	21	23	7	17	30	12	4	25	28	2
Оборот капіталу , X_2 млн. грн.	31,2	25	45	38	5	40	35	11,5	21	48

Таблиця 25

Чисельність працівників, Y тис. чол.	200	180	150	42	62	212	105	125	110	120
Чистий дохід , X_1 млн. грн.	1,6	1,5	1,3	0,2	0,4	1,8	0,4	0,8	0,6	0,9
Оборот капіталу , X_2 млн. грн.	18	20	15	3	5	26	6	7	8	12

Таблиця 26

Чисельність працівників, Y тис. чол.	65	106	96	140	60	130	40	65	23	30
Чистий дохід , X_1 млн. грн.	1,8	2,6	2,3	2,8	1,9	2,5	1,3	2	0,5	0,7
Оборот капіталу , X_2 млн. грн.	13,4	10	12	18	8	21	7	10	3	5

Таблиця 27

Чисельність працівників, Y тис. чол.	120	65	21	50	106	97	110	86	74,5	4
Оборот капіталу, X_1 млн. грн.	31,2	13,4	4,5	10	20	15	25	18	16,5	2
Затрати капіталу, X_2 млн. грн.	19	14	8,5	12	22	19	21	20	16	5

Таблиця 28

Чисельність працівників, Y тис. чол.	86	120	18	125	62	21,2	105	33,5	142	36
Оборот капіталу, X_1 млн. грн.	18	23	7	25	13	10	20	9	27	12
Затрати капіталу, X_2 млн. грн.	10	21	3	19	13	4	12,2	3,2	25	7

Таблиця 29

Чисельність працівників, Y тис. чол.	65	106	96	140	60	131	71	65,5	23	81
Оборот капіталу, X_1 млн. грн.	10	20	12	23	8	21	13,5	6	4,2	15,5
Затрати капіталу, X_2 млн. грн.	14	2,2	7	16	12	1,6	8,6	11,5	19	5

Таблиця 30

Чисельність працівників, Y тис. чол.	43	110	60	85	80	97	80	60	74,5	20
Оборот капіталу, X_1 млн. грн.	31	4,5	20	13,7	16,5	6,8	10	25	15	33
Затрати капіталу, X_2 млн. грн.	19	4	22	9,9	6,1	8	13,2	12	17	25

Завдання 3

Перевірити фактори X_1, X_2, X_3 на наявність мультиколінеарності.

Відомі такі дані про діяльність компаній певної країни.

Таблиця 1

Оборот капіталу, X_1 , млн. гр.	31,2	13,4	4,5	10	20	15	14	18	16,5	25
Затрати капіталу, X_2, млн. гр.	19	5	2	8	22	12	10	20	16	25
Чисельність працівників, X_3 ,тис. чол.	43	10	5	15	66	20	35	50	75	40

Таблиця 2

Оборот капіталу, X_1 , млн. гр.	18	5	7	27	13	10	4	7	24	12
Затрати капіталу, X_2, млн. гр.	20	3	8	19	10	12,6	4	5	13	7
Чисельність працівників, X_3 ,тис. чол.	85	34	50	110	62	70	10,5	34	100	40

Таблиця 3

Оборот капіталу, X_1 , млн. гр.	13,4	20	12	10	13	21	13,5	18	4,2	8
Затрати капіталу, X_2 , млн. гр.	14	22	7	15	12	16	8,6	11,5	1,9	5
Чисельність працівників, X_3 ,тис. чол.	65	106	51	40	60	130	71	65	23	30

Таблиця 4

Оборот капіталу, X_1, млн. гр.	31,2	13,4	25	10	20	40	13,5	18	4,2	2
Затрати капіталу, X_2,млн. гр.	40	14	70	35	22	58	8,6	50	1,9	10
Чисельність працівників, X_3,тис. чол.	350	65	300	100	200	265	37	120	25	35

Таблиця 5

Оборот капіталу, X_1 , млн. гр.	18	20	7	27	13	29	13,5	7	4,2	10
Затрати капіталу, X_2, млн. гр.	60	20	50	80	70	126	40	100	1,9	7
Чисельність працівників, X_3 ,тис. чол.	12	20	15	35	26	33	25	32	10	20

Таблиця 6

Оборот капіталу, X_1, млн. гр.	13,4	20	5	18	8	21	13,5	13,4	4,2	3
Затрати капіталу, X_2, млн. гр.	14	18	7	15	12	16	8,6	11,5	1,9	5
Чисельність працівників, X_3 ,тис. чол.	21	37	15	25	18	23,2	12	29	10	5

Таблиця 7

Оборот капіталу ,X_1 млн. грн.	55	13,4	4,5	10	40	180	140	80	165	110
Затрати капіталу, X_2 млн. грн.	42	14	30	4,8	22	105	99	55	61	71
Чистий дохід , X_3 млн. грн.	3,6	0,8	0,7	1,7	2,6	7,5	4	1,6	7	5,5

Таблиця 8

Оборот капіталу ,X_1 млн. грн.	18	20	7	27	13	10	20	7	25	12
Затрати капіталу, X_2 млн. грн.	20	23	3	19	13	12,6	16	3,2	18	7
Чистий дохід , X_3 млн. грн.	1,1	1,5	0,3	1,9	1	0,7	1,7	0,5	1,8	0,9

Таблиця 9

Оборот капіталу ,X_1 млн. грн.	4	7	9	18	13	21	13,5	13,4	21	15,5
Затрати капіталу, X_2 млн. грн.	20	17	7	15	12	1,6	8,6	11,5	1,9	5
Чистий дохід , X_3 млн. грн.	1,8	1,5	0,9	1,1	1,3	0,9	1,3	1,4	0,6	0,7

Таблиця 10

Оборот капіталу ,X_1 млн. грн.	0,9	0,8	0,7	1,7	2,6	1,2	4	1,6	7	0,4
Затрати капіталу, X_2 млн. грн.	31,2	13,4	4,5	10	20	15	140	18	165	2
Чистий дохід , X_3 млн. грн.	50	65	24	150	250	650	347	86	745	500

Таблиця 11

Оборот капіталу ,X_1 млн. грн.	0,3	0,4	1,3	1,9	1,9	1,4	0,4	0,8	1,8	0,9
Затрати капіталу, X_2 млн. грн.	4	3	7	18	25	20	5	7	15	12
Чистий дохід , X_3 млн. грн.	20	4	120	180	225	212	45	34	142	96

Таблиця 12

Оборот капіталу ,X_1 млн. грн.	1,8	2,6	0,9	1,1	1,9	0,4	1,3	2	0,5	0,7
Затрати капіталу, X_2 млн. грн.	13,4	10	20	18	13	21	23	13,4	25	26
Чистий дохід , X_3 млн. грн.	65	40	96	75	60	130	71	65	120	81

Таблиця 13

Оборот капіталу ,X_1 млн. грн.	31	4,5	110	137	165	150	75	19	51	125
Затрати капіталу, X_2 млн. грн.	19	18,5	47	60	90	80	38	12	25	70
Ринкова капіталізація, X_3 млн. грн.	41	39	20	15	5	10	26	45	31	25

Таблиця 14

Оборот капіталу ,X_1 млн. грн.	180	100	60	137	165	68	13,4	110	45	3
Затрати капіталу, X_2 млн. грн.	110	45	22	99	61	30	13,2	80	40	5
Чистий дохід , X_3 млн. грн.	3	2,1	0,9	3,2	2,5	1,6	1,3	2	0,5	0,3

Таблиця 15

Затрати капіталу,X_1 млн. грн.	23	20	15	18	2	27	9,8	6,8	12,4	25
Оборот капіталу , X_2 млн. грн.	14	17	10	13	1,4	19	5	3,2	7	24
Ринкова капіталізація , X_3 млн. грн.	40,5	38,5	26,5	37	5	42	10	8	15	48

Таблиця 16

Оборот капіталу ,X_1 млн. грн.	2	10	28	18	15	5	9,8	20	25	23
Затрати капіталу, X_2 млн. грн.	2	4,8	26	10	18	8	12,6	16	20	23
Чистий дохід , X_3 млн. грн.	0,1	0,2	1,3	0,3	0,4	0,5	0,6	0,8	0,9	1,2

Таблиця 17

Оборот капіталу ,X_1 млн. грн.	120	50	137	72	165	2	7	100	90	30
Затрати капіталу, X_2 млн. грн.	90	45	99	60	110	1,4	8	81	65	25
Чистий дохід , X_3 млн. грн.	27	13	30	17	36	5	7	25	20	10

Таблиця 18

Затрати капіталу,X_1 млн. грн.	40	125	137	18	165	5	55	90	110	100
Оборот капіталу , X_2 млн. грн.	30	70	99	20	110	1,4	8	60	80	40
Ринкова капіталізація , X_3 млн. грн.	2,6	5	7,5	1,6	7	0,4	1,3	4,5	6	3,5

Таблиця 19

Затрати капіталу,X_1 млн. грн.	2	6,8	27	12	17,5	25	21,5	28	4	15,5
Оборот капіталу , X_2 млн. грн.	1	3,2	11	5	8	12	10	11,5	2	5,8
Ринкова капіталізація , X_3 млн. грн.	33	20	6	15	10	5	7	4	28	13

Таблиця 20

Оборот капіталу ,X_1 млн. грн.	19,5	6,8	27	12	17,5	10	21,5	13,5	4	15,5
Затрати капіталу, X_2 млн. грн.	12,2	3,2	13	7	11	9	15	5	2	10
Чистий дохід , X_3 млн. грн.	2	0,8	2,4	0,9	1,6	1,5	2,1	1,3	0,3	2,1

Таблиця 21

Затрати капіталу,X_1 млн. грн.	100	60	4,5	40	110	15	140	90	165	2
Оборот капіталу , X_2 млн. грн.	80	50	18,5	40	91	25	99	60	110	1,4
Чисельність працівників,X_3 тис. чол.	4	2,3	0,7	1,7	5	1,2	6	3,5	7	0,4

Таблиця 22

Оборот капіталу , X_1 млн. грн.	18	23	4,5	27	13	10	20	7	27	12
Затрати капіталу,X_2 млн. грн.	90	60	130	19	100	126	80	120	13	110
Чистий дохід , X_3 млн. грн.	9	11	0,7	15	5	1,2	14	3,5	18	7

Таблиця 23

Оборот капіталу , X_1 млн. грн.	17	15	12	21	7	20	8	5	2	10
Затрати капіталу,X_2 млн. грн.	14	22	17	23	4	20	8,6	11,5	1,9	16
Чистий дохід , X_3 млн. грн.	1,8	2	1,5	2,7	0,8	2,4	1,3	1	0,3	1,6

Таблиця 24

Чистий дохід , X_1 млн. грн.	21	23	7	17	30	12	4	25	28	2
Оборот капіталу , X_2 млн. грн.	31,2	25	45	38	5	40	35	11,5	21	48
Ринкова капіталізація ,X_3 млн. грн.	35	25	15	27	45	24	13	36,8	30	1,6

Таблиця 25

Чистий дохід , X_1 млн. грн.	1,6	1,5	1,3	0,2	0,4	1,8	0,4	0,8	0,6	0,9
Оборот капіталу , X_2 млн. грн.	18	20	15	3	5	26	6	7	8	12
Ринкова капіталізація ,X_3 млн. грн.	37	35	35	20	26	40	32,5	32	30,5	29,5

Таблиця 26

Чистий дохід , X_1 млн. грн.	1,8	2,6	2,3	2,8	1,9	2,5	1,3	2	0,5	0,7
Оборот капіталу , X_2 млн. грн.	13,4	10	12	18	8	21	7	10	3	5
Ринкова капіталізація ,X_3 млн. грн.	25	37	30	50	29	45	29	23	15	20

Таблиця 27

Оборот капіталу, X_1 млн. грн.	31,2	13,4	4,5	10	20	15	25	18	16,5	2
Затрати капіталу, X_2 млн. грн.	19	14	8,5	12	22	19	21	20	16	5
Ринкова капіталізація ,X_3 млн. грн.	41	28	21	23	37,5	32	37	36,8	30	15

Таблиця 28

Оборот капіталу, X_1 млн. грн.	18	23	7	25	13	10	20	9	27	12
Затрати капіталу, X_2 млн. грн.	10	21	3	19	13	4	12,2	3,2	25	7
Ринкова капіталізація ,X_3 млн. грн.	20	15	35	12	26	33	15	32	10	29,8

Таблиця 29

Оборот капіталу, X_1 млн. грн.	10	20	12	23	8	21	13,5	6	4,2	15,5
Затрати капіталу, X_2 млн. грн.	14	2,2	7	16	12	1,6	8,6	11,5	19	5
Ринкова капіталізація ,X_3 млн. грн.	20	37	30	15	18	46	29,2	25	10	27

Таблиця 30

Оборот капіталу, X_1 млн. грн.	31	4,5	20	13,7	16,5	6,8	10	25	15	33
Затрати капіталу, X_2 млн. грн.	19	4	22	9,9	6,1	8	13,2	12	17	25
Ринкова капіталізація ,X_3 млн. грн.	20	39	24	30	36	35	26	21	31	15

Завдання 4

1. Дослідити вхідні змінні на наявність гетероскедастичності за тестом Гольдфельда-Квандта. Необхідні дані взяти з умови завдання 3.
2. Дослідити побудовану в завданні 2 модель на наявність гетероскедастичності, використовуючи тест рангової кореляції Спірмена.

Завдання 5

Дослідити модель, побудовану в завданні 1 на наявність автокореляції залишків з допомогою критерію Дарбіна-Уотсона. Зробити висновок.

Приклади тестових завдань

1.Визначте правильне означення економетрики

A. Це наука, яка займається обробкою статистичних даних економічних об'єктів.

B. Це наука, яка займається побудовою теоретичних моделей економічних об'єктів.

C. Це наука, яка вивчає кількісні закономірності і взаємозалежності економічних процесів за допомогою математико-статистичних методів і моделей.

D. інша відповідь.

2.Визначте правильне означення математичної моделі

A. Графік процесу.

B. Масштабна копія об'єкта .

C. Система математичних рівнянь.

D. Інша відповідь.

3.Що таке специфікація моделі

A.Побудова моделі.

B. Верифікація моделі.

C. Визначення структури моделі.

D. Інша відповідь.

4.Регресійний аналіз це

A. Дослідження функціональної залежності

B. Дослідження стохастичної залежності.

C. Дослідження кореляційної залежності.

D. Інша відповідь.

6.Як називається рівняння регресії $\hat{y} = a + b \cdot x$

A. Лінійним.

B. Гіперболічним.

C. Степеневим.

D. Показниковим.

7.Як знаходиться SST

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

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

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

D. інша відповідь

8. Для оцінки значущості лінійного рівняння регресії розраховують

A. F – критерій Фішера .

B. t – критерій Стьюдента.

C. коефіцієнт детермінації .

D. інша відповідь.

12. ЛІТЕРАТУРА

Базова

1. Васильченко І. П. Вища математика для економістів підручник / І. П. Васильченко – К. : Знання –Прес, 2002 .– 454с.
2. Высшая математика для экономистов: учебник для студ. вузов, обучающихся по экон. спец. / Н. Ш. Кремер и др., под ред. проф. Н. Ш. Кремера.– М.:ЮНИТИ, 2006. – 479 с.
3. Исследование операций в экономике: уч. пособ. для вузов Под ред. проф. Н. М. Кремера.– М.: ЮНИТИ,2006. – 407с.
4. Ковалев В. Г. Математическое программирование (линейные задачи) учеб. пособие / В. Г. Ковалев, А. Р. Наринян, В. А. Поздев – К.: Европ. ун– т, 2004. – 170 с.
5. Красс М. С. Математика в экономике. Математические методы и модели: учебник/ М. С. Красс, Б. П. Чупринов – М.: Финансы и статистика, 2007.– 544с.
6. Лавріненко Н. М. Основи економіко – математичного моделювання: навч. посіб./ Н. М. Лавріненко, С. М. Латинін, В. В. Фортуна, О. І. Бескровний – Л.: Магнолія, 2006, 2010.–540 с.
7. Орлова И.В.Экономико –математические методы и модели: компьютерное моделирование: Учеб пособие / И. В. Орлова, В. А. Половников – М.: Вузовский учебник, 2007.–365с.
8. Практикум по эконометрике: учеб пособие / И. И. Елисеева и др.; под ред И. И. Елисеевой.– 2-е изд. перераб. и доп. – М.: Финансы и статистика, 2006.– 192 с.
9. Таха Х. Введение в исследование операций / Х. Таха; пер. с англ. и редакция А. А. Минько. – М.: Вильямс, 2007. – 912 с.
10. Эконометрика: учебник / И. И. Елисеева и др. под ред И. И. Елисеевой.– 2- е изд. перераб. и доп. М.: Финансы и статистика, 2006. – 576с.

Допоміжна

11. Вітлінський В. В. Математичне програмування: навч.- метод. посібник для самост. вивч. дисц. /В. В. Вітлінський, С. І. Наконечний, Т. О. Терещенко.–К.: КНЕУ, 2001. 248 с.
12. Конюховский П. В. Математические методы исследования операций / П. В. Конюховский. – СПб Питер: 2001.–192с.
13. Корольов О. А. Економетрія : навч. посіб. / О. А. Корольов.– К.: КНЕУ, 2000. – 660с.
14. Кулинич Е. И. Эконометрия / Е. И. Кулинич.– М.: Финансы и статистика, 2001.– 304с.
15. Наконечний С. І. Економетрія: підручник /С. І. Наконечний ,Т. О. Терещенко, Т. П. Романюк. – Вид. 4-те доп. та перероб. К.: КНЕУ, 2006.– 528 с.

Інформаційні ресурси

16. Вища освіта України і Болонський процес / Навчальна програма. – Київ Тернопіль: ТДПУ ім. В. Гнатюка 2004.– 18с.
17. ІСУЯ 7.5.1– 03.01/УН «Загальні вимоги до організації процесу проведення навчальних занять».
18. ІСУЯ 7.5.1 – 03.02 / УН «Загальні вимоги до організації методичного забезпечення виконання індивідуальних завдань з дисциплін».
19. ІСУЯ 7.5.1 – 03.03 / УН «Загальні вимоги до організації виконання індивідуальних завдань».
20. ІСУЯ 7.5.1 – 03.04 / УН «Загальні вимоги до організації СРС».
21. ІСУЯ 7.5.1 – 03.05 / УН «Загальні вимоги до організації НДРС».
22. ІСУЯ 7.5.1 – 03.06 / УН «Загальні вимоги до організації поточного контролю».
23. ІСУЯ 7.5.1 – 03.07 / УН «Загальні вимоги до організації підсумкового контролю».
24. ІСУЯ 7.5.1 – 03.08 / УН «Критерії забезпеченості дисциплін навчально-методичною літературою».
25. ІСУЯ 7.5.1 – 03.09 / УН «Загальні вимоги до видання навчально-методичної літератури».

ДОДАТОК 1

Таблиця 5% - ого и 1% - ого рівнів ймовірності коефіцієнтів автокореляції залишків (r^0)

k	Додатні значення (r_a)		Від’ємні значення (r_a)	
	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
5	0,253	0,297	-0,753	-0,798
6	0,354	0,447	- 0,708	-0,863
7	0,370	0,510	-0,674	-0,799
5	0,371	0,531	-0,625	-0,764
9	0,366	0,533	-0,593	-0,737
10	0,360	0,525	-0,564	-0,705
11	0,353	0,515	-0,539	-0,679
12	0,348	0,348	-0,516	-0,655
13	0,341	0,495	-0,497	-0,634
14	0,335	0,485	-0,479	-0,615
15	0,328	0,475	-0,462	-0,597
20	0,299	0,432	-0,399	-0,524
25	0,276	0,398	-0,356	-0,473
30	0,257	0,370	-0,324	-0,433
35	0,242	0,347	-0,300	-0,401
40	0,229	0,329	-0,279	-0,376
45	0,218	0,313	-0,262	-0,256
50	0,208	0,301	-0,248	-0,339

ДОДАТОК 2

Критичні значення DW_L і DW_U для коефіцієнта автокореляції залишків критерія Дарбіна-Уотсона для $\alpha = 0{,}05$

Число спостережень	Ч и с л о ф а к т о р і в									
	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
	d_H	d_B	d_H	d_B	d_H	d_B	d_H	d_B	d_H	d_B
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13	0,30	2,59		
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41	0,24	2,82
11	0,93	1,32	0,76	1,60	0,60	1,93	0,44	2,28	0,32	2,65
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	2,18	0,38	2,51
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	2,09	0,45	2,39
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	2,03	0,51	2,30
15	1,08	1,36	0,95	1,54	0,81	1,75	0,69	1,98	0,56	2,22
16	1,11	1,37	0,98	1,54	0,86	1,73	0,73	1,94	0,62	2,16
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,66	2,10
18	1,16	1,39	1,05	1,54	0,93	1,70	0,82	1,87	0,71	2,06
19	1,18	1,40	1,07	1,54	0,97	1,69	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,89	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,65	1,04	1,77	0,95	1,80
26	1,30	1,46	1,23	1,55	1,14	1,65	1,06	1,76	0,98	1,87
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,75	1,00	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,37	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,53	1,70	1,49	17,4	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,52	1,79	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,58	1,72	1,55	1,75	1,53	1,77
90	1,64	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,65	1,69	1,62	1,71	1,60	1,73	1,58	1,76	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78
150	1,72	1,75	1,71	1,76	1,69	1,77	1,68	1,79	1,67	1,80

ПРОДОВЖЕННЯ ДОДАТКУ 2

Число спостережень	Ч и с л о ф а к т о р і в									
	$m=6$		$m=7$		$m=8$		$m=9$		$m=10$	
	d_H	d_B	d_H	d_B	d_H	d_B	d_H	d_B	d_H	d_B
11	0,12	2,89								
12	0,16	2,67	0,11	3,06						
13	0,21	2,49	0,14	2,64	0,09	3,18				
14	0,26	2,35	0,18	2,67	0,12	2,96	0,08	3,29		
15	0,30	2,24	0,23	2,53	0,16	2,82	0,11	3,10	0,07	3,37
16	0,35	2,15	0,27	2,42	0,20	2,68	0,14	2,94	0,09	3,20
17	0,39	2,08	0,31	2,31	0,24	2,57	0,18	2,81	0,13	3,05
18	0,44	2,02	0,37	2,24	0,28	2,47	0,22	2,70	0,16	2,93
19	0,48	1,96	0,40	2,17	0,32	2,36	0,26	2,60	0,20	2,81
20	0,52	1,92	0,44	2,11	0,36	2,31	0,29	2,51	0,23	2,71
21	1,55	1,88	0,47	2,06	0,40	2,24	0,33	2,43	0,27	2,63
22	1,57	1,85	0,51	2,02	0,44	2,19	0,37	2,37	0,30	2,55
23	1,62	1,82	0,55	1,98	0,47	2,14	0,40	2,31	0,34	2,48
24	1,65	1,79	0,58	1,94	0,51	2,10	0,44	2,26	0,38	2,42
25	1,68	1,78	0,61	1,89	0,54	2,06	0,47	2,21	0,41	2,37
26	0,71	1,76	0,64	1,89	0,57	2,03	0,51	2,17	0,44	2,31
27	0,74	1,74	0,67	1,87	0,60	2,00	0,54	2,13	0,47	2,27
28	0,76	1,73	0,70	1,85	0,66	1,97	0,57	2,10	0,50	2,23
29	0,79	1,72	0,72	1,83	0,69	1,95	0,60	2,07	0,53	2,19
30	0,81	1,71	0,76	1,82	0,69	1,93	0,62	2,04	0,56	2,16
31	0,83	1,70	0,77	1,80	0,71	1,91	0,65	2,02	0,59	2,13
32	0,86	1,69	0,79	1,78	0,73	1,89	0,67	2,00	0,62	2,10
33	0,88	1,68	0,82	1,78	0,76	1,87	0,70	1,98	0,64	2,00
34	0,90	1,68	0,84	1,77	0,78	1,86	0,72	1,96	0,67	2,06
35	0,91	1,67	0,86	1,76	0,80	1,85	0,74	1,94	0,69	2,04
36	0,93	1,67	0,88	1,75	0,82	1,84	0,77	1,93	0,71	2,01
40	1,00	1,65	0,95	1,72	0,90	1,80	0,84	1,88	0,79	1,96
45	1,07	1,64	1,02	1,71	0,97	1,77	0,93	1,83	0,86	1,90
50	1,12	1,64	1,06	1,69	1,04	1,75	1,00	1,81	0,96	1,96
55	1,17	1,64	1,13	1,69	1,10	1,73	1,06	1,79	1,02	1,84
60	1,21	1,64	1,18	1,68	1,14	1,73	1,11	1,77	1,07	1,82
65	1,25	1,64	1,22	1,68	1,19	1,72	1,15	1,76	1,12	1,80
70	1,28	1,65	1,25	1,68	1,22	1,72	1,19	1,75	1,16	1,79
75	1,31	1,65	1,28	1,68	1,26	1,72	1,23	1,75	1,20	1,79
80	1,34	1,65	1,31	1,68	1,29	1,71	1,26	1,75	1,23	1,77
85	1,36	1,66	1,34	1,69	1,31	1,71	1,29	1,74	1,26	1,77
90	1,38	1,66	1,36	1,69	1,34	1,71	1,31	1,74	1,29	1,77
95	1,40	1,67	1,38	1,69	1,36	1,72	1,34	1,74	1,31	1,77
100	1,42	1,67	1,40	1,69	1,38	1,72	1,36	1,74	1,34	1,77
150	1,54	1,71	1,53	1,72	1,52	1,74	1,50	1,75	1,49	1,77

ДОДАТОК 3

Значення критерія Пірсона $\chi^2(\alpha, n)$

n \ α	0,995	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,005
1	0,00098	0,00016	0,00098	0,0039	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,61	5,99	7,38	9,21	10,60
3	0,0717	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,34	12,84
4	0,207	0,297	0,484	0,711	1,064	7,78	9,49	11,14	13,28	14,86
5	0,412	0,554	0,831	1,15	1,61	9,24	11,07	12,83	15,09	16,75
6	0,676	0,872	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	0,989	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,96
9	1,73	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,73	26,76
12	3,07	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	34,27
18	6,26	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	37,16
20	7,43	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	40,00
24	9,89	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98	45,56
30	13,79	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	53,67
40	20,71	22,16	24,43	26,51	29,05	51,81	55,76	59,34	63,69	66,77
60	35,53	37,48	40,48	43,19	46,46	74,40	79,08	83,30	88,38	91,95
80	51,17	53,54	57,15	60,39	64,28	96,58	101,88	106,63	112,33	116,32
100	67,33	70,06	74,22	77,93	82,36	118,50	124,34	129,56	135,81	140,17
120	83,85	86,92	91,58	95,70	100,62	140,23	146,57	152,21	158,95	163,64

ДОДАТОК 4

95% - квантилі розподілу Фішера $F(\alpha, k_1, k_2)$

(k_1, k_2 –число ступенів свободи відповідно чисельника і знаменника)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

Продовження додатку 4

k_2 \ k_1	12	15	20	24	30	40	60	120	∞
1	244	246	248	249	250	251	252	253	254
2	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,60	2,53	5,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
30	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

ДОДАТОК 5

Квантилі розподілу Стьюдента $t_{\alpha}(n)$

n \ α	0,80	0,60	0,50	0,40	0,20	0,10	0,05	0,02	0,01
1	0,325	0,727	1,000	1,376	3,078	6,314	12,706	31,821	63,657
2	0,289	0,617	0,816	1,061	1,886	2,920	4,303	6,965	9,925
3	0,277	0,584	0,765	0,978	1,638	2,353	3,182	4,541	5,841
4	0,271	0,569	0,741	0,941	1,533	2,132	2,776	3,747	4,604
5	0,267	0,559	0,727	0,920	1,476	2,015	2,571	3,365	4,032
6	0,265	0,553	0,718	0,906	1,440	1,943	2,447	3,143	3,707
7	0,263	0,549	0,711	0,896	1,415	1,895	2,365	2,998	3,499
8	0,262	0,546	0,706	0,889	1,397	1,860	2,306	2,896	3,355
9	0,261	0,543	0,703	0,883	1,383	1,833	2,262	2,821	3,250
10	0,260	0,542	0,700	0,879	1,372	1,812	2,228	2,764	3,169
11	0,260	0,540	0,697	0,876	1,363	1,796	2,201	2,718	3,306
12	0,259	0,539	0,695	0,873	1,356	1,782	2,179	2,681	3,055
13	0,259	0,538	0,694	0,870	1,350	1,771	2,160	2,650	3,012
14	0,258	0,537	0,692	0,868	1,345	1,761	2,145	2,624	2,977
15	0,258	0,536	0,691	0,866	1,341	1,753	2,131	2,602	2,947
16	0,258	0,535	0,690	0,865	1,337	1,746	2,120	2,583	2,921
17	0,257	0,534	0,689	0,863	1,333	1,740	2,110	2,567	2,898
18	0,257	0,534	0,688	0,862	1,330	1,734	2,101	2,552	2,878
19	0,257	0,533	0,688	0,861	1,328	1,729	2,093	2,539	2,861
20	0,257	0,533	0,687	0,860	1,325	1,725	2,086	2,528	2,845
21	0,257	0,532	0,686	0,859	1,323	1,721	2,080	2,518	2,831
22	0,256	0,532	0,686	0,858	1,321	1,717	2,074	2,508	2,819
23	0,256	0,532	0,685	0,858	1,319	1,714	2,069	2,500	2,807
24	0,256	0,531	0,685	0,857	1,318	1,711	2,064	2,492	2,797
25	0,256	0,531	0,684	0,856	1,316	1,708	2,060	2,485	2,787
26	0,256	0,531	0,684	0,856	1,315	1,706	2,056	2,479	2,779
27	0,256	0,531	0,684	0,855	1,314	1,703	2,052	2,473	2,771
28	0,256	0,530	0,683	0,855	1,313	1,701	2,048	2,467	2,763
29	0,256	0,530	0,683	0,854	1,311	1,699	2,045	2,462	2,756
30	0,256	0,530	0,683	0,854	1,310	1,697	2,042	2,457	2,750
40	0,255	0,529	0,681	0,851	1,303	1,684	2,021	2,423	2,704
60	0,254	0,527	0,679	0,848	1,296	1,671	2,000	2,390	2,660
100	0,254	0,526	0,677	0,845	1,290	1,660	1,984	2,364	2,626
200	0,254	0,525	0,676	0,843	1,286	1,652	1,972	2,345	2,601
∞	0,253	0,524	0,675	0,842	1,282	1,645	1,960	2,326	2,576

Навчальне видання

Серебреников *Вадим Михайлович,*
Квітка *Тетяна Володимирівна,*
Конайгора *Ольга Константинівна*

Кафедра вищої математики та інформаційних систем

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОДЕЛІ:
ЕКОНОМЕТРИКА

Навчально-методичний посібник для вивчення дисципліни

розділ «Кореляційно-регресійний аналіз»

Формат 60х84/8. Ум. др. арк.

Донецький національний університет
економіки і торгівлі імені
Михайла Туган-Барановського
50042, Дніпропетровська обл.,
м.Кривий Ріг, вул. Курчатова, 13.
Свідоцтво суб'єкту видавничої
Справи ДК № 4929 від 07.07.2015 р.