

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО”
НАВЧАЛЬНО-НАУКОВИЙ
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ

**П.І.Бідюк
В.Я.Данилов
О.Л.Жиров**

ПРИКЛАДНА СТАТИСТИКА

Навчальний посібник

Рекомендовано Методичною радою КПІ імені Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за спеціальностями 122 Комп'ютерні науки, 124 Системний аналіз

Електронне мережне видання

Київ
КПІ ім. Ігоря Сікорського
2023

Рецензент

Цюцюра С.В., доктор технічних наук, професор
кафедри інформаційних технологій
КНУБА

Відповідальний
редактор

Кузнецова Н.В., доктор технічних наук, доцент

*Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 6 від 30.03.2023 р.)
за поданням Вченої ради факультету/навчально-наукового інституту
(протокол № 3 від 27.03.2023 р.)*

Навчальний посібник призначений для засвоєння здобувачами вищої освіти основ прикладної статистики, як інструментарію статистичного аналізу даних. В навчальному посібнику викладено теорію, наведені важливі приклади використання методів та методології прикладної статистики для побудови статистичних моделей, побудови інтервального прогнозу поведінки системи та оцінювання ризиків.

Призначений для здобувачів ступеня бакалавра спеціальностей 122 Комп'ютерні науки, 124 Системний аналіз.

Реєстр. № НП 22/23-604. Обсяг 8,5 авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
<https://kpi.ua>

Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р

@ П.І.Бідюк, В.Я.Данилов, О.Л.Жиров
@КПІ ім. Ігоря Сікорського, 2023

ЗМІСТ

ВСТУП	7
Розділ 1. ЛІНІЙНА РЕГРЕСІЯ.....	8
1.1.Оцінка тісноти та значущості зв'язку між змінними моделі.....	11
1.1.1.Коефіцієнт детермінації	11
1.1.2. Коефіцієнт кореляції	12
1.1.3.Перевірка адекватності моделі за критерієм Фішера	12
1.1.4.Зв'язок між коефіцієнтом детермінації (R^2) та F-відношенням Фішера.....	13
1.2.Покрокова побудова лінійної моделі з інтервальним прогнозом.....	14
1.3.Покрокове виконання комп'ютерного практикуму до розділу 1.....	15
1.3.1. Виконання практикуму до розділу 1 на Excel.....	22
1.3.2. Виконання комп'ютерного практикуму до розділу 1 з використанням блоку аналізу Excel.....	26
1.3.3. Склад пакету аналізу Excel	26
1.3.4.Використання таблиць Excel разом з різними програмними продуктами	27
1.3.5. Виконання завдання 1 з використанням блоку аналізу, інтервальний прогноз для заданого рівня надійності	29
1.3.6.Контрольні питання для підготовки до виконання практикуму.....	33
1.3.7.Варіанти індивідуальних даних для практикуму №1.....	35
Розділ 2. НЕЛІНІЙНА ПАРНА РЕГРЕСІЯ.....	39
2.1. Короткі теоретичні відомості.....	39
2.1.1. Мета та основні завдання комп'ютерного практикуму до розділу 2	42
2.1.2.Порядок і рекомендації щодо виконання практикуму №2.....	43
2.2.Візуальний аналіз вибору моделі.....	47
2.3.Інтервальний прогноз за нелінійною моделлю	50
2.4.Приклад аналізу нелінійної залежності	50
2.5.Варіанти індивідуальних даних для практикуму №2.....	54
2.6. Контрольні питання для підготовки до виконання практикуму.....	58
Розділ 3. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ.....	61
3.1. Короткі теоретичні відомості	61
3.1.1. Гетероскедастичність та її наслідки	62
3.1.2. Виявлення гетероскедастичності.....	63
3.1.3. Тест рангової кореляції Спірмена	63
3.1.4. Тест Гольдфельда –Квандта на гетероскедастичність.....	63

3.2. Мультиколінійність.....	63
3.2.1. Ознаки наявності мультиколінійності.....	63
3.2.2. Наслідки мультиколінійності	64
3.2.3. Правило відбору факторів в множині мультиколінійних.....	64
3.3. Перевірка значущості коефіцієнта детермінації.....	65
3.4. Перевірка значущості коефіцієнта кореляції	65
3.5. Оцінка статистичної значущості параметрів моделі..	66
3.6. Прогнозування за загальною лінійною моделлю	67
3.6.1 Мета та основні завдання комп'ютерного практикуму до розділу 3	68
3.6.2. Специфікація моделі	68
3.6.3. Оцінка рівняння регресії	70
3.6.4. Матриця парних коефіцієнтів кореляції.....	72
3.6.5. Модель регресії в стандартному масштабі.....	75
3.6.6. Порядок і рекомендації щодо виконання практикуму.....	75
3.6.7. Аналіз параметрів рівняння множинної регресії	76
3.6.8. Показники тісноти зв'язку факторів з результатами.....	77
3.6.9. Частинні коефіцієнти еластичності...	77
3.6.10. Стандартизовані частинні коефіцієнти регресії.....	78
3.6.11. Частинні коефіцієнти кореляції	79
3.6.12. Індекс множинної кореляції (множинний коефіцієнт кореляції).....	80
3.6.13. Коефіцієнт детермінації.....	81
3.6.14. Оцінка значення результативної ознаки при заданих значеннях факторів	82
3.6.15. Перевірка гіпотез щодо коефіцієнтів рівняння регресії,.....	83
3.6.16. Перевірка загальної якості рівняння множинної регресії.....	84
3.6.17. Перевірка на наявність гетероскедастичності.....	86
3.6.18. Перевірка на наявність автокореляції залишків	87
3.6.19. Виявлення автокореляції	88
3.6.20. Завдання на дослідження мультиколінійності алгоритмом Феррара—Глобера	89
3.6.21. Порядок і рекомендації щодо виконання практикуму.....	90
3.6.22. Контрольні питання для підготовки до виконання практикуму №3	95
3.6.23. Варіанти вихідних даних завдання №3	95

Розділ 4. ПОБУДОВА ТА ДОСЛІДЖЕННЯ НЕОКЛАСИЧНИХ ВИРОБНИЧИХ ФУНКЦІЙ	96
4.1. Короткі теоретичні відомості	96
4.2. Економічний аналіз неокласичних систем	96
4.3. Мета та основні завдання практикуму №4	99

4.4. Порядок і рекомендації щодо виконання практикуму	102
4.5.Індивідуальні вихідні дані варіантів практикуму до розділу №4	103
4.6.Контрольні питання для підготовки до виконання практикуму	104
 Розділ 5. ОДНОЧАСНІ СТРУКТУРНІ РІВНЯНЯ.....	105
5.1.Короткі теоретичні відомості.....	105
5.2.Алгоритм вибору методу оцінювання	107
5.2.1.Мета та основні завдання комп'ютерного практикуму №5.....	108
5.2.2.Порядок і рекомендації щодо виконання	108
5.2.3. Вихідні дані індивідуальних завданю №5	109
Контрольні питання для підготовки до виконання практикуму №5.....	112
 Розділ 6. МЕТОДИКА БОКСА-ДЖЕНКІНСА ПОБУДОВИ МАТЕМАТИЧНОЇ МОДЕЛІ	114
6.1.Аналіз процесу	114
6.2.Попередня обробка даних	115
6.3.Обробка екстремальних (аномальних) значень	116
6.4.Аналіз наявності нелінійностей	118
6.5.Формування елементів структури моделі	121
6.6.Оцінювання коефіцієнтів моделей-кандидатів	126
6.7.Діагностика моделей – вибір кращої з множини оцінених кандидатів	129
6.8. Приклади побудови математичних моделей за допомогою експериментальних даних	135
6.8.1.Регресійна модель індексу споживчих цін	134
6.8.2.Модель авторегресії для відхилень ІСЦ від середнього	139
6.8.3.Авторегресія з ковзним середнім для ІСЦ	141
6.8.4. Врахування впливу на ІСЦ агрегату МЗ.....	143
6.9.Визначення впливу ВВП на ІСЦ	143
6.10.Запитання і завдання до розділу №6	145
ДОДАТКИ	147
Д1. Приклади і методика побудови задач експрес-тестування та модульної контрольної роботи.....	148
Д.2. Державна статистична служба України.....	152
Д.3.Національні статистичні служби Європи як першоджерела реальних даних для засвоєння і використання вивчених методів прикладної статистики.....	152
Д.4.Програми і середовища прикладної статистики	154
Д.5. Деякі джерела загальнооступних вихідних кодів програм з прикладної статистики	156
Д.6. Англо-український словник з прикладної статистики	157

Д.7. Статистичні таблиці.....	180
Д.7.1 Значення функції Лапласа	180
Д.7.2 Таблиця значень F-критерія Фішера на рівні значущості 0.05	182
Д.7.3 Критичні значення t--критерія Стьюдента	182
Д.7.4 Квантилі розподілу Хі-квадрат	183
Д.7.5 Статистики Дарбіна-Уотсонана 5%-ому рівні значущості	184
Д.7.6 Статистики Дарбіна-Уотсонана 1%-ому рівні значущості	186
Література	187

ВСТУП

Прикладна статистика є однією з фундаментальних дисциплін в освітній програмі бакалаврів з системного аналізу, оскільки абсолютна більшість бакалаврських дипломів аналізує часові ряди фінансових і виробничих процесів. Прикладна статистика є складовою частиною системного аналізу, оскільки еквівалентною назвою прикладної статистики є статистичний аналіз даних. Прикладна статистика включає:

- вивчення та оволодіння методами збору та обробки статистичних даних у формі часових перерізів рядів;
- вивчення та оволодіння методами побудови статистичних і математичних моделей на основі функціонального підходу;
- вивчення критеріальної бази для дослідження якості даних і адекватності створених моделей;
- уміння аналізувати вплив структури даних на розв'язання задачі статистичного моделювання, прогнозування та формування статистичних висновків.
- уміння перевіряти відповідність вхідних статистичних даних постановці задачі статистичного і математичного моделювання на основі отриманих даних;
- уміння розробляти статистичні моделі досліджуваних процесів в економіці та фінансах;
- уміння доводити коректність та ефективність побудованих статистичних і математичних моделей на розв'язуванні прикладних задач;
- уміння порівнювати різні методи моделювання, прогнозування, оцінювання ризиків та формування статистичних висновків.

В ІПСА КПІ ім.Ігоря Сікорського основними підручниками з предмету прикладна статистика є [1-3]. Для проведення практикумів з предмету використовуються різні першоджерела, тому виникла потреба в даному навчальному посібнику, в якому створено більше двохсот задач різної тематики на застосування методів прикладної статистики.

Теоретичною основою прикладної статистики є математична статистика, рекомендуємо підручники [4-6], що найтісніше пов'язані з підручниками з прикладної статистики.

Розділ 1. ЛІНІЙНА РЕГРЕСІЯ

Регресійні методи дозволяють описувати зв'язки між факторами, які досліджуються, навіть тоді, коли ці зв'язки не мають точного функціонального опису, в умовах невизначеностей та впливу збурень.

Модель парної регресії запишеться так:

$$y = f(x, u) \quad (1.1)$$

де x – незалежна змінна (пояснювальна, факторна ознака),

y – залежна змінна (результуючий фактор),

u – випадкова складова (невраховані фактори, збурення).

Наприклад, звичайна парна лінійна регресія:

$$y = a_0 + a_1 x + u.$$

В загальному випадку лінійної регресії факторних ознак буде m :

$x_1, \dots, x_m$. Тоді загальна лінійна регресія матиме вигляд:

$$y = a_0 + a_1 x_1 + \dots + a_m x_m + u.$$

Незалежні змінні (факторні ознаки $x_1, \dots, x_m$) найпростіше вважати детермінованими, тобто, що вони є наперед заданими змінними, або вхідними показниками для економічної системи. Це припущення суттєво спрощує викладки, тому що умовні характеристики (наприклад умовні математичні сподівання) співпадатимуть з безумовними.

Випадкові складові u називають ще стохастичними складовими, похибками або залишками. Вони є наслідками помилок спостережень, містять у собі вплив усіх випадкових факторів моделі та факторів, які не входять у модель.

Постановка задачі

Побудова звичайної регресійної моделі полягає в оцінюванні параметрів моделі a_0, a_1 . Отримані оцінки, наприклад з системи нормальних рівнянь, будемо позначати $\hat{a}_0, \hat{a}_1$. Регресійна модель матиме вид

$$\hat{y}(x) = \hat{a}_0 + \hat{a}_1 x,$$

причому значення результуючого показника ми можемо оцінити в довільних точках, в тому числі в точках спостережень:

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i,$$

де $x_i, i = 1, n$ - спостереження.

Згідно з основною ідеєю дисперсійного аналізу, загальна сума квадратів відхилень змінної y від свого середнього значення $\bar{y}$ розкладається на дві частини – «пояснену» і «непояснену»:

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2, \quad (1.2)$$

де $\sum (y_i - \bar{y})^2$ – загальна сума квадратів відхилень;

$\sum (\hat{y}_i - \bar{y})^2$ – пояснена (або факторна сума квадратів відхилень) регресійною моделлю $\hat{y}_i$ сума квадратів відхилень ;

$\sum (y_i - \hat{y}_i)^2$ – залишкова (непояснена) сума квадратів відхилень, що характеризує вплив неврахованих у моделі факторів (сумування проводиться по кількості спостережень $i = \overline{1, n}$). Границі сумування вказувати не будемо.

Схема дисперсійного аналізу має вигляд, представлений у таблиці (1.1)

(n – число спостережень, m – число незалежних факторів X_1-X_m).

Таблиця 1.1

Загальна, факторна та залишкова дисперсії

Компоненти дисперсії	Сума квадратів	Число ступенів вільності	Дисперсія з врахуванням ступенів вільності
Загальна	$\sum (y_i - \bar{y})^2$	$n - 1$	$S_{заг}^2 = \frac{\sum (y_i - \bar{y})^2}{n - 1}$
Факторна	$\sum (\hat{y}_i - \bar{y})^2$	m	$S_{факт}^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{m}$
Залишкова	$\sum (y_i - \hat{y}_i)^2$	$n - m - 1$	$S_{зал}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - m - 1}$

Надалі цю таблицю будемо використовувати і для множинної лінійної регресії з числом факторів $m > 1$.

Перейдемо до звичайної парної лінійної регресії. Для неї кількість факторів $m = 1$.

Визначення дисперсії на один ступінь вільності приводить дисперсії до співставного виду. Порівнюючи факторну й залишкову дисперсії на один ступінь вільності, одержимо величину F - критерію Фішера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{зал}}^2}. \quad (1.3)$$

Фактичне значення - критерію Фішера (1.3) порівнюється з табличним значенням $F_{\text{табл}}(\alpha; k_1; k_2)$ при рівні значущості α й ступенях вільності $k_1 = m$ й $k_2 = n - m - 1$. При цьому, якщо фактичне значення - критерію більше табличного, то визнається статистична значущість рівняння в цілому.

Для парної лінійної регресії $m = 1$, тому

$$F = \frac{S_{\text{факт}}^2}{S_{\text{зал}}^2} = \frac{\sum (\hat{y}_i - \bar{y})^2 \cdot (n - 2)}{\sum (y_i - \hat{y}_i)^2}. \quad (1.4)$$

Коефіцієнтом детермінації парної регресії називається величина

$$r_{xy}^2 = \frac{\sum_{i=1}^n (y_{\text{розрах}} - y_{\text{сер}})^2}{\sum_{i=1}^n (y_{\text{факт}} - y_{\text{сер}})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2}.$$

Величина - критерію пов'язана з коефіцієнтом детермінації r_{xy}^2 , і її можна розрахувати за наступною формулою:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2). \quad (1.5)$$

У парній лінійній регресії оцінюється значущість не тільки рівняння в цілому, але й окремих його параметрів. Із цією метою по кожному з параметрів визначається його стандартна помилка: m_{a_1} і m_{a_0} .

Стандартна помилка коефіцієнта регресії визначається за формулою:

$$m_{a_1} = \sqrt{\frac{S_{\text{зал}}^2}{\sum (x_i - \bar{x})^2}} = \frac{S_{\text{зал}}}{\sigma_x \cdot \sqrt{n}},$$

де $S_{\text{зал}}^2 = \frac{\sum (y_i - \hat{y}_{x_i})^2}{n - 2}$ – залишкова дисперсія на один ступінь вільності.

Величина стандартної помилки разом з t -розподілом Стюдента при $n - 2$ ступенях вільності застосовується для перевірки істотності коефіцієнта регресії й для розрахунків його довірчого інтервалу.

Для оцінки істотності коефіцієнта регресії його величина порівнюється з його стандартною помилкою, тобто визначається фактичне значення t -

критерію Стюдента: $t_{a_0} = \frac{\hat{a}_0}{m_{\hat{a}_0}}, t_{a_1} = \frac{\hat{a}_1}{m_{\hat{a}_1}}$ які потім порівнюються з табличним

значенням при певному рівні значущості α й числі ступенів вільності $(n-2)$. Довірчий інтервал для коефіцієнта регресії береться в виді $(a_0 - t_{\text{табл}} \cdot m_{a_0}; a_0 + t_{\text{табл}} \cdot m_{a_0}), (a_1 - t_{\text{табл}} \cdot m_{a_1}; a_1 + t_{\text{табл}} \cdot m_{a_1})$.

1.1. Оцінка тісноти та значущості зв'язку між змінними моделі

Після вибору виду рівняння регресії та знаходження його параметрів розпочинають наступний етап – кореляційний аналіз, тобто дають оцінку тісноти та значущості зв'язку змінних у регресійній моделі.

Тісноту зв'язку між залежною змінною y та незалежною змінною x оцінюють за допомогою статистичних характеристик: **коефіцієнта детермінації та коефіцієнта кореляції**. За допомогою цих коефіцієнтів перевіряється відповідність побудованої регресійної моделі (яка є обраною нами для перевірки на адекватність теоретичною моделлю) фактичним даним.

Під поняттям “тіснота зв'язку” розуміється оцінка впливу незалежної змінної x на залежну змінну y .

Після встановлення тісноти зв'язку між змінними моделі характеризують **значущість зв'язку**, яка в кореляційному аналізі частіше всього здійснюється за допомогою F- критерію Фішера.

1.1.1 Коефіцієнт детермінації

Коефіцієнт детермінації показує, якою мірою варіація залежної змінної (результативного показника) Y визначається варіацією незалежної змінної (вхідного показника) X . Тобто дається відповідь на запитання, чи справді зміна значення Y лінійно залежить саме від зміни значення X , а не відбувається під впливом різних випадкових факторів. Коефіцієнт детермінації використовується як при лінійному, так і при нелінійному зв'язку між змінними та розраховується за формулою:

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

де $Y_{\text{розн}}$ – теоретичні значення залежної змінної на підставі побудованої регресійної моделі; $Y_{\text{сер}}$ – загальна середня фактичних даних залежної змінної; $Y_{\text{факт}}$ – фактичні індивідуальні значення залежної змінної.

Коефіцієнт детермінації приймає значення від 0 до 1.

1.1.2 Коефіцієнт кореляції

Найпростішим критерієм, який дає **кількісну оцінку зв'язку** між двома показниками, є коефіцієнт кореляції (або індекс кореляції). Він розраховується за такою формулою:

$$R = \pm \sqrt{1 - \frac{\sum_{i=1}^n (y_{\text{факт}} - y_{\text{розрах}})^2}{\sum_{i=1}^n (y_{\text{факт}} - \bar{y})^2}} \equiv \pm \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{x_i})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Чим ближче коефіцієнт кореляції до одиниці, тим тісніший зв'язок між незалежною та залежною змінними.

Для спрощення розрахунків тісноту кореляційного зв'язку характеризують коефіцієнтом кореляції, який розраховується за формулою:

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

Значення R лежить у діапазоні від -1 до $+1$. При $R=0$ змінні не можуть мати лінійного кореляційного зв'язку. Ступінь тісноти їх лінійної залежності зростає при наближенні R до ± 1 . Кореляційний зв'язок між показниками відсутній при $R=\pm 1$. Коли $R>0$, то зв'язок між показниками прямий (додатній), якщо $R<0$ – обернений (від'ємний).

1.1.3 Перевірка адекватності моделі за критерієм Фішера

Тестування значущості змінної X за критерієм Фішера складається з наступних етапів:

- формулюється нуль-гіпотеза $H_0: a_1 = 0$;
- задається рівень значущості α (наприклад, $\alpha=0.05$ відповідає 5ти відсотковому);
- обчислюється F -відношення;
- за таблицями F -розподілу Фішера знаходимо F -критичне значення при заданому рівні значущості з ступенями вільності f_1 та f_2 ;
- цю гіпотезу відкидаємо з 5%-вим ризиком помилитися, якщо:

$$F_{\text{розрах}} > F_{0,95},$$

де $F_{0,95}$ – значення F при 5%-вому ризик помилки (знаходимо за таблицями F -критерію Фішера з відповідними ступенями вільності і заданим рівнем значущості).

Отже, достовірність моделі оцінюють порівнянням розрахункового ($F_{\text{розр}}$) та табличного значень критерію Фішера. Наприклад, нехай отримано $F_{\text{розр}} = 15,3$. Значення $F_{\text{табл}}$ визначають за спеціальними таблицями залежно від ступенів вільності f_1 та f_2 (див. дод. Б.2):

$$\begin{aligned} f_1 &= (n - m - 1), \\ f_2 &= (n - 1), \end{aligned}$$

де n – кількість спостережень; m – кількість незалежних змінних, для парної регресії $m=1$.

Задаючи рівень значущості $\alpha \in [0,1]$ ми задаємо прийнятний рівень можливої помилки. Скажімо, якщо $\alpha = 0.05$, то це означає, що в 5 відсотках випадків ми згодні примиритись з можливістю не справдження основної гіпотези, адже, якщо модель статистично значуща, то в 95% основна гіпотеза справджуватиметься.

За означенням, надійність (або рівень довіри) γ взаємозв'язаний, з рівнем значущості α простим співвідношенням $\gamma = 1 - \alpha$. Так рівень значущості $\alpha = 0.05$ відповідає надійності $\gamma = 1 - 0.05 = 0.95$.

Впевніться, що при надійності $\gamma = 1 - \alpha = 1 - 0.05 = 0.95$ та ступенях вільності $f_1 = (8 - 1 - 1) = 6$; $f_2 = (8 - 1) = 7$ табличне значення критерію Фішера буде дорівнювати $F_{\text{табл}} = 3.87$.

Оскільки значення $F_{\text{розр}} > F_{\text{табл}}$ ($15.3 > 3.87$) то маємо право зробити висновок про адекватність побудованої моделі, при цьому ми припускаємо виконання передумов для побудови регресійної моделі. Наприклад, присутність лінійного зв'язку між показниками. Перевірка неявних гіпотез проводиться додатково, так наявність лінійної залежності можемо перевірити за допомогою тесту Чоу.

1.1.4 Зв'язок між коефіцієнтом детермінації (R^2) та F-відношенням Фішера

Між коефіцієнтом детермінації R^2 та оцінкою фактичного значення F-відношення Фішера є тісний зв'язок:

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

Отже, тестування адекватності отриманої моделі (тобто перевірку загальної статистичної значущості моделі) можливе із використанням коефіцієнта детермінації. Для парної регресії $m=1$.

1.2 Покрокова побудова лінійної моделі з інтервальним прогнозом

Розглянемо виміри вхідних і вихідних даних деякої системи:

Таблиця 1.

n	$x(i)$	$y(i)$
1	0.07	5.58
2	0.26	6.57
3	2.13	9.89
4	3.34	11.36
5	5.67	13.49
6	5.83	14.1
7	6.88	15.28
8	8.67	17.14
9	10.79	23.41
10	12.07	24.07
11	13.21	26.32
12	14.16	25.63
13	14.77	29.53
14	14.79	28.13
15	14.9	29.04
...16	15.76	29.23
17	16.51	29.74

На основі спостережених показників x_i та y_i , $n = 17$ з таблиці :

- 1) - побудувати лінійну регресію y на x ;
- 2) - оцінити середню відносну похибку моделі, оцінити середню еластичність $\bar{K}_{y/x} = \hat{b}_1 \frac{\bar{x}}{\bar{y}}$;
- 3) - визначити зону надійності на рівні значущості $\alpha = 0.1$;
- 4) - перевірити наявність автокореляції залишків за критерієм Дарбіна-Уотсона;
- 5) - зробити трендовий та інтервальний прогноз $удля$ таких значень x_{pk} , де

$$x_{pk} = \max_i(x_i) + k \cdot (\max_i(x_i) - \min_i(x_i)) / n, \text{ де } k = \overline{1, \dots, 3};$$

б) – зробити висновки.

1.3 Покрокове виконання комп'ютерного практикуму до розділу 1

1-й крок:

Будуємо графік спостережених пар взаємозв'язаних економічних показників

$(x_i; y_i), i=1, \dots, n;$

Впевнюємось в доцільності використання саме моделі звичайної регресії

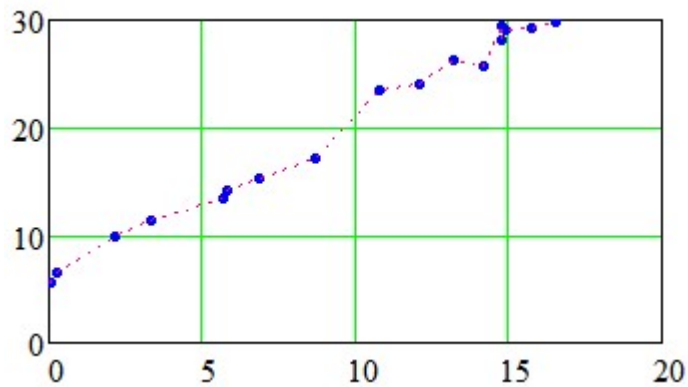


Рис.1.1. Спостережені дані

На графіку видно доцільність специфікації залежності:

$$y = b_0 + b_1x + u;$$

Для виконання наступних кроків введемо наступні позначення:

$$\bar{\delta} = \frac{\sum_{i=1}^n \delta_i}{n}; S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1}}; \quad \delta_i = \frac{\hat{u}_i}{y_i} \cdot 100\%; \quad ;$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n};$$

$$DW = \frac{\sum_{i=1}^n (u_i - u_{i-1})^2}{\sum_{i=1}^n u_i^2}; \quad \Delta \hat{y}_i = t_{\alpha,k} \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{1 + \frac{\tilde{x}_i^2}{D(x)}}; \quad D(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n};$$

$$\Delta y_{pi} = t_{\alpha,k} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{pi} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}};$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}; \quad \tilde{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-3}.$$

2-й крок:

2.1) визначити параметри моделі методом найменших квадратів (МНК) за формулами:

$$\hat{b}_0 = \frac{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$\hat{b}_1 = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$\hat{b}_1 = 1.505796 \quad \hat{b}_0 = 5.75698$$

2.2) побудувати лінійну регресію для чого:

обчислити значення $\hat{y}_i$ для кожного значення x_i і занести в таблицю у якості додаткового стовпчика, побудувати графік регресійної функції (рис.1.2)

$$\hat{y}(x) = \hat{b}_0 + \hat{b}_1 x$$

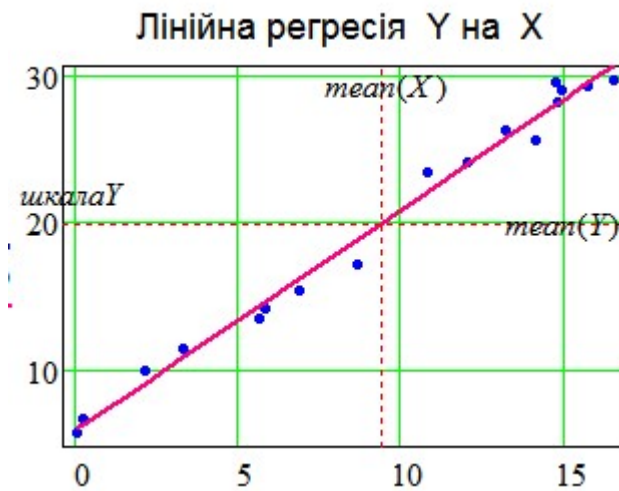


Рис.1.2. Регресія у на х .

(коефіцієнт $\hat{b}_1$ називається *коефіцієнтом регресії*, напрямом зв'язку між y і x визначає знак коефіцієнта регресії $\hat{b}_1$, у нашому випадку даний зв'язок є прямим (позитивним);

3-й крок:

3.1) обчислити залишкову дисперсію за формулою:

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1}}$$
 (в блокові аналізу S це стандартна похибка), де n – довжина вибірки, m – число факторів ($m=1$ в звичайній лінійній регресії, бо для нашого випадку y залежить від одного фактора x)

$$S = 0.9742;$$

3.2) обчислити відносну похибку розрахункових значень регресії за формулою:

$$\delta_i = \frac{\hat{u}_i}{y_i} \cdot 100\%,$$

а середнє значення відносної похибки, як

$$\bar{\delta} = \frac{\sum_{i=1}^n \delta_i}{n}, \quad \bar{\delta} = -0.02455;$$

4-й крок:

4.1) обчислити коефіцієнти еластичності за формулою

$$\overline{K}_{y/x} = \hat{b}_1 \frac{\overline{x}}{\overline{y}}, \quad \text{де}$$

$$\overline{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \overline{y} = \frac{\sum_{i=1}^n y_i}{n},$$

$$\overline{x} = 9.401, \overline{y} = 19.912,$$

$$\overline{K}_{y/x} = 0.71088;$$

5-й крок:

5.1) обчислити центровані значення $\tilde{x}_i$ за формулою

$$\tilde{x}_i = x_i - \overline{x};$$

5.2) знайти коефіцієнт Стьюдента $t_{\alpha, \kappa}$, де $\alpha=1-p=0.1$, $\kappa=n-2$

(у нашому випадку $t_{\alpha, \kappa}=1.75$);

5.3) обчислити дисперсію:

$$D(x) = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n},$$

$$D(x) = 30.534 ;$$

5.4) для побудови довірчого коридору (надійної зони) за t -розподілом в області спостережень обчислити $\Delta \hat{y}_i$ за формулою:

$$\Delta \hat{y}_i = t_{\alpha, \kappa} \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{1 + \frac{\tilde{x}_i^2}{D(x)}};$$

5.5) з'єднати неперервною лінією на графіку всі значення $(\hat{y}_i - \Delta \hat{y}_i)$ і

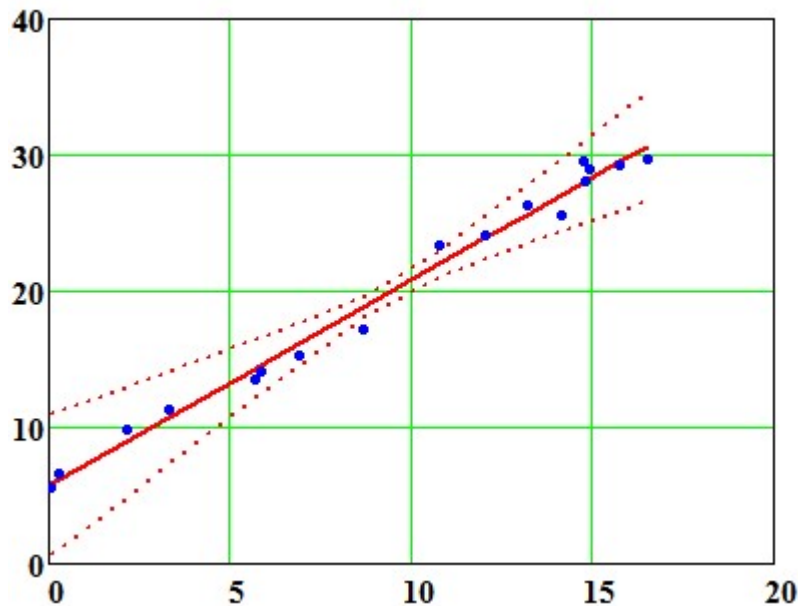


Рис.1.3.Довірчий коридор.

$(\hat{y}_i + \Delta \hat{y}_i)$ та отримані дані занести у таблицю (отримуємо надійну зону),
будуємо графік рис.1.3;

6-й крок:

6.1) обчислити оцінки збурень за формулою

$$u_i = y_i - \hat{y}_i, \text{ де } i=1, 2, \dots, n;$$

6.2) визначити DW- статистику за формулою:

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

6.3) із додаткуБ.5 в кінці цієї книги знайти нижню d_L і верхню d_U межі статистики DW - Дарбіна-Уотсона:

$$d_L=1.13, d_U=1,38;$$

6.4) зробити висновок про автокореляцію:

в даному випадку, оскільки $d_U = 1.38 < 2.229 < 4 - d_U = 2.62$, то ряд не містить автокореляції.

7-й крок:

7.1) для трендових прогнозних значень в отриману формулу $\hat{y}(x)$ підставити значення x_{p_i} :

$x_{p1} =$	17.477
$x_{p2} =$	18.444
$x_{p3} =$	19.411

та отримати

$\hat{y}(x_{p1}) =$	32.074
$\hat{y}(x_{p2}) =$	33.53
$\hat{y}(x_{p3}) =$	34.986

7.2) знайти межі надійних інтервалів прогнозованих значень за формулою

$$\Delta y_{p_i} = t_{\alpha, \kappa} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{p_i} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} =$$

=	1.859	
	1.884	
	1.911	

7.3) записати межі надійних інтервалів індивідуальних прогнозних значень

$$(y_{p_i} - \Delta y_{p_i}; y_{p_i} + \Delta y_{p_i}):$$

30.215	< 32.074 <	33.933
31.647	< 33.53 <	35.414
33.075	< 34.986 <	36.897

8) Побудувати результуючий графік з усіма отриманими результатами. Зробити остаточні висновки.

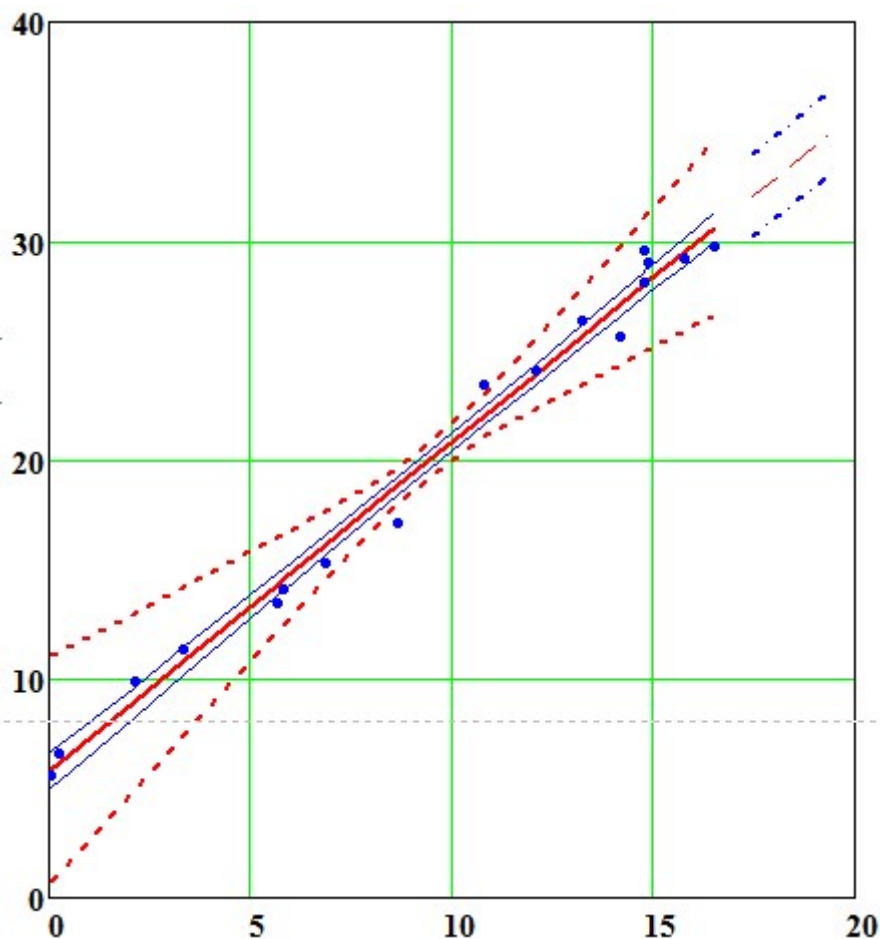


Рис.1.4. Результуючий графік, що містить точки вимірів, лінію регресії, довірчий коридор, довірчу область, прогноз за трендом та інтервальний прогноз.

9) зробити короткі економетричні та економічні висновки отриманих результатів.

Вже зараз, на отриманому графікові (рис. 1.4), видно, що існує область в яку виміри попадають з імовірністю більшою за 0.9. Розгляд обґрунтування статистичної значущості всієї моделі та кожного з її параметрів продовжимо в двох наступних прикладах виконання практикуму №1 побудови регресійної моделі.

Вперше виконувати побудову лінійної регресії рекомендується без використання спеціалізованого програмного забезпечення..

1.3.1 Виконання практикуму до розділу 1 на EXCEL

Таблиця 1.3

Обчислення завдання 1 на EXCEL без використання блоку аналізу

	A	B	C	D	E	F	G	H	I	J	K
1	n	X(i)	Y(i)	(Xi)^2	X(i)*Y(i)	Y^(i)	U(i)	U(i)^2	$\delta(i)$	$X_i = x_i - x_c$	X_i^2
2	1	0.07	5.580	0.00	0.3906	5.862386	-0.2824	0.079742	-5.06068	-	87.0598768
3	2	0.26	6.570	0.07	1.7082	6.148488	0.4215	0.177673	6.415715	-	83.5503533
4	3	2.13	9.890	4.54	21.0657	8.964327	0.9257	0.856871	9.359687	-	52.8614533
5	4	3.34	11.360	11.16	37.9424	10.78634	0.5737	0.329085	5.049818	-	36.7307298
6	5	5.67	13.490	32.15	76.4883	14.29485	-0.8048	0.647778	-5.96624	-	13.9172886
7	6	5.83	14.100	33.99	82.203	14.53577	-0.4358	0.189899	-3.09059	-	12.7491003
8	7	6.88	15.280	47.33	105.1264	16.11686	-0.8369	0.700335	-5.47683	-	6.3533651
9	8	8.67	17.140	75.17	148.6038	18.81224	-1.6722	2.796372	-9.75633	-	0.5337592
10	9	10.79	23.410	116.42	252.5939	22.00452	1.4055	1.975362	6.003741	1.3894118	1.9304651
11	10	12.07	24.070	145.68	290.5249	23.93194	0.1381	0.01906	0.573562	2.6694118	7.1257592
12	11	13.21	26.320	174.50	347.6872	25.64855	0.6714	0.450843	2.551095	3.8094118	14.5116180
13	12	14.16	25.630	200.51	362.9208	27.07906	-1.4491	2.09977	-5.65376	4.7594118	22.6520003
14	13	14.77	29.530	218.15	436.1581	27.99759	1.5324	2.348268	5.189319	5.3694118	28.8305827
15	14	14.79	28.130	218.74	416.0427	28.02771	0.1023	0.010463	0.363633	5.3894118	29.0457592
16	15	14.90	29.040	222.01	432.696	28.19335	0.8467	0.71682	2.915469	5.4994118	30.2435298
17	16	15.76	29.230	248.38	460.6648	29.48833	-0.2583	0.066736	-0.88379	6.3594118	40.4421180
18	17	16.51	29.740	272.58	491.0074	30.61768	-0.8777	0.770322	-2.95118	7.1094118	50.5437356
19	sum	159.81	338.510	2021.39	3963.824	338.51	0.0000	14.2354	-0.41738	0.000000	519.0814941
20		x _c =	9.401	b ₀ =	5.75698	S=	0.9742	δ^* =	-0.02455	D(x)=	30.5342055
21		y _c =	19.912	b ₁ =	1.505796	m=	1.0000	a ₁ =	0.710884	t(α,k)=	1.75

Таблиця 1.3 (продовження)

	L	M	N	O	P	Q	R	S	T
1	ΔY	$Y^{(i)} - \Delta Y$	$Y^{(i)} + \Delta Y$	$U_i - U_{i-1}$	$(U_i - U_{i-1})^2$	σ^2	$Y_i - Y_c$	$(Y_i - Y_c)^2$	$Y_i^{(i)} - Y_c$
2	0.811432094	5.0509541	6.6738183			-0.2823862	-14.332353	205.416341	-14.05
3	0.799232032	5.3492555	6.9477196	-0.70	0.49547334	0.421512469	-13.342353	178.018382	-13.7639
4	0.683331073	8.2809958	9.647658	0.5041606	0.25417793	0.925673088	-10.022353	100.447558	-10.948
5	0.613696734	10.172644	11.400037	-0.352014	0.12391366	0.573659371	-8.552353	73.142741	-9.12601
6	0.498887508	13.795959	14.793734	-1.378506	1.90027811	-0.80484638	-6.422353	41.246617	-5.61751
7	0.492288465	14.043485	15.028062	0.3690726	0.13621456	-0.43577382	-5.812353	33.783447	-5.37658
8	0.454463987	15.662396	16.571324	-0.401086	0.16087021	-0.8368601	-4.632353	21.458694	-3.79549
9	0.417076624	18.395159	19.229312	-0.835376	0.6978525	-1.67223576	-2.772353	7.685941	-1.10012
10	0.426348705	21.578176	22.430873	3.0777115	9.4723081	1.405475742	3.497647	2.09217	4.377181
11	0.45919738	23.472746	24.391141	-1.267419	1.60635211	0.138056272	4.157647	4.01959	16.15711
12	0.502211569	25.14634	26.150763	0.533392	0.28450706	0.671448308	6.407647	5.7362	32.90398
13	0.545706281	26.533352	27.624765	-2.120507	4.4965484	-1.44905833	5.717647	7.16671	51.36167
14	0.576532653	27.421062	28.574127	2.9814642	8.88912853	1.532405829	9.617647	8.08524	65.37113
15	0.577576571	27.450134	28.605287	-1.430116	2.04523157	0.1022899	8.217647	8.11536	65.85902
16	0.583353361	27.609994	28.776701	0.7443624	0.55407537	0.84665229	9.127647	8.28099	68.57487
17	0.63039999	28.857933	30.118733	-1.104985	1.22099175	-0.25833267	9.317647	9.57598	91.69939
18	0.673768656	29.943911	31.291449	-0.619347	0.38359113	-0.87768001	9.827647	10.7053	114.604
19	9.745503682	328.7645	348.2555	-2.003091	32.7215143	-2.6645E-15	0.000000	3.6E-14	1176.977
20	DW=	2.2986021	R^2=	0.9880497	f=	2.695			
21			R^2^=	0.9863425	s2=	0.9490			

..

Таблиця 1.3 (продовження)

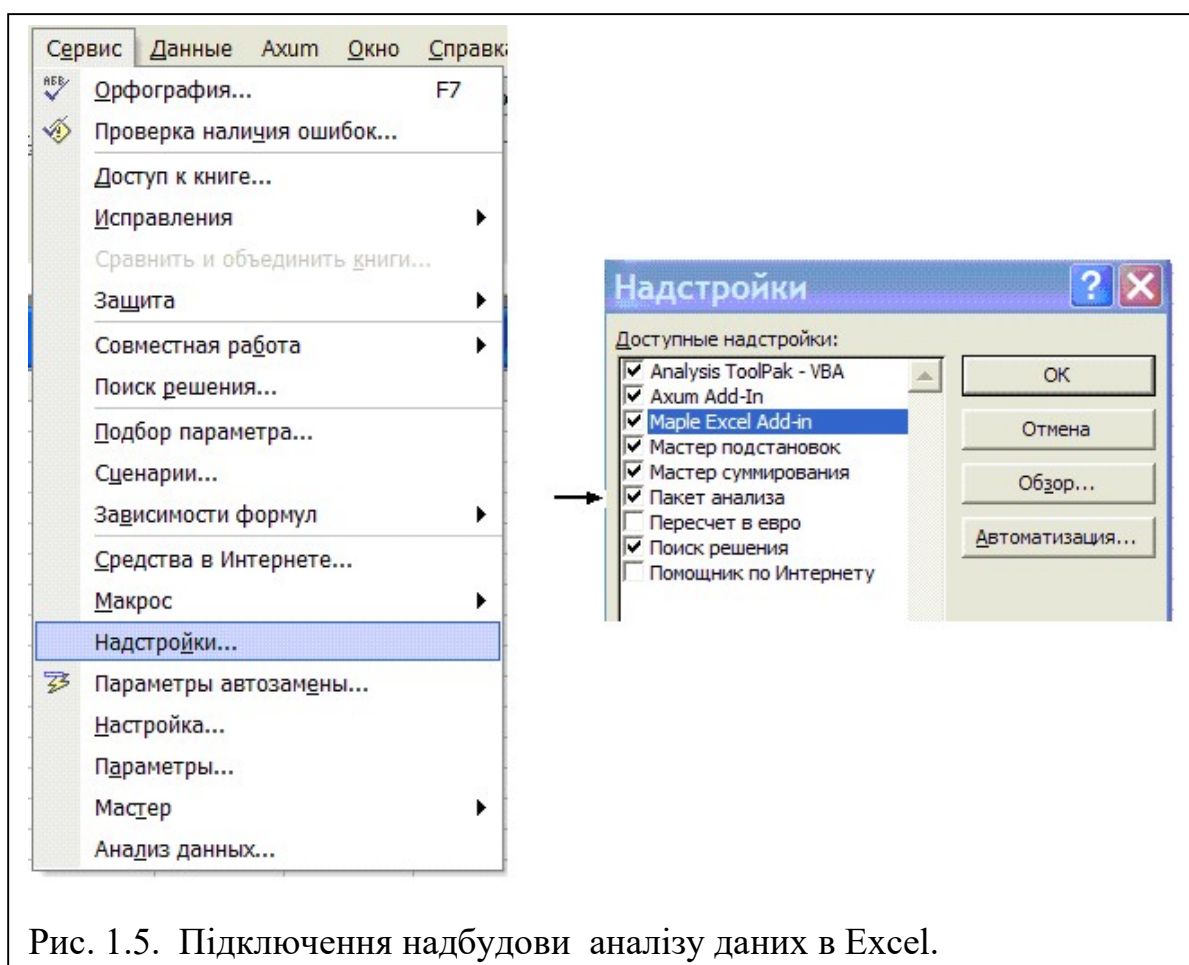
	V	W	X	Y	Z	AA
1	Y_i^2	$X_i Y_i$	Y_{left_tk}	Y_{right_tk}	Y_{left_fk}	Y_{right_fk}
2	31.136400	0.3906	2.50427828	9.220494	-4.480586	16.2053586
3	43.164900	1.7082	2.857713888	9.439261	-3.987095	16.2840704
4	97.812100	21.0657	6.334818336	11.59384	0.8654405	17.0632133
5	129.049600	37.9424	8.582576741	12.9901	3.9987479	17.5739334
6	181.980100	76.4883	12.89973196	15.68996	9.997894	18.5917988
7	198.810000	82.203	13.19513123	15.87642	10.406595	18.664953
8	233.478400	105.1264	15.12618309	17.10754	13.065575	19.1681453
9	293.779600	148.6038	18.32330414	19.30117	17.306326	20.3181451
10	548.028100	252.5939	21.35859313	22.65046	20.015056	23.9939921
11	579.364900	290.5249	22.89272645	24.97116	20.731155	27.1327329
12	692.742400	347.6872	24.2265247	27.07058	21.268709	30.0283948
13	656.896900	362.9208	25.32960101	28.82852	21.69073	32.4673869
14	872.020900	436.1581	26.03576311	29.95943	21.955155	34.0400338
15	791.296900	416.0427	26.05889563	29.99652	21.963762	34.0916587
16	843.321600	432.696	26.18610437	30.20059	22.011038	34.3756572
17	854.392900	460.6648	27.17964817	31.79702	22.377584	36.5990809
18	884.467600	491.0074	28.04500588	33.19035	22.693844	38.5415163
19	7931.743300	3963.824	307.1366001	369.8834	241.87993	435.140072
20						
21						

1.3.2 Виконання комп'ютерного практикуму до розділу 1 з використанням блоку аналізу Excel

Підготовка Excel до проведення аналізу даних

Підключення надбудов Excel

В меню Сервіс-Надбудови перевірте, чи встановлена надбудова <Пакет аналізу>. Послідовність цих дій зображена на Рис.1.5.



Щоб підключити відповідну надбудову потрібно встановити мітку, як це показано на Рис. 1.5.

1.3.3 Склад пакета аналізу Excel

Щоб продивитись вміст пакету аналізу Excel, зайдіть з головного меню <Сервіс>-<Аналіз даних> . В віконці, яке з'явиться натисніть кнопку Довідка (<Справка>)

....

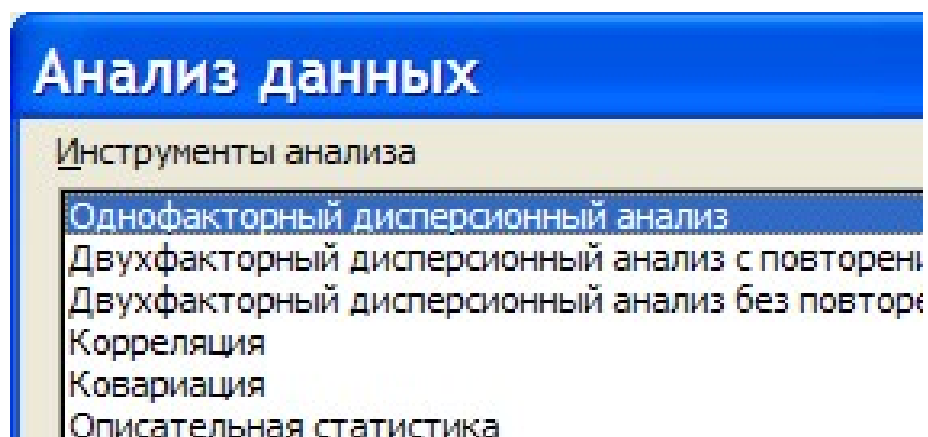


Рис.1.6. Вибір типу обробки.

Відкрийте довідку Excel та ознайомтесь з призначенням інструментів.

1.3.4 Використання таблиць Excel разом з різними програмними продуктами

Проблеми сумісності форматів даних

При передачі даних з одного програмного середовища в інше (забезпечення для обробки іншим пакетом прикладних програм) часто постає проблема сумісності. Додаткові проблеми вносяться кирилізацією програмних продуктів.

Оскільки стандарти MS Office підтримуються більшістю програмних продуктів, то будемо орієнтуватись на ці стандарти. Всі проблеми ми не вирішимо, тому будемо намагатись забезпечити сумісність числових даних.

Будемо вважати, що дані готуються для обробки в виді таблиць Excel, або в виді звичайних текстових файлів.

Встановлення міжнародних стандартів в числових форматах

В Excel використаємо можливість встановити розділові знаки в числових даних, які сумісні з міжнародним стандартом. В меню Сервіс-Параметри-Міжнародні встановимо згідно Рис. 1.7.

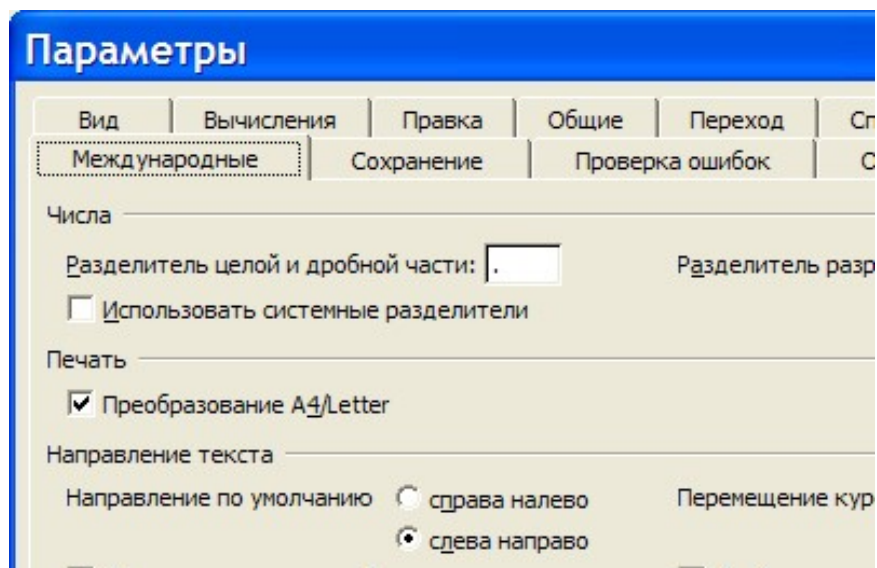


Рис.1.7. Вибір міжнародного стандарту розділових знаків.

Використання загального формату в Excel

Для сумісності форматів Excel, Maple, MathCad та інших середовищ програмування можливе використання загального формату в таблицях Excel. На Рис.1.8 зображено відкриту таблицю Test1.xls. Відкрийте за допомогою Excel файли Test1.xls, Test2.xls.

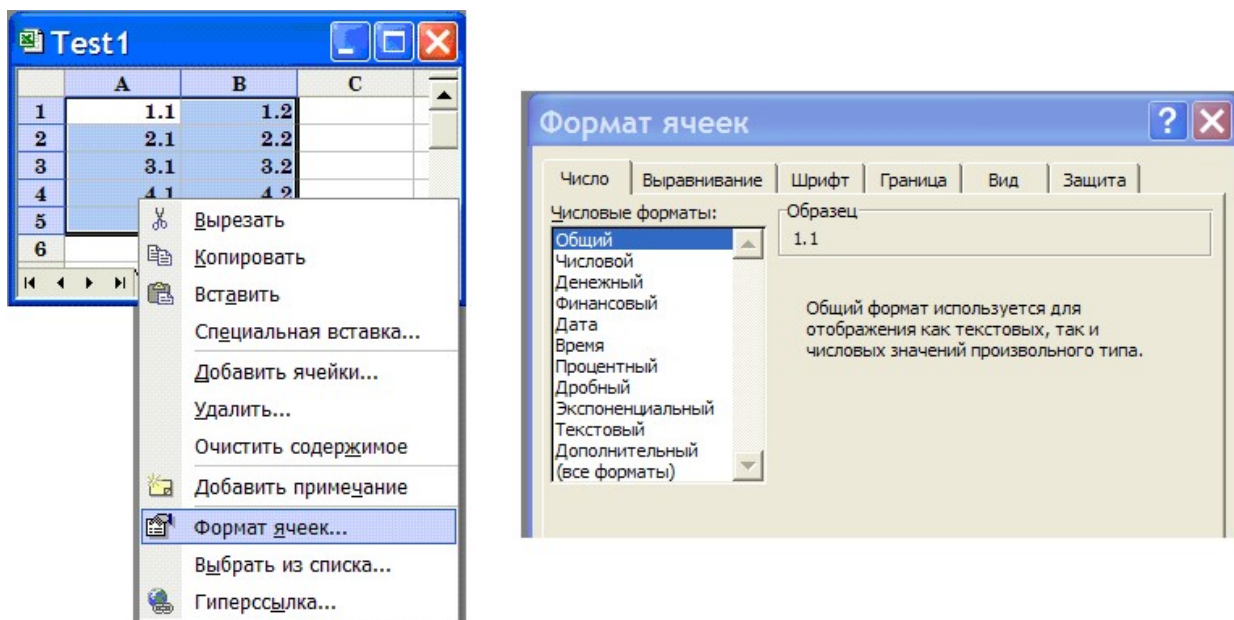


Рис. 1.8. Використання загального формату для вирішення проблеми сумісності передачі даних між різним програмним забезпеченням.

Використання такого формату - не самий вдалий метод забезпечення сумісності через те, що відбувається виключення проведення повноцінного

контролю за типами даних, які проводились би в Excel автоматично. Цей недолік компенсується простотою передачі даних між програмними продуктами (наприклад між MathCad та Excel). Спокусившись простотою передачі даних через загальний формат Excel, ми беремо на себе відповідальність за додатковий контроль правильності передачі даних.

1.3.5. Виконання завдання 1 з використанням блоку аналізу

На окремий лист Excel вставляємо вхідні дані:

Таблиця 1.4

Розрахунки завдання №1 з використанням блоку аналізу

	В	С
2	X(i)	Y(i)
3	0.07	5.580
4	0.26	6.570
5	2.13	9.890
6	3.34	11.360
7	5.67	13.490
8	5.83	14.100
9	6.88	15.280
10	8.67	17.140
11	10.79	23.410
12	12.07	24.070
13	13.21	26.320
14	14.16	25.630
15	14.77	29.530
16	14.79	28.130
17	14.90	29.040
18	15.76	29.230
19	16.51	29.740
20	17.477	
21	18.444	
22	19.411	

В клітинках B20:B22 одразу вводимо точки для прогнозу. Кількість точок прогнозу – довільне натуральне число, яке не обов'язково рівне трьом. Запускаємо блок аналізу:

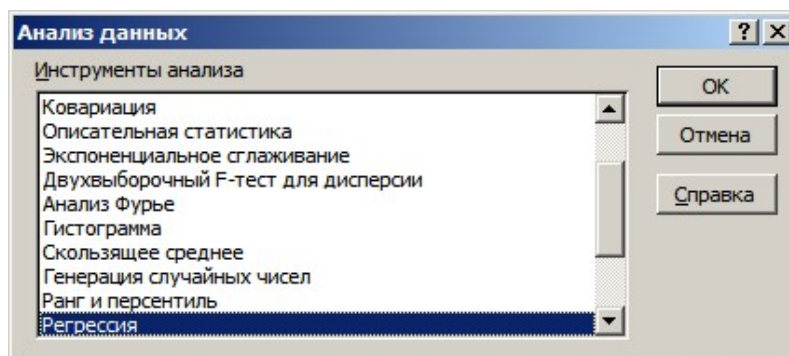


Рис.1.9.Вибір виду обробки (регресія)

При виборі <Регрессия> з'являється вікно:

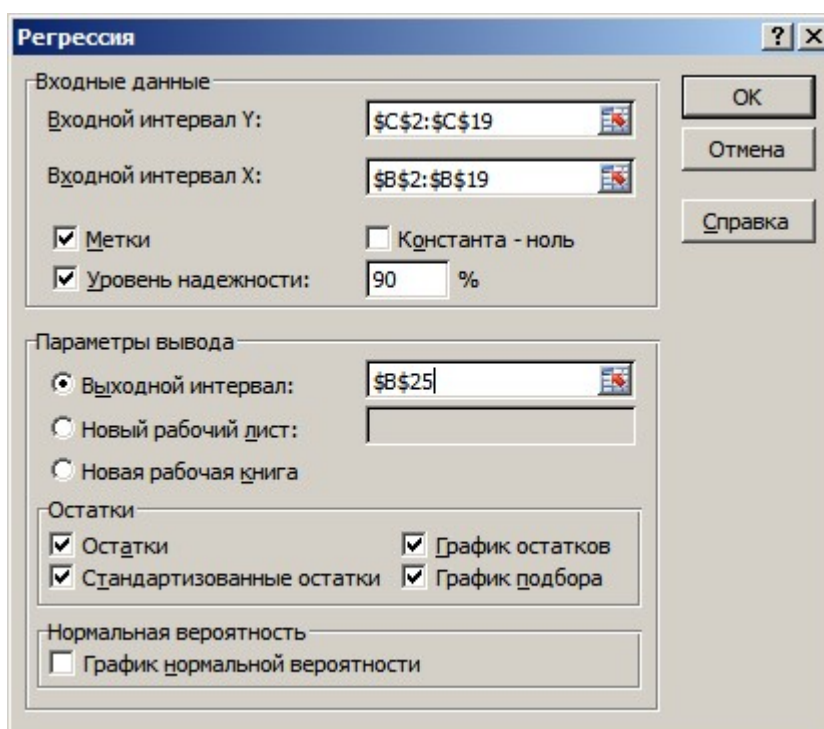


Рис.1.10. Задаємо параметри так, як показано на рисунку.

В цьому віконці все набираємо так, як на попередньому рисунку. Натискаємо <Enter>. Отримуємо набір таблиць:

Таблиця 1.5.

Регресійна статистика

	<i>B</i>	<i>C</i>	
27	<i>Регрессионная статистика</i>		
28	Множественный R	0.994006871	
29	R-квадрат	0.988049659	R^2
30	Нормированный R-квадрат	0.98725297	
31	Стандартная ошибка	0.974179916	S
32	Наблюдения	17	

Таблиця 1.6

Дисперсійний аналіз

	B	C	D	E	F	G
34						
35		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
36	Регрессия	1	1176.977	1176.977	1240.194343	7.75025E-16
37	Остаток	15	14.2354	0.949027		
38	Итого	16	1191.213			

Таблиця 1.7

Коефіцієнти регресії

	B	C	D	E	F
40		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
41	Y-пересечение	5.75698045	0.466253324	12.3473231	2.92401E-09
42	X(i)	1.50579646	0.042758395	35.2163931	7.75025E-16

Таблиця 1.7(проовження)

Коефіцієнти регресії

	G	H	I	J
40	Нижние 95%	Верхние 95%	Нижние 90.0%	Верхние 90.0%
41	4.7631850	6.7507759	4.9396149	6.5743460
42	1.4146591	1.5969339	1.43083884	1.5807541

Використання електронних таблиць EXCEL дозволяє обійтись без таблиць із критичними значеннями *t*-критерію й *F*-Критерію. У результатах розрахунків з'являються нові значення **Значущість *F* (G36)** і **Значущість *t* (F40:F42)**, яке визначає розрахунковий рівень значущості *F* і *t*-критеріїв згідно заданих вхідних даних. Якщо це значення менше заданого рівня значущості (в прикладі $\alpha = 0.1$), то модель вважається статистично значущою.

В комірці G:36 отримано $7.75025E-16 < 0.1$, тому гіпотеза про рівність нулю всіх коефіцієнтів відхиляється. Це означає, що **отримана модель статистично значуща**.

Трендові прогнознi значення будуються тепер дуже просто, для цього досить набрати $=\$C\$41+\$C\$42*B3$ в клітинці G:49 та розмножити вниз до

повного дублювання прогнозних значень C49:C65 та додатково продовжити ще на три клітинки для моментів майбутнього часу, в яких ми хочемо зробити прогноз. Для побудови інтервального прогнозу в нас уже є стандартна похибка S , дійсно вона міститься в клітинці C:31 (таблиці регресійної статистики). Обраховувати $t_{\alpha,k}$ ми вже вміємо з першого способу побудови регресії (це забирає кілька секунд часу).

Таблиця 1.8

Залишки і прогнози

	C	D	E	F
48	<i>Наблюдение</i>	<i>Предсказанное Y(i)</i>	<i>Остатки</i>	<i>Стандартные остатки</i>
49	1	5.862386204	0.282386204	-0.29937716
50	2	6.148487531	0.421512469	0.446874544
51	3	8.964326912	0.925673088	0.981370113
52	4	10.78634063	0.573659371	0.608176006
53	5	14.29484638	-0.804846381	-0.853273358
54	6	14.53577382	-0.435773815	-0.461993984
55	7	16.1168601	-0.836860098	-0.887213315
56	8	18.81223576	-1.672235763	-1.772852879
57	9	22.00452426	1.405475742	1.490042116
58	10	23.93194373	0.138056272	0.14636301
59	11	25.64855169	0.671448308	0.711848826
60	12	27.07905833	-1.44905833	-1.536247035
61	13	27.99759417	1.532405829	1.624609488
62	14	28.0277101	0.1022899	0.108444603
63	15	28.19334771	0.84665229	0.897594695
64	16	29.48833267	-0.258332666	-0.273876341
65	17	30.61768001	-0.877680012	-0.930489331

Таблиця 1.9

Інтервальний прогноз для рівня надійності 90%

Xp	Yp(i)-ΔY	Yp	Yp(i)+ΔY
17.477	30.21513224	32.07378519	33.93243814
18.444	31.74335022	33.52989037	35.31643051
19.411	33.19793544	34.98599554	36.77405564

Прогноз за трендом зображено на наступному рис.1.11.

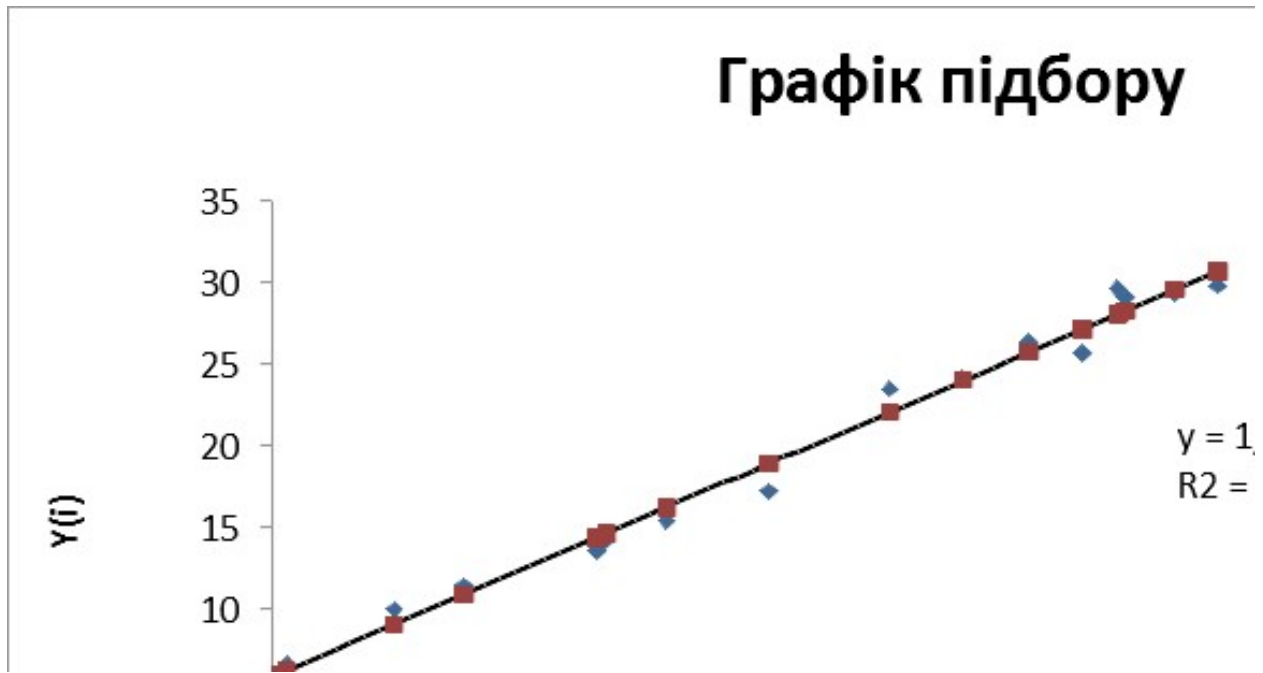


Рис.1.11. Графік побудованої регресії.

Економетрична модель $\hat{Y} = 5.76 + 1.51X\sqrt{a^2 + b^2}$ кількісно описує зв'язок факторів Y та X .

Параметр $\hat{b}_1 = 1.51$ характеризує граничну ефективність ресурсу X , показники X та Y тісно взаємозв'язані, 98 відсотків варіації Y пояснюється зміною X .

Визначимо коефіцієнт еластичності:

$$\bar{K}_{y/x} = 0.71088.$$

На основі коефіцієнта еластичності можемо стверджувати, що при збільшенні вкладень в ресурс X на один відсоток Y зросте на 0,71%.

1.3.6.Контрольні питання для підготовки до виконання практикуму

1. Екзогенні змінні. Ендогенні змінні. Проста економетрична модель. Стохастична складова. Сукупність спостережень.
2. Основні види розподілів випадкової величини. Емпірична функція розподілу. Гістограма.
3. Передумови застосування методу найменших квадратів.
4. Метод найменших квадратів.
5. Функція правдоподібності.
6. Метод максимальної правдоподібності.

7. Система нормальних рівнянь.
8. Оцінка дисперсії залишків.
9. Коефіцієнти детермінації і кореляції.
10. Перевірка значущості та інтервали довіри.

Рекомендована література: [1, 2, 3, 7,8].

Таблиця 1.10

Варіанти індивідуальних даних для практикуму №1

Варіант 1																	
X	0.07	0.26	2.13	3.34	5.67	5.83	6.88	8.67	10.8	12.07	13.21	14.16	14.77	14.79	14.9	15.8	16.51
Y	0.58	1.57	4.89	6.36	8.49	9.1	10.3	12.1	18.4	19.07	21.32	20.63	24.53	23.13	24.04	24.2	24.74
Варіант 2																	
X	1.39	2.15	2.5	2.62	2.67	5.39	5.46	5.52	5.86	6.4	9.37	10.21	12.27	13.42	13.69	13.9	15.34
Y	2.72	5.5	3.49	4.77	3.91	7.21	7.2	7.67	8.2	9.09	9.52	11.47	15.06	16.31	16.64	16.4	17.34
Варіант 3																	
X	0.29	0.3	1.62	1.85	2.31	2.41	6.02	6.11	7.08	8.35	12.02	12.78	12.97	15.42	16.28	16.6	16.71
Y	2.83	3.49	2.24	2.23	4.72	3.52	7.5	6.28	6.98	5.9	10.46	9.87	11.2	13.13	14.91	12.5	12.59
Варіант 4																	
X	0.43	1.32	2.01	2.07	2.75	2.95	3.82	6.64	6.74	7.04	7.19	7.46	10.01	12.92	12.99	15.9	16.72
Y	2.95	5.84	6.77	4.77	6.43	7.1	5.95	10.1	8.8	10.61	9.87	10.1	11.35	13.58	12.8	14	17.48
Варіант 5																	
X	0.02	0.66	0.87	2.4	3.15	3.36	5.97	8.72	9.39	12.11	13.75	14.97	15.47	15.79	16.08	16.4	16.59
Y	5.34	5.46	7.08	8.3	8.14	5.94	9.8	10.7	10.9	12.49	15.63	14.94	15.12	17.25	15.73	17.4	15.86
Варіант 6																	
X	0.15	0.22	0.65	2.42	3.99	4.37	6.7	8.02	8.56	9.73	11.32	11.36	11.94	13.28	13.93	15	15.76
Y	5.51	5.76	4.81	7.36	10.4	8.18	10.9	10.6	10.7	14.4	15.4	15.15	15.02	12.37	15.79	17.1	16.91
Варіант 7																	
X	0.59	0.97	1.57	2.76	5.01	5.52	7.98	8.5	8.67	11.8	12.61	14.08	14.5	14.57	14.64	16.3	16.89
Y	7.75	5.94	8.31	9.4	10.1	9.96	12.9	11.4	12.8	13.62	16.18	16.29	16.74	17	15.93	18.3	15.33
Варіант 8																	
X	0.59	0.97	1.57	2.76	5.01	5.52	7.98	8.5	8.67	11.8	12.61	14.08	14.5	14.57	14.64	16.3	16.89
Y	7.6	5.71	7.89	8.67	8.83	8.53	10.8	9.24	10.5	10.57	12.89	12.62	12.96	13.2	12.13	14.1	10.97
Варіант 9																	

X	1.18	4.21	4.35	4.36	6.44	9.01	9.54	10.1	10.5	11.47	14.08	14.21	14.43	15.59	15.77	16	16.02
Y	9.89	9.99	11.8	11.4	14.3	13.4	15.1	15.7	15.3	15.18	15.5	16.06	17.9	18.43	17.64	19.1	18.42
Варіант 10																	
X	1.34	2.43	2.9	3.37	3.56	4.1	5.37	6.23	7.07	7.63	9.06	9.84	9.96	12.14	12.83	13.2	13.24
Y	8.02	11.8	10	12.6	12.2	11.6	11.5	11.7	12.6	11.96	13.25	13.13	14.33	14.99	14.93	13.7	15.82
Варіант 11																	
X	0.45	2.39	5.79	5.89	5.93	8.27	8.72	11.1	11.7	12.26	13.7	14.06	15.07	15.83	15.89	16.3	16.54
Y	10.4	12.8	16.6	15.3	13.9	16.6	16.1	18.1	16.3	19.94	20.01	20.63	17.9	19.95	19.44	19.2	20.78
Варіант 12																	
X	1.25	2.3	3.45	5.22	7.07	7.25	8.27	8.87	9.99	10.12	10.37	11.81	12.36	13.23	14.13	14.4	14.99
Y	13.5	15.7	14.5	15.7	17	17.1	17.8	15.2	17.6	18.58	17.69	19.06	19.91	20.43	19.25	19.8	21.31
Варіант 13																	
X	3.56	3.66	4.3	4.53	6.44	7.26	7.67	7.85	8.28	8.31	10.45	11.09	12.12	13.83	14.66	14.8	15.82
Y	15.6	15.9	15.7	17.1	16.9	16.8	16.8	18	15.4	18.06	18.63	19.69	21.73	22.3	21.23	22.2	23.14
Варіант 14																	
X	0.77	1	3.76	3.79	3.91	4.34	6.05	6.35	6.55	7.8	10.39	10.97	11.86	12.43	13.39	14.7	15.84
Y	14.7	14.6	16.6	16.1	17.5	17.2	16.4	18.9	17.6	18.23	19.95	20.7	21.34	22.53	22.34	22.8	23.16
Варіант 15																	
X	0.44	2.17	2.29	5.05	5.82	6.18	6.34	7.71	10.8	10.91	11.46	14.21	14.44	15.39	15.54	16.3	16.68
Y	16.2	17.3	16.3	18.9	20.1	17.6	18.3	19.3	21.2	19.47	21.54	23.02	22.93	25.06	24.8	23.8	24.59
Варіант 16																	
X	1	1.29	1.72	1.76	2.09	2.16	2.87	4.24	8.07	10.94	12.63	14.1	14.21	15.43	15.89	16.8	16.81
Y	17.1	18	16.2	17.6	18.6	18.5	19.1	19.3	19.6	21.6	24.6	23.4	24.04	24.95	24.39	24.1	26.65
Варіант 17																	
X	1.92	2.45	3.77	4.33	5.91	6.83	7.27	8.29	8.31	8.73	9.14	11.64	11.69	13.6	13.93	14.5	15.59
Y	18.6	19.4	20.5	19.7	20.3	20.1	21.8	20.8	21.3	23.05	22.08	23.86	24.28	25.95	23.76	25.8	23.78
Варіант 18																	
X	0.77	1	3.76	3.79	3.91	4.34	6.05	6.35	6.55	7.8	10.39	10.97	11.86	12.43	13.39	14.7	15.84
Y	15	15.5	14.2	13.9	15.2	15.4	14.5	17	15.6	15.4	17.25	18.32	20.09	17.53	19.45	19.7	20.78

Варіант 19																	
X	1.18	4.21	4.35	4.36	6.44	9.01	9.54	10.1	10.5	11.47	14.08	14.21	14.43	15.59	15.77	16	16.02
Y	10.2	10.3	10.5	11.3	12.7	12.2	12.8	12.2	12.9	11.35	14	14.09	13.21	15.28	14.4	14.9	13.6
Варіант 20																	
X	0.78	0.87	0.89	1.36	2.43	2.72	2.84	5.15	5.43	5.95	6.31	6.34	6.44	9.7	14.19	14.4	15.44
Y	19.7	21.5	20	20.9	21.4	20.3	21.6	24.4	23	22.2	24.1	24.45	24.11	26.34	28.06	28	28.33
Варіант 21																	
X	0.13	0.65	1.07	2.2	4.14	5.16	6.55	7.01	8.35	9.02	9.08	12.71	12.78	14.62	15.18	16.7	16.72
Y	23	20.1	21.3	20	23.4	23.7	23	25.3	25	26.84	24.2	27.05	28.26	28.96	29.43	31.6	29.8
Варіант 22																	
X	1.45	2.29	2.45	3.7	4.21	4.41	4.63	9.77	11.6	12.22	13.33	14.5	14.52	15.15	15.35	15.8	16.32
Y	22.9	21.9	23.6	24.3	24.1	26.3	24.9	28.1	28.1	28.19	30.12	29.92	28.4	30.54	30.49	32.5	32.62
Варіант 23																	
X	1.14	2.29	5.32	5.34	6.25	7.72	8.11	9.41	10.5	10.95	11.05	11.19	11.46	12.73	14.12	15.4	15.62
Y	22	24.7	26.7	25	26.7	25.6	27.6	28.3	27.7	28.72	29.62	29.62	29.57	29.87	30.02	30.9	31.56
Варіант 24																	
X	0.45	2.39	5.79	5.89	5.93	8.27	8.72	11.1	11.7	12.26	13.7	14.06	15.07	15.83	15.89	16.3	16.54
Y	10.9	10.6	13.2	13.4	13.5	14.3	13.2	13.9	16	13.88	16.64	15.45	15.64	15.66	15.86	17.2	14.17
Варіант 25																	
X	0.64	2.29	2.96	3.47	4.21	4.44	6.05	8.24	9.31	9.72	10.33	11.06	13.38	13.5	13.66	13.7	14.53
Y	24.2	24.8	26.8	28.4	27.4	27.7	29.3	28.7	29.4	31.05	29.87	30.16	34.21	31.28	33.1	31.9	32.56
Варіант 26																	
X	1.29	2.08	3.02	7.04	8.76	8.9	9.73	10.1	10.5	10.62	11.2	11.36	11.75	12.13	12.58	12.8	13.98
Y	26.5	27.7	29.1	29.6	32.3	31.1	30.9	30.8	31.3	31.87	31.69	32.6	31.65	32.79	31.97	32.5	33.46
Варіант 27																	
X	0.63	1.06	1.6	2.66	2.97	3.18	7.79	8.83	8.95	9.17	9.63	12.21	12.48	12.82	12.92	16.4	16.44
Y	28.2	26.7	25.6	27.4	28.2	28.3	31.4	31	31.2	32.36	31.97	34.24	32.82	34.67	32.45	36.2	35.24

Варіант 28																	
X	0.77	1	3.76	3.79	3.91	4.34	6.05	6.35	6.55	7.8	10.39	10.97	11.86	12.43	13.39	14.7	15.84
Y	15	15.5	14.2	13.9	15.2	15.4	14.5	17	15.6	15.4	17.25	18.31	20.08	17.53	19.45	19.7	20.78
Варіант 29																	
X	0.08	0.53	0.57	0.84	1.31	1.89	2.78	3.57	3.89	4.41	4.95	6.22	8.65	10.1	11.09	12.3	14.94
Y	29.6	29	30.2	29.6	29	32	29.2	31.4	30.3	32.01	29.5	32.25	34.45	33.37	34.13	35.8	38.54
Варіант 30																	
X	0.02	0.33	1.23	2.23	3.16	8.85	8.92	9.49	10.1	10.4	10.56	10.66	12.09	12.83	13.25	16.2	16.54
Y	31.3	30.7	32.1	31	30.5	35.6	34.9	33.5	34.2	36.32	33.87	37.28	35.77	39.06	38.3	41.3	37.07

Розділ 2. НЕЛІНІЙНА ПАРНА РЕГРЕСІЯ

2.1.Короткі теоретичні відомості

Нелінійні моделі і методи їх аналізу значно складніші, ніж лінійні. В економічній теорії проблема ускладнюється ще і складністю інтерпретації результатів. Вивчення нелінійної парної регресії почнемо з моделей, які допускають використання методів лінеаризації.

В методах оцінювання параметрів розглядають парні нелінійні регресії двох видів:

- 1) нелінійні за факторами, але лінійні за невідомими параметрами, які підлягають оцінці;
- 2) нелінійні за факторами і параметрами.

Регресії, нелінійні за факторами, але лінійні за оцінюваними параметрами, називаються **квазілінійними**.

Розглянемо загальний випадок квазілінійних парних регресій. Нехай задано пари вимірів показників Y, X , які заносимо в таблицю 2.1 :

Таблиця 2.1

Наявні пари вимірів показників

X	x_1	x_2	...	x_n
Y	y_1	y_2	...	y_n

Парну квазілінійну регресію можна записати в загальному вигляді:

$$\hat{Y} = a + b \cdot \varphi(X).$$

Заміною величин

$$z_i = \varphi(x_i),$$

$i = \overline{1, n}$ нелінійна парна регресія зводиться до лінійної парної регресії:

$$\hat{Y} = a + b \cdot Z.$$

Формули для оцінок параметрів (згідно до нормальної системи рівнянь для парної лінійної регресії) набувають вигляду:

$$b = \frac{n \sum_{i=1}^n z_i y_i - \sum_{i=1}^n z_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n z_i^2 - \left(\sum_{i=1}^n z_i \right)^2}, \quad a = \frac{\sum_{i=1}^n z_i^2 \cdot \sum_{i=1}^n y_i - \sum_{i=1}^n z_i \cdot \sum_{i=1}^n z_i y_i}{n \cdot \sum_{i=1}^n z_i^2 - \left(\sum_{i=1}^n z_i \right)^2}.$$

Коефіцієнт еластичності в нелінійній регресії

В економічних задачах для оцінки впливу на показник будь-якого фактора часто використовують коефіцієнт еластичності. Якщо відомо статистичний ряд з базисними даними показника і фактора, то коефіцієнт еластичності для значення фактора знаходять за формулою

$$K_{x_i} = \frac{\Delta y_i}{y_i} : \frac{\Delta x_i}{x_i},$$

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

Якщо між фактором і показником знайдена стохастична залежність, то коефіцієнт еластичності для значення фактора аналогічно можна знайти за

$$\text{формулою } K_{x_i} = \frac{\Delta \hat{y}(x_i)}{\hat{y}(x_i)} : \frac{\Delta x_i}{x_i}.$$

Якщо зробити граничний перехід при $\Delta x_i \rightarrow 0$, то одержимо формулу для точкової оцінки коефіцієнта еластичності:

$$K_{x_i} = \frac{x_i}{\hat{y}(x_i)} \lim_{\Delta x_i \rightarrow 0} \frac{\Delta \hat{y}(x_i)}{\Delta x_i} = \frac{x_i \hat{y}'(x_i)}{\hat{y}(x_i)}.$$

Коефіцієнт еластичності показує, на скільки відсотків зміниться показник, якщо фактор зміниться на один відсоток.

Для парної квазілінійної регресії $\hat{Y} = a_0 + a_1\varphi(X)$ коефіцієнт еластичності знаходиться за формулою $K = \frac{a_1\varphi'(X)X}{\hat{Y}}$.

Для звичайної лінійної регресії $\hat{Y} = a_0 + a_1X$ коефіцієнт еластичності знаходиться за формулою $K_X = \frac{(a_0 + a_1X)'X}{\hat{Y}} = \frac{a_1X}{\hat{Y}}$.

Таблиця 2.2.

Різновиди парних рівнянь регресії

Функції		Аналітичний вираз
1.	Лінійна	$Y = a_0 + a_1X$
2.	Параболічна	$Y = a_0 + a_1X^2$
3.	Гіперболічна	$Y = a_0 + a_1 \frac{1}{X}$
4.	Степенева	$Y = a_0X^{a_1}$
5.	Показникова	$Y = a_0a_1^X$
6.	Модифікована експоненціальна	$Y = a_0e^{a_1X}$
7.	Показниково-степенева	$Y = a_0X^{a_1}a_2^X$
8.	Екологічна	$Y = a_0e^{-a_1^2(X-a_2)^2}$
9.	Логістична	$Y = \frac{a_0}{1 + a_1e^{a_2X}}$
10.	Гомперця	$Y = e^{a_0a_1^X + a_2}$
11.	Іраціональна	$Y = \sqrt{a_0 + a_1X + a_2X^2}$
12.	Обернена до квадратичної	$Y = \frac{1}{a_0 + a_1X + a_2X^2}$

13.	Дробово-раціональна	$Y = \frac{X}{a_0 + a_1 X + a_2 X^2}$
14.	Функція Джонсона	$\log Y = -\frac{a_0}{a_1 + X} + a_2$
15.	Функції Торнквіста	$Y = \frac{a_0 X}{a_1 + X}$
		$Y = \frac{a_0 (X - a_1)}{X + a_2}$
		$Y = \frac{a_0 X (X - a_1)}{X + a_2}$

2.1.1. Мета та основні завдання комп'ютерного практикуму до розділу 2

На базі статистичних даних показників змінних $Y(t)$
за $n=17$ місяців спостереження показника

Таблиця 2.3

t	Y
2	9.51
3	11.62
4	11.22
5	15.22
6	13.99
7	15.18
8	14.98
9	17.88
10	16.78
11	18.94
12	20.98
13	15.71
14	20.74
15	24.7
16	20.78
17	20.74
18	19.75

побудувати графік тренду зміни $Y(t)$, обрати найефективнішу форму однофакторної моделі серед 8-ми нелінійних і однієї лінійної залежностей з таблиці 2.4.

Вказані в таблиці 2.4 залежності зберігають властивості середніх величин.

Якщо під середніми величинами розуміти одну з середніх: арифметичну, геометричну, гармонічну, то при виборі специфікації моделей в виді однієї з вказаних в таб. 2.4 функцій матиме місце деяке додаткове рівняння між узагальненими середніми. Наприклад, для звичайної лінійної регресії, що відповідає специфікації моделі через функцію $f1(a_0, a_1, t)$, матиме місце рівняння для звичайних середніх

$$\bar{y} = \hat{a}_0 + \hat{a}_1 \bar{x}.$$

Таблиця 2.4

Моделі, що зберігають властивості середніх

1	$f1(a_0, a_1, t) := a_0 + a_1 \cdot t$
2	$f2(a_0, a_1, t) := a_0 + a_1 \cdot \ln(t)$
3	$f3(a_0, a_1, t) := a_0 + \frac{a_1}{t}$
4	$f4(a_0, a_1, t) := a_0 \cdot a_1^t$
5	$f5(a_0, a_1, t) := (a_0 \cdot t^{a_1})$
6	$f6(a_0, a_1, t) := e^{a_0 + \frac{a_1}{t}}$
7	$f7(a_0, a_1, t) := \frac{1}{a_0 + a_1 \cdot t}$
8	$f8(a_0, a_1, t) := \frac{1}{a_0 + a_1 \cdot \ln(t)}$
9	$f9(a_0, a_1, t) := \frac{t}{a_0 + a_1 \cdot t}$

2.1.2. Порядок і рекомендації щодо виконання комп'ютерного практикуму №2

Позначимо перший стовпчик вихідних даних Таб. 2.3 через X , другий Y . Незавжди бачити що побудова всіх дев'яти моделей Таб. 2.4 зводиться до побудови звичайних лінійних регресій на парах векторів вибраних серед

$$X_i, i := \overline{1, n}, \quad Y_i, i := \overline{1, n}, \quad X1_i = (1 / X_i), i := \overline{1, n}, \quad Y1_i = (1 / Y_i), i := \overline{1, n}$$

$$\ln X_i = \ln(X_i), i := \overline{1, n}, \quad \ln Y_i = \ln(Y_i), i := \overline{1, n}.$$

Обчислимо ці стовпчики і будемо їх використовувати для побудови відповідних регресій.

	1		1		1		1
1	2	1	9.51	1	0.5	1	0.105
2	3	2	11.62	2	0.333	2	0.086
3	4	3	11.22	3	0.25	3	0.089
4	5	4	15.22	4	0.2	4	0.066
5	6	5	13.99	5	0.167	5	0.071
6	7	6	15.18	6	0.143	6	0.066
7	8	7	14.98	7	0.125	7	0.067
8	9	8	17.88	8	0.111	8	0.056
9	10	9	16.78	9	0.1	9	0.06
10	11	10	18.94	10	0.091	10	0.053
11	12	11	20.98	11	0.083	11	0.048
12	13	12	15.71	12	0.077	12	0.064
13	14	13	20.74	13	0.071	13	0.048
14	15	14	24.7	14	0.067	14	0.04
15	16	15	20.78	15	0.063	15	0.048
16	...	16	...	16	...	16	...

	1		1
1	0.693	1	2.252
2	1.099	2	2.453
3	1.386	3	2.418
4	1.609	4	2.723
5	1.792	5	2.638
6	1.946	6	2.72
7	2.079	7	2.707
8	2.197	8	2.884
9	2.303	9	2.82
10	2.398	10	2.941
11	2.485	11	3.044
12	2.565	12	2.754
13	2.639	13	3.032
14	2.708	14	3.207
15	2.773	15	3.034
16	...	16	...

Рис. 2.1. Дані побудови регресій із Табл. 2.4.

Номер моделі позначатимемо k . Модель $k=1$

Найшвидше оцінювати коефіцієнти a_{0_k}, a_{1_k} через оператори intercept та slope.

Щоб визначити найкращу з моделей будемо для кожної з 9-ти моделей обраховувати суми $\sum_{i=1}^n (Y_i - \hat{Y}_{i,k})^2$. Мінімальна з таких сум серед моделей буде вказувати на найбільш ефективну модель (тобто таку, яка має найменшу дисперсію ухилів).

Порядок обчислень викладено в Таблиці 2.5.

У таблиці 2.6. вказані пари коефіцієнтів для моделей 1-9.

Таблиця 2.5

Вибір найбільш ефективної моделі

$k = 1$	$a_{0_k} = \text{intercept}(X, Y) = 9.823$	$a_{1_k} = \text{slope}(X, Y) = 0.716$	$\sum_i (Y_i - Y_{m_i, k})^2 =$ 60.053
$k = 2$	$a_{0_k} = \text{intercept}(\ln X, Y) = 4.695$	$a_{1_k} = \text{slope}(\ln X, Y) = 5.74$	-----“----- 51.777
$k = 3$	$a_{0_k} = \text{intercept}(X1, Y) = 21.189$	$a_{1_k} = \text{slope}(X1, Y) = -28.655$	-----“----- 84.786
$k = 4$	$a_{0_k} = \exp(\text{intercept}(X, \ln Y)) = 10.493$	$a_{1_k} = \exp(\text{slope}(X, \ln Y)) = 1.046$	-----“----- 73.75
$k = 5$	$a_{0_k} = \exp(\text{intercept}(\ln X, \ln Y)) = 7.416$	$a_{1_k} = \exp(\text{slope}(\ln X, \ln Y)) = 0.373$	-----“----- 50.479
$k = 6$	$a_{0_k} = \text{intercept}(X1, \ln Y) = 3.086$	$a_{1_k} = \text{slope}(X1, \ln Y) = -1.93$	-----“----- 70.604
$k = 7$	$a_{0_k} = \text{intercept}(X, Y1) = 9.822$	$a_{1_k} = \text{slope}(X, Y1) = 0.716$	-----“----- 108.688
$k = 8$	$a_{0_k} = \text{intercept}(X, Y) = 5.42$	$a_{1_k} = \text{slope}(X, Y) = 5.401$	-----“----- 57.629
$k = 9$	$a_{0_k} = \text{intercept}(X1, Y1) = -25.341$	$a_{1_k} = \text{slope}(X1, Y1) = 20.703$	-----“----- 62.031

Таблиця 2.6

Пари коефіцієнтів для моделей $k=1-k=9$

Модель k		
1	9.823	0.716
2	4.695	5.74
3	21.189	-28.655
4	10.493	1.046
5	7.416	0.373
6	3.086	-1.93
7	9.822	0.716
8	5.42	5.401
9	-25.341	20.703

2.2. Візуальний аналіз вибору моделі

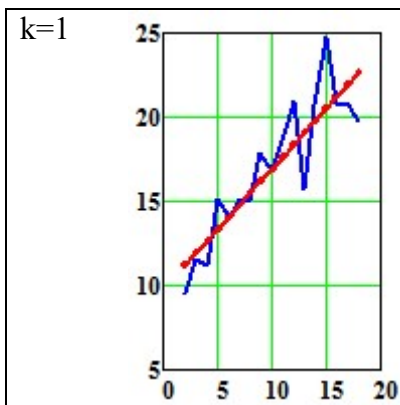


Рис.2.2 а). Модель №1

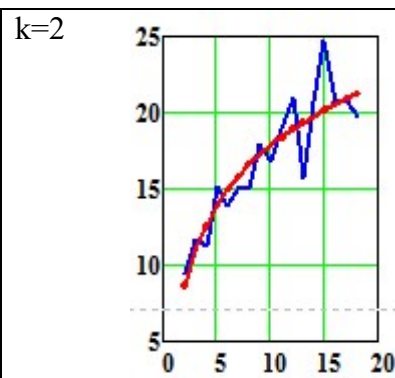


Рис.2.2 б). Модель №2

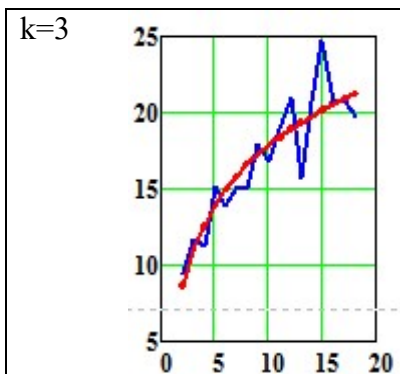


Рис. 2.2 в). Модель №3

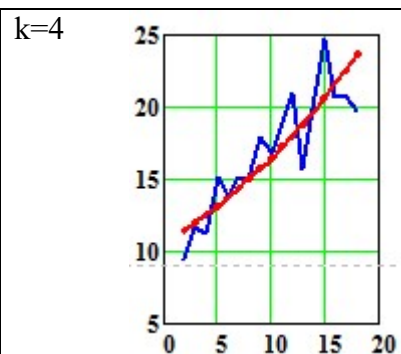


Рис.2.2 г). Модель №4

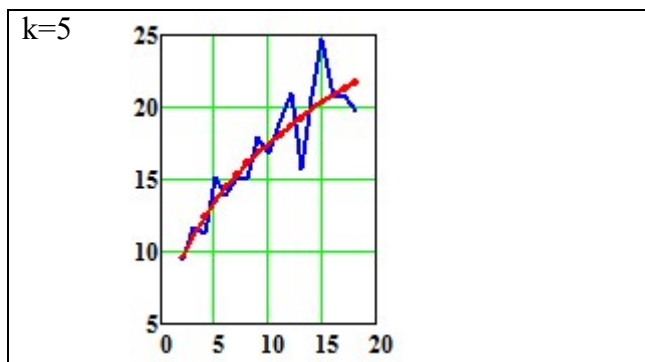


Рис.2.2 д). Модель №5

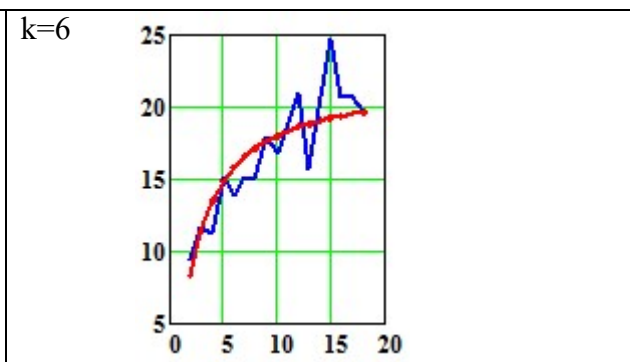


Рис.2.2 е). Модель №6

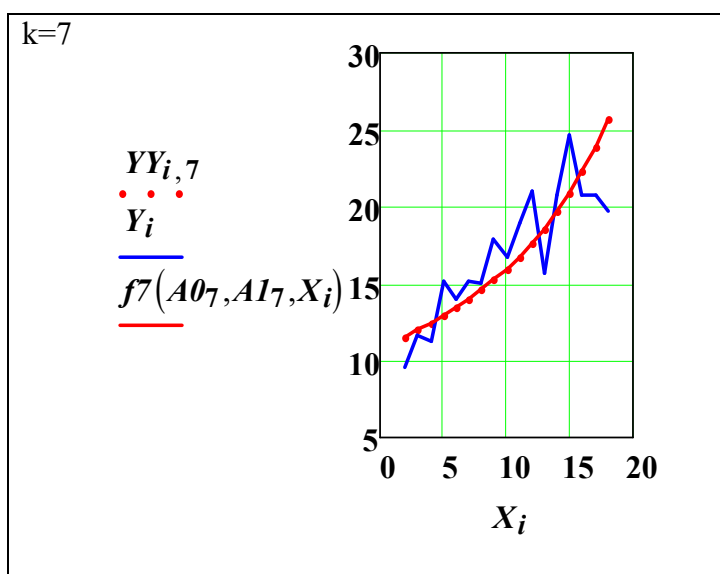


Рис.2.2 ж). Модель №7

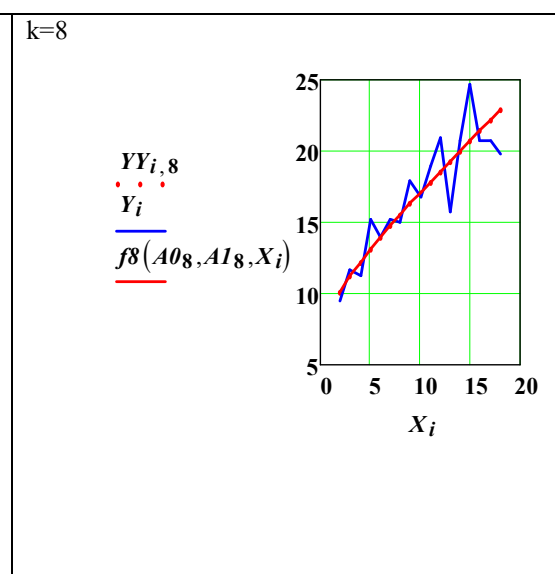


Рис.2.2 з). Модель №8

k=9

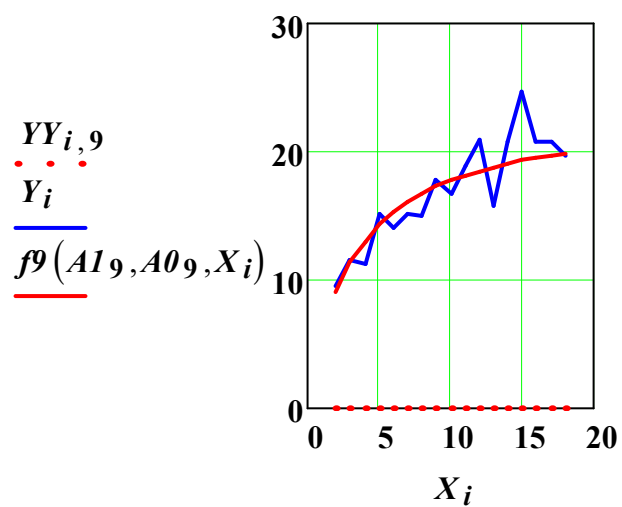


Рис.2.2 и). Модель №9

Візуальний метод є лише допоміжним при виборі моделі. Власне, цей приклад демонструє, що вибір потрібно робити серед всіх дев'яти моделей і це зробити досить складно. Щоб уникнути високоїмовірної помилки при візуальному виборі найкращої моделі, випишемо два стовпчики чисел. В першому – номери моделей, а в другому - обчислені для моделей із відповідними номерами величини сум квадратів відхилень конкретної k-ї моделі $Ym_{i,k}$ від реального виміру Y_i :

$$\sum_{i=1}^n (Y_i - Ym_{i,k})^2.$$

Таблиця 2.7

Оцінка ефективності моделей (упорядкована за номером моделі)

Номер моделі k	$\sum_{i=1}^n (Y_i - Ym_{i,k})^2$
1	60.053
2	51.777
3	84.786
4	73.75
5	50.479
6	70.604
7	108.688
8	57.629
9	62.031

Таблиця 2.8

Оцінка ефективності моделей (упорядкована за дисперсіями)

Номер моделі k	$\sum_{i=1}^n (Y_i - Ym_{i,k})^2$
5	50.479
2	51.777
8	57.639
1	60.053
9	62.031
6	70.604
4	73.75
3	84.786
7	108.688

Таблиця 2.8 утворюється перевпорядкуванням рядків таблиці 2.7 за зростанням сум квадратів ухилів, які є оцінками дисперсій з точністю до сталого множника. Найефективнішою моделлю, тобто моделлю з мінімальною дисперсією ухилів, виявилась п'ята модель. Слід відмітити, що моделі 5,2,1 є досить близькими за ефективністю. Для важливих реальних задач доцільно відслідковувати динаміку змін за кількома моделями.

2.3.Інтервальний прогноз за нелінійною моделлю

Прогнозування для нелінійних моделей викладається за [7].

Попередньо, оцінюючи найефективнішу із моделей, ми вирішили проблему лінеаризації задачі. Для обраної п'ятої моделі процес побудови квазілінійної моделі полягає в тому, щоб прологарифмувати виміри X та Y . Далі ми повинні не просто знайти коефіцієнти квазілінійної і основної моделі, а і дослідити:

- 1) статистичну значущість всієї моделі;
- 2) статистичну значущість кожного з коефіцієнтів квазілінійної моделі;
- 3) зробити інтервальну оцінку прогнозних значень за квазілінійною моделлю в лінеаризованій області;
- 4) повернутись назад до нелінійної області, бо саме в основній нелінійній області розмірності економічних показників правильно інтерпретуються;
- 5) зробити висновки за отриманими результатами.

2.4. Приклад аналізу нелінійної залежності

Аналіз нелінійної залежності за 5-ю моделлю

Перетворення вихідних даних для лінеаризації задачі

Таблиця 2.9

Лінеаризація

	C	D	E	F	G
19	X	Y		lnX	lnY
20	2	9.51	1	0.693147	2.252344
21	3	11.62	2	1.098612	2.452728
22	4	11.22	3	1.386294	2.417698
23	5	15.22	4	1.609438	2.72261
24	6	13.99	5	1.791759	2.638343
25	7	15.18	6	1.94591	2.719979
26	8	14.98	7	2.079442	2.706716
27	9	17.88	8	2.197225	2.883683
28	10	16.78	9	2.302585	2.820188
29	11	18.94	10	2.397895	2.941276
30	12	20.98	11	2.484907	3.04357
31	13	15.71	12	2.564949	2.754297
32	14	20.74	13	2.639057	3.032064
33	15	24.7	14	2.70805	3.206803
34	16	20.78	15	2.772589	3.033991
35	17	20.74	16	2.833213	3.032064
36	18	19.75	17	2.890372	2.983153
37	19		18	2.944439	
38	20		19	2.995732	
39	21		20	3.044522	

В таблиці червоним виділено точки прогнозу. Поки що клітинки D:37- D:39 та G:37- G:39 пусті. Більш інтенсивним кольором виділено дані квазілінійної моделі, які ми будемо обробляти точно так, як в завданні 1. За виключенням певних особливостей, наприклад при оцінюванні коефіцієнтів еластичності, ми фактично повторимо ті ж кроки, що і в завданні 1.

Обробка проводиться в іншому просторі, тому в кінці обробки потрібно повернутись в основний простір.

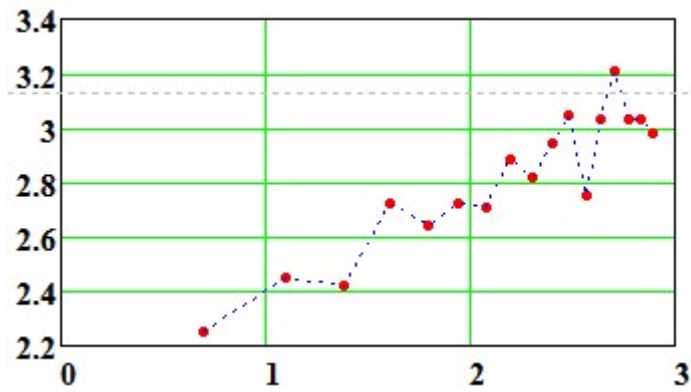


Рис. 2.3. Графік перетворених даних.

Навіть візуально помітно, що лінеарізація пройшла вдало.

Самостійно побудуйте описові статистики для перетворених даних.

Вкажемо лише, що коефіцієнт кореляції між векторами $\text{corr}(\ln X, \ln Y) = 0.929$ вказує на суттєвий зв'язок між факторами.

Будуємо звичайну лінійну регресію між $\ln Y$ та $\ln X$.

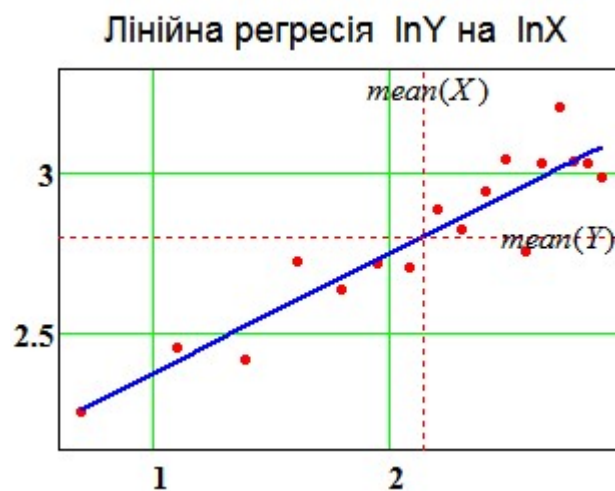


Рис 2.4. Побудова регресії квазілінійної моделі.

На графіку оператори $\text{mean}(X)=2.41$ $\text{mean}(Y)=2.802$ обчислюють середні в лінеаризованій області.

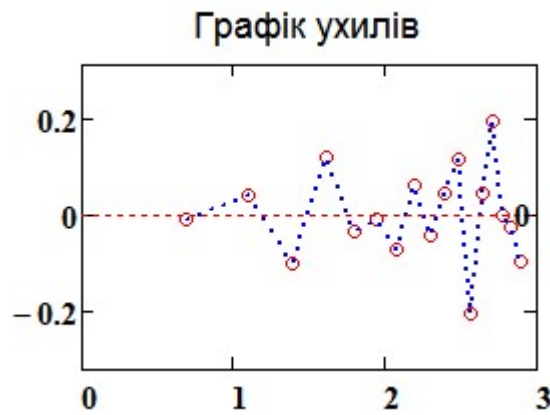


Рис. 2.5. Ухили квазілінійної моделі.

Коефіцієнт детермінації $R^2 = 0.864$.

Коефіцієнт множинної детермінації $1 - (1 - R^2) \frac{n-1}{m-3} = 0.844$ вказує на те,

що приблизно 85% відхилень пояснюється побудованою моделлю.

Значення статистики Дарбіна-Уотсона $DW = 2.556$. При порівнянні з табличним значення подвійна нерівність :

$$1.38 < 2.556 < 2.62.$$

Вказана подвійна нерівність дозволяє прийняти гіпотезу про відсутність автокореляції залишків.

Наступний графік довірчої області, дозволяє зробити попередні висновки, про загальну статистичну значущість моделі:

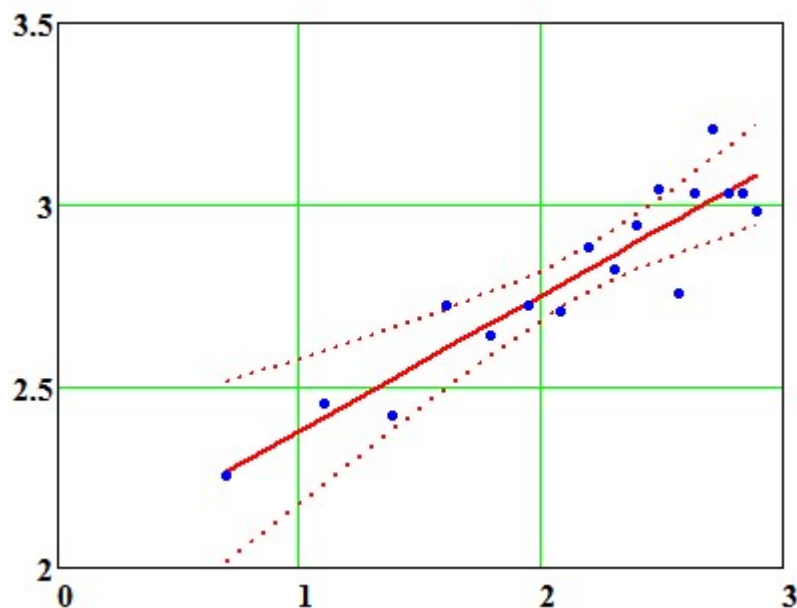


Рис.2.6. Довірча область дозволяє зробити попередні висновки про статистичну значущість побудованої квазілінійної моделі.

Трендові та інтервальні оцінки в лінеаризованій області подамо у виді таблиці 2.10.

Таблиця 2.10

Прогнозні інтервали оцінені в лінеаризованій області

Прогнозні точки		Прогнозні інтервали в лінеаризованій області		
2.944		2.917	3.102	3.288
2.996		2.935	3.121	3.308
3.045		2.952	3.14	3.327

Повернення до нелінійної області вказане в таблиці 2.11.

Таблиця 2.11

Повернення в нелінійну область

Прогнозні точки X		Інтервальний прогноз за нелінійною моделлю В центрі трендове значення Y		
19		18.482	22.248	26.781
20		18.82	22.678	27.327
21		19.146	23.094	27.857

Для наглядності зобразимо прогноз в виді графіка на рис. 2.7.

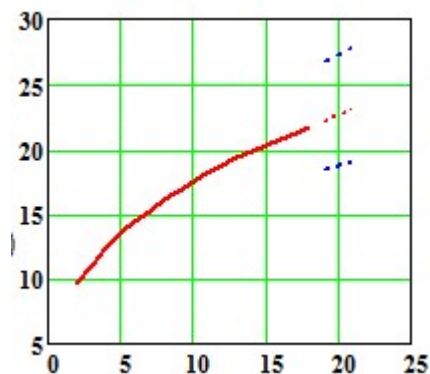


Рис.2.7. Графік трендових та інтервальних оцінок на рівні надійності 90% за нелінійною моделлю в вихідній (нелінійній) області .

Відмітимо, що інтервали несиметричні відносно точок трендового прогнозу .

Таблиця 2.12

2.5.Варіанти індивідуальних даних для практикуму №2

[illegible]

Варіант 9																		
	X	1.18	4.21	4.35	4.36	6.44	9.01	9.54	10.1	10.5	11.47	14.08	14.21	14.43	15.59	15.77	15.95	16
	Y	14	2.93	3	2.88	1.08	0.28	0.22	0.17	0.15	0.08	0.03	0.02	0.02	0.01	0.01	0.01	0.01
Варіант 10																		
	X	1.34	2.43	2.9	3.37	3.56	4.1	5.37	6.23	7.07	7.63	9.06	9.84	9.96	12.14	12.83	13.21	13.2
	Y	13.4	8.24	5.97	4.58	4.72	3.33	1.77	1.04	0.69	0.52	0.27	0.17	0.17	0.06	0.04	0.03	0.03
Варіант 11																		
	X	0.45	2.39	5.79	5.89	5.93	8.27	8.72	11.1	11.7	12.26	13.7	14.06	15.07	15.83	15.89	16.25	16.5
	Y	23.4	9.23	1.41	1.29	1.26	0.41	0.32	0.09	0.06	0.05	0.02	0.02	0.01	0.01	0.01	0.01	0
Варіант 12																		
	X	1.25	2.3	3.45	5.22	7.07	7.25	8.27	8.87	9.99	10.12	10.37	11.81	12.36	13.23	14.13	14.4	15
	Y	16.3	9.78	4.97	1.98	0.77	0.65	0.36	0.27	0.15	0.14	0.11	0.06	0.05	0.03	0.02	0.01	0.01
Варіант 13																		
	X	3.56	3.66	4.3	4.53	6.44	7.26	7.67	7.85	8.28	8.31	10.45	11.09	12.12	13.83	14.66	14.76	15.8
	Y	5.28	4.99	3.39	2.82	0.98	0.69	0.5	0.46	0.33	0.37	0.11	0.09	0.05	0.02	0.01	0.01	0.01
Варіант 14																		
	X	0.77	1	3.76	3.79	3.91	4.34	6.05	6.35	6.55	7.8	10.39	10.97	11.86	12.43	13.39	14.72	15.8
	Y	23.3	21.5	4.88	4.6	4.36	3.27	1.21	1.05	1	0.5	0.12	0.09	0.05	0.04	0.02	0.01	0.01
Варіант 15																		
	X	0.44	2.17	2.29	5.05	5.82	6.18	6.34	7.71	10.8	10.91	11.46	14.21	14.44	15.39	15.54	16.33	16.7
	Y	33.7	12.5	11	2.41	1.41	1.25	1.04	0.49	0.09	0.08	0.06	0.01	0.01	0.01	0.01	0.005	0.01
Варіант 16																		
	X	1	1.29	1.72	1.76	2.09	2.16	2.87	4.24	8.07	10.94	12.63	14.1	14.21	15.43	15.89	16.78	16.8
	Y	24.7	20.5	15.6	15.3	13.3	12	9.29	3.49	0.39	0.08	0.03	0.01	0.01	0.01	0.005	0.005	0.01
Варіант 17																		
	X	1.92	2.45	3.77	4.33	5.91	6.83	7.27	8.29	8.31	8.73	9.14	11.64	11.69	13.6	13.93	14.5	15.6
	Y	15.3	11	4.99	3.87	1.42	0.93	0.66	0.34	0.36	0.28	0.23	0.05	0.06	0.02	0.01	0.01	0.01

Варіант 18																		
	X	2.03	2.7	3.32	8.21	8.87	9.36	11.1	11.5	11.5	11.77	11.9	13	13.57	15.06	16.43	16.68	16.9
	Y	15.9	10.4	6.71	0.36	0.26	0.21	0.07	0.06	0.06	0.05	0.04	0.02	0.02	0.01	0.005	0.005	0.01
Варіант 19																		
	X	0.09	0.33	0.84	1.07	1.69	3.47	4.71	5.49	6.2	7.89	8.03	8.28	8.36	11.44	12.77	14.07	15.4
	Y	45.3	43.7	32.1	30.2	18.6	6.84	2.96	1.96	1.26	0.51	0.45	0.39	0.38	0.05	0.03	0.01	0.01
Варіант 20																		
	X	0.78	0.87	0.89	1.36	2.43	2.72	2.84	5.15	5.43	5.95	6.31	6.34	6.44	9.7	14.19	14.38	15.4
	Y	32.4	34.3	31.2	22.6	12.2	11	10.6	2.71	2.02	1.58	1.25	1.33	1.18	0.17	0.01	0.01	0.01
Варіант 21																		
	X	0.13	0.65	1.07	2.2	4.14	5.16	6.55	7.01	8.35	9.02	9.08	12.71	12.78	14.62	15.18	16.7	16.7
	Y	53	38.8	31.1	14.6	5.17	2.53	1.08	0.78	0.37	0.25	0.23	0.03	0.03	0.01	0.01	0.005	0.01
Варіант 22																		
	X	1.45	2.29	2.45	3.7	4.21	4.41	4.63	9.77	11.6	12.22	13.33	14.5	14.52	15.15	15.35	15.81	16.3
	Y	24.1	15.4	13.5	5.94	4.84	4.07	3.43	0.15	0.05	0.04	0.02	0.01	0.01	0.01	0.01	0.005	0.01
Варіант 23																		
	X	1.14	2.29	5.32	5.34	6.25	7.72	8.11	9.41	10.5	10.95	11.05	11.19	11.46	12.73	14.12	15.44	15.6
	Y	31.2	16	2.36	2.66	1.43	0.56	0.43	0.19	0.1	0.08	0.07	0.07	0.05	0.02	0.01	0.01	0.01
Варіант 24																		
	X	2.54	2.77	3.46	3.95	5.58	7.52	9.85	9.89	10.5	10.88	11.42	13.44	14.33	15.62	15.76	15.93	16.8
	Y	13.5	11.6	7.42	5.7	2.03	0.68	0.15	0.15	0.1	0.08	0.06	0.02	0.01	0.005	0.005	0.005	0.01
Варіант 25																		
	X	0.64	2.29	2.96	3.47	4.21	4.44	6.05	8.24	9.31	9.72	10.33	11.06	13.38	13.5	13.66	13.69	14.5
	Y	45.6	16.5	10.9	7.76	4.77	4.2	1.66	0.42	0.21	0.17	0.11	0.07	0.02	0.02	0.02	0.02	0.01
Варіант 26																		
	X	1.29	2.08	3.02	7.04	8.76	8.9	9.73	10.1	10.5	10.62	11.2	11.36	11.75	12.13	12.58	12.79	14
	Y	30.5	20.9	10.6	0.83	0.32	0.29	0.16	0.14	0.11	0.11	0.07	0.06	0.05	0.04	0.03	0.03	0.01

Варіант 27																		
	X	0.63	1.06	1.6	2.66	2.97	3.18	7.79	8.83	8.95	9.17	9.63	12.21	12.48	12.82	12.92	16.43	16.4
	Y	53.2	34.4	26	13.6	12.1	10.4	0.6	0.32	0.3	0.24	0.19	0.04	0.03	0.03	0.03	0.01	0.01
Варіант 28																		
	X	1.35	2.55	3.34	4.24	5.4	6.69	7.58	7.83	8.03	8.32	9.92	10.2	10.82	11.91	12.07	12.84	14.4
	Y	30	15.8	9.57	5.73	2.46	1.15	0.67	0.63	0.52	0.43	0.16	0.13	0.08	0.05	0.04	0.02	0.01
Варіант 29																		
	X	0.08	0.53	0.57	0.84	1.31	1.89	2.78	3.57	3.89	4.41	4.95	6.22	8.65	10.1	11.09	12.32	14.9
	Y	89.4	59.7	54.9	45.3	34.2	23.4	14.8	8.2	7.13	4.93	3.75	1.76	0.36	0.14	0.08	0.04	0.01
Варіант 30																		
	X	0.02	0.33	1.23	2.23	3.16	8.85	8.92	9.49	10.1	10.4	10.56	10.66	12.09	12.83	13.25	16.15	16.5
	Y	85.7	65.8	40.1	19.7	12.2	0.3	0.28	0.22	0.16	0.13	0.1	0.11	0.04	0.02	0.02	0.01	0.01

2.6.Контрольні питання для підготовки до виконання практикуму

- 1) Дати визначення середніх величин (арифметичної, геометричної, гармонічної);
- 2) Вказати нерівності для трьох середніх величин (арифметичної, геометричної, гармонічної);
- 3) Показати що функції 1)-9) зберігають властивості середніх величин, наприклад має місце властивість оцінок МНК $\bar{Y} = \hat{a}_0 + \hat{a}_1 \bar{X}$.
- 4) Дайте означення квазілінійної моделі.

- 5) Довести, що нелінійна модель $\hat{Y} = \frac{a}{X} + b$

є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = -\frac{a}{a + bX}.$$

- 6) Довести, що нелінійна модель $\hat{Y} = a \ln X + b$

є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = \frac{a}{a \ln X + b}.$$

- 7) Довести, що нелінійна модель $\hat{Y} = ae^X + b$

є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = \frac{aXe^X}{ae^X + b}.$$

- 8) Довести, що нелінійна модель $\hat{Y} = a\sqrt{X} + b$

є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = \frac{aX}{2aX + 2b\sqrt{X}}.$$

- 9) Довести, що нелінійна модель $\hat{Y} = aX^2 + b$

є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = \frac{2aX^2}{aX^2 + b}.$$

- 10) Довести, що нелінійна модель $\hat{Y} = aX^3 + b$

є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = \frac{3aX^3}{aX^3 + b}.$$

11) Довести, що нелінійна модель є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = \frac{-aX}{aX + b}.$$

12) Довести, що нелінійна модель $\hat{Y} = \frac{1}{ae^{-X} + b}$ є квазілінійною та, що її коефіцієнт еластичності знаходиться за формулою

$$K = \frac{aX}{a + be^X}.$$

13)-20) Віднайти рівняння між узагальненими середніми, яке є аналогом рівняння $\bar{y} = \hat{a}_0 + \hat{a}_1\bar{x}$ при специфікації моделі через функцію $f1(a_0, a_1, t)$.

Розділ 3. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ

3.1. Короткі теоретичні відомості

Введемо позначення

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + u.$$

Множинна лінійна регресія є узагальненням парної лінійної регресії на випадок, коли залежна змінна взаємозв'язана більш ніж з однією незалежною змінною. Ясно, що коли ми не врахуємо вплив невідомого фактору на результуючий, то непояснена частина збурень буде великою і побудована модель або зовсім статистично незначущою, або суттєво гіршою, ніж модель, яка додатковий фактор враховуватиме. Цікаво, що саме такі міркування приводять до отримання нових наукових і практичних результатів.

Схема аналізу множинної лінійної регресії подібна до аналізу парної регресії.

Для звичайної та множинної лінійних регресій має місце теорема Гауса-Маркова:

Припустимо, що $Y = XA + u$, де X – детермінована матриця $n \times k$, що має максимальний ранг і

- 1) $E(u) = 0$;
- 2) $D(u) = \sigma^2$;
- 3) $Cov(u_i, u_k) = 0, i \neq k$;
- 4) $Cov(x_{ij}, u_i) = 0$;

то оцінка методу найменших квадратів $\hat{A} = (X^T X)^{-1} X^T Y$ є (за теоремою Гауса-Маркова) незміщеною, ефективною та слушною.

Крім умов 1)-4), необхідних для застосування теореми Гауса-Маркова при побудові множинної лінійної регресії, додатково будемо вимагати виконання наступних передумов:

- 5) модель є лінійною щодо параметрів;
- 6) між пояснюючими змінними немає строгої або сильної залежності;
- 7) помилка u має гаусівський розподіл з параметрами 0 і σ .

Порушення хоча б одного із припущень МНК приводить до погіршення якості моделі. У цьому випадку оцінку не можна вважати незміщеною, ефективною і слушною. Так порушення другої умови приводить до гетероскедастичності, третього – до автокореляції, шостого – до мультиколінійності. Виявлення цих явищ є необхідним елементом аналізу

множинної лінійної регресії. Фахівець повинен володіти теоретичними та практичними знаннями щодо методів виявлення автокореляції, гетероскедастичності й мультиколінійності, а також методів усунення або суттєвого зменшення негативного впливу таких явищ.

Не менш важливою задачею фахівця є правильна специфікація моделі, тому що як включення несуттєвих змінних так і пропуск суттєвих змінних, дуже впливають на якість моделі. Оскільки включення факторів в модель виконується, як правило, послідовно (фактор за фактором), то навіть порядок включення факторів може бути важливим.

3.1.1. Гетероскедастичність і її наслідки

У другій передумові вимагається, щоб дисперсія випадкового залишку в кожному спостереженні була постійною (саме дисперсія залишку, а не залишок). Розширимо пояснення.

Коли ми записуємо модель $Y = XA + u$, перші дві умови Гауса-Маркова вказують, що випадкові члени $u_1, \dots, u_n$ в n спостереженнях з'являються на основі імовірнісних розподілів, що мають нульове математичне очікування і одну й ту ж незмінну дисперсію. Збурення моделі в кожний з моментів спостереження повинні бути однаково розподілені. Така сталість розподілів називаються гомоскедастичністю. При порушенні гомоскедастичності ми матимемо нестационарний, гетероскедастичний випадковий процес.

Математично гомоскедастичність і гетероскедастичність можуть визначатися в такий спосіб:

гомоскедастичність: $\text{Var}(e_i) = \sigma^2$ постійна для всіх спостережень;

гетероскедастичність: $\text{Var}(e_i) = \sigma_i^2$, де σ_i^2 не обов'язкові однакові для всіх номерів i .

Для чого вводиться вимога про відсутність гетероскедастичності? По-перше, бажано, щоб дисперсія МНК оцінок була найменшою, щоб забезпечувати максимальну точність. При відсутності гетероскедастичності звичайні коефіцієнти регресії мають найменшу дисперсію серед усіх незміщених оцінок, які відшукуються у виді лінійних функцій від спостережень. Якщо має місце гетероскедастичність, то оцінки МНК є неефективними. Гетероскедастичність може приводити до заниження оцінок стандартної похибки, що призведе до недооцінки ризику. Цілком імовірно, що стандартні помилки будуть занижені, а отже, t -статистика — завищена, буде отримана неправильна уявлення про точність оцінки рівняння регресії. Можливо, буде прийнята невірна гіпотеза, що коефіцієнт значущо відрізняється від нуля на даному рівні значущості, тоді як у дійсності це не так.

3.1.2.Виявлення гетероскедастичності

Для виявлення гетероскедастичності використовуються тести

рангової кореляції Спірмена, Гольдфельда-Квандтата тест Глейзера [8].

3.1.3.Тест рангової кореляції Спірмена

При виконанні тесту рангової кореляції Спірмена передбачається, що дисперсія випадкового члена буде або збільшуватися, або зменшуватися в міру збільшення x_i , і тому в регресії, оціненої за допомогою МНК, абсолютні величини залишків і значення пояснюючих змінних будуть корельованими. Дані про x_i та залишки впорядковуються, і коефіцієнт рангової кореляції визначається як

$$r_{x_i,u} = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)},$$

де D_i — різниця між рангом x_i і рангом u . Якщо припустити, що коефіцієнт кореляції для генеральної сукупності дорівнює нулю, то коефіцієнт рангової кореляції має нормальний розподіл з математичним очікуванням 0 і дисперсією $\frac{1}{(n-1)}$.

3.1.4.Тест Гольдфельда –Квандта на гетероскедастичність

В тесті Гольдфельда –Квандта робиться припущення про залежність збурень u від одного з факторів x_i , найчастіше беруться найпростіші залежності [7,8].

3.2.Мультиколінійність

3.2.1. Ознаки наявності мультиколінійності

Існує кілька таких ознак, за якими може бути встановлено наявність мультиколінійності:

1. Коефіцієнт детермінації достатньо високий, але деякі з коефіцієнтів регресії статистично незначущі, тобто вони мають низькі t -статистики.
2. Парна кореляція між малозначущими пояснювальними змінними достатньо висока.

Проте ця ознака буде надійною лише у випадку двох пояснювальних змінних, а при більшій їх кількості доцільніше використовувати частинні коефіцієнти кореляції.

3. Високі частинні коефіцієнти кореляції.

Частинні коефіцієнти кореляції визначають силу лінійної залежності між двома змінними без врахування впливу на них інших змінних. Проте при вивченні багатовимірних зв'язків в ряді випадків парні коефіцієнти кореляції можуть давати зовсім хибні уявлення про характер зв'язку між двома змінними.

3.2.2 Наслідки мультиколінійності:

- 1). Великі дисперсії (стандартні помилки) оцінок. Це утруднює знаходження дійсних обумовлених величин і розширює інтервальні оцінки, погіршуючи їх точність.
- 2). Зменшуються t -статистики коефіцієнтів, що може привести до не виправданого висновку про істотність впливу відповідної пояснюючої змінної на залежну змінну.
- 3). Оцінки коефіцієнтів по МНК і їх стандартні помилки стають дуже чутливими до найменших змін даних, тобто вони стають нестійкими.
- 4). Утрудняється визначення внеску кожної з пояснюючої змінних.
- 5). Можливе одержання невірної знака в коефіцієнті регресії.

3.2.3 Правило відбору факторів в множині мультиколінійних

Продемонструємо правило на прикладі.

Нехай, наприклад, в залежності $y = f(x_1, x_2, x_3)$ матриця парних коефіцієнтів кореляції наступна:

Таблиця 3.1

	y	x_1	x_2	x_3
y	1	0,8	0,7	0,6
x_1	0,8	1	0,8	0,5
x_2	0,7	0,8	1	0,2
x_3	0,6	0,5	0,2	1

Видно, що фактори x_1 і x_2 дублюють один одного. В модель доцільно включити фактор x_2 , а не x_1 , хоча кореляція x_2 з результатом y слабша, ніж кореляція фактора x_1 з y ($r_{yx_2} = 0,7 < r_{yx_1} = 0,8$), але значно слабша міжфакторна кореляція $r_{x_2x_3} = 0,2 < r_{x_1x_3} = 0,5$. Тому в даному випадку в рівняння множинної регресії включаються фактори x_2 , x_3 .

3.3. Перевірка значущості коефіцієнта детермінації

Для перевірки статистичної значущості коефіцієнта детермінації R^2 висувається нульова гіпотеза $H_0: R^2=0$. Це означає, що досліджуване рівняння не пояснює зміну залежної змінної (Y) під впливом відповідних незалежних змінних. У такому разі всі коефіцієнти при незалежних змінних мають дорівнювати нулю. При цьому нульову гіпотезу можна подати у вигляді

$$H_0: \beta_0 = \beta_1 = \dots = \beta_m = 0.$$

Альтернативною до неї є H_A : значення хоча б одного параметра моделі відмінне від нуля ($\beta_i \neq 0$), тобто хоча б один із факторів впливає на змінювання залежної змінної.

Для перевірки цих гіпотез застосовують F -критерій Фішера з m і $n-m-1$ ступенями вільності. За отриманими в моделі значеннями коефіцієнта детермінації R^2 обчислюють експериментальне значення F -статистики:

$$F_{\text{експ}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

яке порівнюють з табличним значенням розподілу Фішера при заданому рівні значущості α (як правило, $\alpha = 0,05$ або $\alpha = 0,01$). Якщо $F_{\text{експ}} > F_{\text{таб}}$, то нульова гіпотеза відхиляється, тобто існує такий коефіцієнт у регресійному рівнянні, який суттєво відрізняється від нуля, а відповідний фактор впливає на досліджувану змінну. Відхилення нуль-гіпотези свідчить про адекватність побудованої моделі. У протилежному випадку модель вважається неадекватною.

3.4. Перевірка значущості коефіцієнта кореляції

Коефіцієнт кореляції, як вибіркова характеристика, перевіряється на значущість за допомогою t -критерію Стюдента. Фактичне значення t -статистики обчислюється за відповідною формулою

і порівнюється з табличним значенням t -розподілу з $n-m-1$ ступенями вільності та при заданому рівні значущості $\alpha/2$ (такий рівень зумовлений тим, що критична область складається з двох проміжків).

Якщо абсолютна величина експериментального значення t -статистики перевищує табличне, тобто $|t_{\text{експ}}| > t_{\text{табл}}$, можна зробити висновок, що коефіцієнт кореляції достовірний (значущий), а зв'язок між залежною змінною та всіма незалежними факторами суттєвий.

3.5. Оцінка статистичної значущості параметрів моделі

Окрім загальних показників адекватності моделі існують також оцінки, що дають змогу встановити якість окремих частин рівняння, зокрема одного чи кількох коефіцієнтів регресії.

Як і в попередніх випадках, рішення відносно якості коефіцієнтів приймають на основі відповідних статистичних критеріїв.

Статистичну значущість кожного параметра моделі можна перевірити за допомогою t -критерію. При цьому нульова гіпотеза має вигляд

$$H_0: \beta_j = 0,$$

альтернативна гіпотеза

$$H_A: \beta_j \neq 0.$$

Експериментальне значення t -статистики для кожного параметра моделі обчислюється за формулою

$$t_j = \frac{\beta_j}{\sqrt{\sigma_u^2 C_{jj}}} = \frac{\beta_j}{S_{\beta_j}}$$

де C_{jj} — діагональний елемент матриці $(X'X)^{-1}$;

S_{β_j} — стандартна похибка оцінки параметра моделі:

$$S_{\beta_j} = \sigma_u \sqrt{C_{jj}}.$$

Експериментальне значення t_j -критерію порівнюється з табличним значенням $t_{\text{табл}}$ з $n-m-1$ ступенями вільності при заданому рівні значущості $\alpha/2$ (критична область розбивається на два фрагменти, межі яких задаються квантилем $\alpha/2$). Якщо значення t -статистики потрапляє до критичної області (за абсолютним значенням перевищує $t_{\text{табл}}$), приймається альтернативна гіпотеза про значущість відповідного параметра. Інакше робиться висновок про статистичну незначущість параметра β_j , а це означає, що відповідна незалежна змінна не впливає суттєво на зміну залежної змінної.

3.6. Прогнозування за загальною лінійною моделлю

Якщо побудована модель адекватна за F -критерієм, то її застосовують для прогнозування залежної змінної. Про прогнозування за моделлю говорять тоді, коли в часових рядах прогнозний період настає пізніше, ніж базовий. Якщо регресія побудована за просторовими даними, прогноз стосується тих елементів генеральної сукупності, що перебувають за межами застосованої вибірки.

Припустимо, що ми побудували модель та оцінили параметри методом найменших квадратів. На підставі побудованої моделі можна знайти прогнозні значення матриці залежних змінних Y_{np} , які відповідають очікуваним значенням матриці незалежних змінних X_{np} .

Прогноз на перспективу буває двох видів: точковий та інтервальний.

Незміщена оцінка точкового прогнозу може розглядатися як точкова оцінка математичного сподівання прогнозного значення Y_{np}

$$M[Y_{np}(X_{np})] = BX_{np},$$

а також як індивідуальне значення Y_{np} для матриці незалежних змінних X_{np} , що лежать за межами базового періоду $Y_{np} = BX_{np}$.

Дисперсія похибки прогнозу дорівнює

$$\sigma_n^2 = \sigma_u^2 X'_{np} (X'X)^{-1} X_{np} = X'_{np} \text{var}(B) X_{np},$$

де σ_u^2 – дисперсія залишків u , яка розраховується за формулою (2.9); $\text{var}(B)$ – дисперсійно-коваріаційна матриця, яка записується у вигляді:

Елементи на головній діагоналі матриці $\sigma_{\beta_j}^2$ та за її межами $\sigma_{\beta_j \beta_k}$ розраховуються за формулами:

$$\sigma_{\beta_j}^2 = \sigma_u^2 c_{jj}, \quad \sigma_{\beta_j \beta_k} = \sigma_u^2 c_{jk},$$

де c_{jj}, c_{jk} – елементи матриці похибок $(X'X)^{-1}$.

Тоді дисперсія прогнозу буде:

$$\sigma_n^2 = X'_{np} \text{var}(B) X_{np}$$

Середньоквадратична (стандартна) похибка прогнозу:

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

Довірчий інтервал для прогнозних значень

t_α – табличне значення t -критерія Стюдента з $(n-m-1)$ ступенями вільності
 α – рівень значущості.

Для використання t -критерія Стюдента необхідно обрати бажаний рівень значущості α (0,05 або 0,01) та число ступенів вільності $(n-m-1)$.

Інтервальний прогноз математичного сподівання $M(Y_{np})$ буде в межах:

$$Y_{np} - t_\alpha \sigma_n \leq M(Y_{np}) \leq Y_{np} + t_\alpha \sigma_n$$

Визначення інтервального прогнозу індивідуального значення Y_{np} базується на знаходженні середньоквадратичної помилки прогнозу:

$$\sigma_{n(i)}^2 = \sigma_u^2 + \sigma_n^2 = \sigma_u^2 + \sigma_u^2 X'_{np} (X'X)^{-1} X_{np} = \sigma_u \sqrt{1 + X'_{np} (X'X)^{-1} X_{np}}$$

Тоді інтервальний прогноз індивідуального значення буде відповідати такому довірчому інтервалу:

$$Y_{np} - t_\alpha \sigma_{n(i)} \leq Y_{np} \leq Y_{np} + t_\alpha \sigma_{n(i)}.$$

3.6.1. Мета та основні завдання комп'ютерного практикуму до розділу 3

3.6.2. Специфікація моделі .

Розглянемо рівняння множинної регресії, яке може бути представлене у вигляді:

$$Y = f(\beta, X) + u$$

де $X = (X_1, X_2, \dots, X_m)^T$ - вектор незалежних пояснюючих змінних; β - вектор параметрів, які потрібно визначити; u - випадкова помилка; Y - залежна змінна.

Теоретичне рівняння множинної лінійної регресії має вигляд:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m + u$$

β_0 - вільний член, що визначає значення Y для випадку, коли всі пояснюючі змінні X_j рівні 0.

Починаючи будувати модель ми неявно робимо припущення, про виконання передумов, нагадаємо їх:

1. Математичне очікування випадкового відхилення u_i рівне 0 для всіх спостережень ($M(u_i) = 0$).
2. Гомоскедастичність (сталість дисперсій відхилень). Дисперсія випадкових відхилень u_i постійна: $D(u_i) = S^2$ для будь-яких i .
3. Відсутність автокореляції.
4. Випадкове відхилення є незалежним від пояснюючих змінних.
5. Модель є лінійною щодо параметрів.
6. Мультиколінійність відсутня. Між пояснюючими змінними відсутня строга лінійна залежність.
7. Помилки мають нормальний розподіл. Виконання даної передумови важлива для перевірки статистичних гіпотез і побудови довірчих інтервалів.

Емпіричне рівняння множинної регресії представимо у вигляді:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_m X_m + \hat{u},$$

тут $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$ - емпіричні коефіцієнти регресії (оцінки теоретичних значень $\beta_0, \beta_1, \dots, \beta_m$ коефіцієнтів регресії); $\hat{u}$ - оцінка відхилення u .

При виконанні передумов МНК щодо помилок u_i , оцінки $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$ параметрів $\beta_0, \beta_1, \dots, \beta_m$ множинної лінійної регресії по МНК є незміщеними, ефективними й слушними (тобто Blue-Оцінками).

Для оцінки параметрів рівняння множинної регресії застосовують МНК.

3.6.3. Оцінка рівняння регресії

Визначимо вектор оцінок коефіцієнтів регресії. Згідно з методом найменших квадратів, вектор s виходить із вираження:

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

Матриця X

1	95.98	18.23	10.31
1	90.93	20.35	11.65
1	107.29	18.05	12.31
1	116.18	21.78	12.72
1	128.9	24.86	13.83
1	147.09	32	16.14
1	151.94	36.24	17.01
1	165.06	46.4	18.76
1	183.09	54.66	21.45
1	202.93	76.64	26.27
1	228.06	84.4	29.76
1	255.09	99.66	32.45
1	289.93	132.64	36.27
1	318.39	168.62	42.87
1	350.9	171.59	49.72
1	379.87	184.99	53.24
1	404.32	207.51	60.29
1	436.73	320.84	82.45

Матриця Y

64.54
87.52
94.58
107.62
125.05
142.28
156.55
179.28
209.3
245.52
279.28
329.3

385.52
430.05
484.87
522.03
570.39
741.82

Матриця X^T

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
95.98	90.93	107.29	116.18	128.9	147.09	151.94	165.06	183.09	202.93	228.06	255.09	289.93	318.39	350.9	379.87	404.32	436.73	
18.23	20.35	18.05	21.78	24.86	32	36.24	46.4	54.66	76.64	84.4	99.66	132.64	168.62	171.59	184.99	207.51	320.84	
10.31	11.65	12.31	12.72	13.83	16.14	17.01	18.76	21.45	26.27	29.76	32.45	36.27	42.87	49.72	53.24	60.29	82.45	

Множимо матриці, $(X^T X)$

18	4052.68	1719.46	547.5
4052.68	1129926.79	546021.55	161229.18
1719.46	546021.55	288261.03	81586.92
547.5	161229.18	81586.92	23630.09

У матриці $(X^T X)$ число 18, що лежить на перетинанні 1-й рядка й 1-го стовпця, отримане як сума добутків елементів 1-й рядка матриці X^T і 1-го стовпця матриці X . Знаходимо добуток

$$X^T Y = \begin{array}{|c|} \hline 5155.5 \\ \hline 1534803.43 \\ \hline 775095.39 \\ \hline 224078.26 \\ \hline \end{array}$$

Знаходимо обернену матрицю $(X^T X)^{-1} =$

1.35	-0.00428	0.0281	-0.0991
------	----------	--------	---------

-0.00428	9.3E-5	3.9E-5	-0.000672
0.0281	3.9E-5	0.000948	-0.00419
-0.0991	-0.000672	-0.00419	0.0214

Вектор оцінок коефіцієнтів регресії рівний

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{array}{|c|} \hline -34.12 \\ \hline 0.76 \\ \hline 0.56 \\ \hline 3.13 \\ \hline \end{array}$$

Рівняння регресії (оцінка рівняння регресії)

$$\hat{y} = -34.12 + 0.76x_1 + 0.56x_2 + 3.13x_3$$

3.6.4. Матриця парних коефіцієнтів кореляції.

Число спостережень $n = 18$. Число незалежних змінних у моделі рівно 3, а число регресорів із урахуванням одиничного вектора дорівнює числу невідомих коефіцієнтів. З урахуванням ознаки Y , розмірність матриці стає рівним 5. Матриця незалежних змінних X має розмірність (18 х 5). Матриця $(X^T X)$ визначається безпосереднім множенням або за наступними попередньо обчислених сумами.

Матриця складена з Y і X

1	64.54	95.98	18.23	10.31
1	87.52	90.93	20.35	11.65
1	94.58	107.29	18.05	12.31
1	107.62	116.18	21.78	12.72
1	125.05	128.9	24.86	13.83
1	142.28	147.09	32	16.14
1	156.55	151.94	36.24	17.01
1	179.28	165.06	46.4	18.76
1	209.3	183.09	54.66	21.45
1	245.52	202.93	76.64	26.27
1	279.28	228.06	84.4	29.76
1	329.3	255.09	99.66	32.45
1	385.52	289.93	132.64	36.27
1	430.05	318.39	168.62	42.87
1	484.87	350.9	171.59	49.72

1	522.03	379.87	184.99	53.24
1	570.39	404.32	207.51	60.29
1	741.82	436.73	320.84	82.45

Транспонована матриця.

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
64.54	87.52	94.58	107.62	125.05	142.28	156.55	179.28	209.3	245.52	279.28	329.3	385.52	430.05	484.87	522.03	570.39	741.82
95.98	90.93	107.29	116.18	128.9	147.09	151.94	165.06	183.09	202.93	228.06	255.09	289.93	318.39	350.9	379.87	404.32	436.73
18.23	20.35	18.05	21.78	24.86	32	36.24	46.4	54.66	76.64	84.4	99.66	132.64	168.62	171.59	184.99	207.51	320.84
10.31	11.65	12.31	12.72	13.83	16.14	17.01	18.76	21.45	26.27	29.76	32.45	36.27	42.87	49.72	53.24	60.29	82.45

Перемножаємо дві останні матриці:

18	5155.5	4052.68	1719.46	547.5
5155.5	2132229.87	1534803.43	775095.39	224078.26
4052.68	1534803.43	1129926.79	546021.55	161229.18
1719.46	775095.39	546021.55	288261.03	81586.92
547.5	224078.26	161229.18	81586.92	23630.09

Отримана матриця може бути розрахована безпосередньо через суми:

$\sum n$	$\sum y$	$\sum x_1$	$\sum x_2$	$\sum x_3$
$\sum y$	$\sum y^2$	$\sum x_1 y$	$\sum x_2 y$	$\sum x_3 y$
$\sum x_1$	$\sum y x_1$	$\sum x_1^2$	$\sum x_2 x_1$	$\sum x_3 x_1$
$\sum x_2$	$\sum y x_2$	$\sum x_1 x_2$	$\sum x_2^2$	$\sum x_3 x_2$
$\sum x_3$	$\sum y x_3$	$\sum x_1 x_3$	$\sum x_2 x_3$	$\sum x_3^2$

Знайдемо парні коефіцієнти кореляції.

Ознаки x і y	$\sum x_i$		$\sum y_i$		$\sum x_i y_i$	
Для y і x_1	4052.68	225.15	5155.5	286.42	1534803.43	85266.86
Для y і x_2	1719.46	95.53	5155.5	286.42	775095.39	43060.86
Для y і x_3	547.5	30.42	5155.5	286.42	224078.26	12448.79
Для x_1 і x_2	1719.46	95.53	4052.68	225.15	546021.55	30334.53
Для x_1 і x_3	547.5	30.42	4052.68	225.15	161229.18	8957.18
Для x_2 і x_3	547.5	30.42	1719.46	95.53	81586.92	4532.61

Ознаки x і y					
Для y і x_1	12081.69	36422.71	109.92	190.85	0.99
Для y і x_2	6889.37	36422.71	83	190.85	0.99
Для y і x_3	387.61	36422.71	19.69	190.85	0.99
Для x_1 і x_2	6889.37	12081.69	83	109.92	0.97
Для x_1 і x_3	387.61	12081.69	19.69	109.92	0.97
Для x_2 і x_3	387.61	6889.37	19.69	83	1

Матриця парних коефіцієнтів кореляції:

-	y	x_1	x_2	x_3
y	1	0.99	0.99	0.99
x_1	0.99	1	0.97	0.97
x_2	0.99	0.97	1	1
x_3	0.99	0.97	1	1

Аналіз першого рядка цієї матриці дозволяє зробити відбір факторних ознак, які можуть бути включені в модель множинної кореляційної залежності. Факторні ознаки, у яких $r_{yx_i} < 0.5$ виключають із моделі.

Колінеарність – залежність між факторами. У якості критерію мультиколінеарності може бути прийняте виконання наступних нерівностей:

$$r(x_i y) > r(x_k x_j) ; r(x_k y) > r(x_k x_j).$$

Якщо одна з нерівностей не виконується, то виключається той параметр x_k або x_j , зв'язок якого з результативним показником Y виявляється найменш тісним.

При оцінці мультиколінеарності факторів слід ураховувати, що чим ближче до 0 визначник матриці міжфакторної кореляції, тим сильніша мультиколінеарність факторів і ненадійніші результати множинної регресії.

Для відбору найбільш значущих факторів x_i враховуються наступні умови:

- зв'язок між результативною ознакою й фактором повинний бути вищим міжфакторного зв'язку;
- зв'язок між факторами повинен бути не більшим 0.7;

- при високому міжфакторному зв'язку відбираються фактори з меншим коефіцієнтом кореляції між собою.

Більш об'єктивну характеристику тісноти зв'язку дають частинні коефіцієнти кореляції, що вимірюють вплив фактору x_i на результат при незмінному рівні інших факторів.

3.6.5. Модель регресії в стандартному масштабі

Модель регресії в стандартному масштабі припускає, що всі значення досліджуваних ознак приводяться до стандартних (стандартизованих) значень за формулами:

$$t_j = \frac{x_{ji} - \bar{x_j}}{S(x_j)}$$

де x_{ji} - значення змінної x_j в i -ому спостереженні.

Таким чином, початок відліку кожної стандартизованої змінної прив'язується до її середнього значення, а в якості одиниці зміни приймається її середнє квадратичне відхилення S .

Якщо зв'язок між змінними в природньому масштабі лінійний, то зміна початку відліку й одиниці виміру цієї властивості не порушать, так що й стандартизовані змінні будуть зв'язані лінійним співвідношенням:

$$t_y = \sum \beta_j t_{xj}$$

3.6.6. Порядок і рекомендації щодо виконання практикуму

Для оцінки β - коефіцієнтів застосуємо МНК. При цьому система нормальних рівнянь буде мати вигляд:

$$\begin{aligned} rx_{1y} &= \beta_1 + rx_{1x_2} \cdot \beta_2 + \dots + rx_{1x_m} \cdot \beta_m \\ rx_{2y} &= rx_{2x_1} \cdot \beta_1 + \beta_2 + \dots + rx_{2x_m} \cdot \beta_m \\ &\dots \\ rx_{my} &= rx_{mx_1} \cdot \beta_1 + rx_{mx_2} \cdot \beta_2 + \dots + \beta_m \end{aligned}$$

Для наших даних:

$$\begin{aligned} 0.990615873815 &= \beta_1 + 0.967525478136\beta_2 + 0.974529079396\beta_3 \\ 0.991161346661 &= 0.967525478136\beta_1 + \beta_2 + 0.995659600045\beta_3 \\ 0.994568011011 &= 0.974529079396\beta_1 + 0.995659600045\beta_2 + \beta_3 \end{aligned}$$

Звідки знаходимо:

$$\beta_1 = 0.43865889782842; \beta_2 = 0.24554383055896; \beta_3 = 0.32260408680531;$$

Стандартизована форма рівняння регресії має вигляд:

$$y^0 = 0.439x_1 + 0.246x_2 + 0.323x_3$$

Знайдені з даної системи β -коефіцієнти дозволяють визначити значення коефіцієнтів у регресії в природньому масштабі за формулами:

$$b_j = \beta \frac{S(y)}{S(x_j)}$$

.

3.6.7. Аналіз параметрів рівняння множинної регресії

Перейдемо до статистичного аналізу отриманого рівняння регресії:

- перевірці значущості рівняння і його коефіцієнтів;
- дослідженню абсолютних і відносних помилок апроксимації.

Для незміщеної оцінки дисперсії зробимо наступні обчислення:
Незміщена помилка $u = Y - Y(x) = Y - X*s$ (абсолютна помилка апроксимації)

Y	$Y(x)$	U	$(Y - Y_{сер})^2$
64.54	81.52	-16.98	49229.26
87.52	83.06	4.46	39559.88
94.58	96.29	-1.71	36801.31
107.62	106.44	1.18	31968.25
125.05	121.34	3.71	26039.2
142.28	146.45	-4.17	20775.38
156.55	155.26	1.29	16865.35
179.28	176.46	2.82	11478.27
209.3	203.27	6.03	5946.98
245.52	245.86	-0.34	1672.54
279.28	280.3	-1.02	50.93
329.3	317.91	11.39	1838.98
385.52	375.02	10.5	9821.47
430.05	437.64	-7.59	20630.53
484.87	485.5	-0.63	39383.73
522.03	526.14	-4.11	55513.64
570.39	579.53	-9.14	80640.85
741.82	737.49	4.33	207392.2
			655608.74

$$se^2 = (Y - X*s)^T(Y - X*s) = 807.89$$

Незміщена оцінка дисперсії рівна:

$$s^2 = \frac{1}{n-m-1} s_e^2 = \frac{1}{18-4} 807.89 = 57.71$$

Оцінка середньоквадратичного відхилення рівна (стандартна помилка для оцінки Y):

$$S = \sqrt{S^2} = \sqrt{57.71} = 7.6$$

Знайдемо оцінку коваріаційної матриці вектора $k = S \cdot (X^T X)^{-1}$

10.24	-0.0325	0.21	-0.75
-0.0325	0.000707	0.0003	-0.00511
0.21	0.0003	0.0072	-0.0319
-0.75	-0.00511	-0.0319	0.16

Дисперсії параметрів моделі визначаються співвідношенням $S_i^2 = K_{ii}$, тобто це елементи, що лежать на головній діагоналі.

3.6.8 Показники тісноти зв'язку факторів з результатом

Якщо факторні ознаки різні за своєю сутністю (або) мають різні одиниці виміру, то коефіцієнти регресії b_j при різних факторах є несумірні. Тому рівняння регресії доповнюють сумірними показниками тісноти зв'язку фактору з результатом, що дозволяють ранжувати фактори за силою впливу на результат.

До таких показників тісноти зв'язку відносять:

частинні коефіцієнти еластичності, β -коефіцієнти, частинні коефіцієнти кореляції.

3.6.9. Частинні коефіцієнти еластичності.

З метою розширення можливостей змістовного аналізу моделі регресії використовуються частинні коефіцієнти еластичності, які визначаються за формулою:

$$E_i = b_i \frac{\overline{x_i}}{\overline{y}}$$

Частинний коефіцієнт еластичності показує, наскільки відсотків у середньому змінюється ознака-результат упри збільшенні ознаки-фактору x_j на 1% від свого середнього рівня при фіксованому положенні інших факторів моделі.

$$E_1 = 0.76 \frac{225.15}{286.42} = 0.6$$

Частинний коефіцієнт еластичності $|E_1| < 1$. Отже, його вплив на результативну ознаку Y незначний.

$$E_2 = 0.56 \frac{95.53}{286.42} = 0.19.$$

Частинний коефіцієнт еластичності $|E_2| < 1$. Отже, його вплив на результативну ознаку Y незначний.

Частинний коефіцієнт еластичності $|E_3| < 1$. Отже, його вплив на результативну ознаку Y незначний.

3.6.10. Стандартизовані частинні коефіцієнти регресії.

Стандартизовані частинні коефіцієнти регресії - β_j -коефіцієнти (β_j) показують, на яку частину свого середнього квадратичного відхилення $S(y)$ зміниться ознака-результат у зі зміною відповідного фактору x_j на величину свого середнього квадратичного відхилення (S_{x_j}) при незмінному впливі інших факторів (що входять у рівняння).

По максимальному β_j можна робити висновки про те, який з факторів сильніше впливає на результат Y .

За коефіцієнтами еластичності й β -коефіцієнтам можуть бути зроблені протилежні висновки. Причини цього:

- а) варіація одного фактору дуже велика;
- б) різнонаправлений вплив факторів на результат.

Коефіцієнт β_j може також інтерпретуватися як показник прямого (безпосереднього) впливу j -ого фактору (x_j) на результат (y). У множинній регресії j -ий фактор виявляє не тільки прямий, але й опосередкований вплив на результат (тобто вплив через інші фактори моделі).

Непрямий вплив вимірюється величиною: $\sum \beta_i r_{x_j, x_i}$, де m - число факторів у моделі. Повний вплив j -ого фактору на результат рівне сумі

прямого й непрямого впливів вимірює коефіцієнт лінійної парної кореляції даного фактору й результату - $r_{x_j, y}$.

Так для нашого прикладу безпосередній вплив фактору x_1 на результат Y у рівнянні регресії вимірюється β_j і становить 0.43865889782842; непрямий (опосередкований) вплив даного фактору на результат визначається як:

$$r_{x_1 x_2} \beta_2 = 0.967525478136 * 0.24554383055896 = 0.2376.$$

3.6.11. Частинні коефіцієнти кореляції.

Коефіцієнт частинної кореляції відрізняється від простого коефіцієнта лінійної парної кореляції тим, що він вимірює парну кореляцію відповідних ознак (y і x_i) за умови, що вплив на них інших факторів (x_j) усунуте.

$$r_{yx1/x2} = \frac{r_{yx1} - r_{yx2} \cdot r_{x1 x2}}{\sqrt{(1-r_{yx2}^2)(1-r_{x1 x2}^2)}} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.94$$

Тіснота зв'язку сильна.

$$r_{yx1/x3} = \frac{r_{yx1} - r_{yx3} \cdot r_{x1 x3}}{\sqrt{(1-r_{yx3}^2)(1-r_{x1 x3}^2)}} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.92$$

Тіснота зв'язку сильна.

$$r_{yx2/x1} = \frac{r_{yx2} - r_{yx1} \cdot r_{x2 x1}}{\sqrt{(1-r_{yx1}^2)(1-r_{x2 x1}^2)}} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.95$$

Тіснота зв'язку сильна.

$$r_{yx2/x3} = \frac{r_{yx2} - r_{yx3} \cdot r_{x2 x3}}{\sqrt{(1-r_{yx3}^2)(1-r_{x2 x3}^2)}} = \frac{0.99 - 0.99 \cdot 1}{\sqrt{(1-0.99^2)(1-1^2)}} = 0.0939$$

Тіснота зв'язку низька.

$$r_{yx3/x1} = \frac{r_{yx3} - r_{yx1} \cdot r_{x3 x1}}{\sqrt{(1-r_{yx1}^2)(1-r_{x3 x1}^2)}} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.95$$

Тіснота зв'язку сильна.

$$r_{yx3/x2} = \frac{r_{yx3} - r_{yx2} \cdot r_{x3 x2}}{\sqrt{(1-r_{yx2}^2)(1-r_{x3 x2}^2)}} = \frac{0.99 - 0.99 \cdot 1}{\sqrt{(1-0.99^2)(1-1^2)}} = 0.62$$

Тіснота зв'язку помірна.

$$r_{x1 x2/y} = \frac{r_{x1 x2} - r_{x1 y} \cdot r_{x2 y}}{\sqrt{(1-r_{x1 y}^2)(1-r_{x2 y}^2)}} = \frac{0.97 - 0.99 \cdot 0.99}{\sqrt{(1-0.99^2)(1-0.99^2)}} = -0.79$$

Тіснота зв'язку сильна

$$r_{x1 x2/x3} = \frac{r_{x1 x2} - r_{x1 x3} \cdot r_{x2 x3}}{\sqrt{(1-r_{x1 x3}^2)(1-r_{x2 x3}^2)}} = \frac{0.97 - 0.97 \cdot 1}{\sqrt{(1-0.97^2)(1-1^2)}} = -0.13$$

Тіснота зв'язку низька.

$$r_{x1 x3/y} = \frac{r_{x1 x3} - r_{x1 y} \cdot r_{x3 y}}{\sqrt{(1-r_{x1 y}^2)(1-r_{x3 y}^2)}} = \frac{0.97 - 0.99 \cdot 0.99}{\sqrt{(1-0.99^2)(1-0.99^2)}} = -0.75$$

Тіснота зв'язку сильна.

$$r_{x1 x3/x2} = \frac{r_{x1 x3} - r_{x1 x2} \cdot r_{x3 x2}}{\sqrt{(1-r_{x1 x2}^2)(1-r_{x3 x2}^2)}} = \frac{0.97 - 0.97 \cdot 1}{\sqrt{(1-0.97^2)(1-1^2)}} = 0.48$$

Тіснота зв'язку не сильна.

$$r_{x2 x3/y} = \frac{r_{x2 x3} - r_{x2 y} \cdot r_{x3 y}}{\sqrt{(1-r_{x2 y}^2)(1-r_{x3 y}^2)}} = \frac{1 - 0.99 \cdot 0.99}{\sqrt{(1-0.99^2)(1-0.99^2)}} = 0.72$$

Тіснота зв'язку сильна.

$$r_{x2 x3/x1} = \frac{r_{x2 x3} - r_{x2 x1} \cdot r_{x3 x1}}{\sqrt{(1-r_{x2 x1}^2)(1-r_{x3 x1}^2)}} = \frac{1 - 0.97 \cdot 0.97}{\sqrt{(1-0.97^2)(1-0.97^2)}} = 0.93$$

3.6.12 Індекс множинної кореляції (множинний коефіцієнт кореляції).

Тісноту спільного впливу факторів на результат оцінює індекс множинної кореляції.

На відмінну від парного коефіцієнта кореляції, який може набувати від'ємних значень, він приймає значення від 0 до 1. Тому R не може бути використаний для інтерпретації напрямку зв'язку. Чим щільніше фактичні значення y_i розташовуються щодо лінії регресії, тим менша залишкова дисперсія й більша величина $R_{y(x1, \dots, xm)}$.

Таким чином, при значенні R близькому до 1, рівняння регресії краще описує фактичні дані й фактори сильніше впливають на результат. При значенні R близькому до 0 рівняння регресії погано описує фактичні дані й фактори слабо впливають на результат.

$$r_{yx1/x2} = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{\sqrt{(1-r_{yx2}^2)(1-r_{x1x2}^2)}} \\ r_{yx1/x2} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.94$$

Тіснота зв'язку сильна.

$$r_{yx1/x3} = \frac{r_{yx1} - r_{yx3} \cdot r_{x1x3}}{\sqrt{(1-r_{yx3}^2)(1-r_{x1x3}^2)}} \\ r_{yx1/x3} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.92$$

Тіснота зв'язку сильна.

$$r_{yx2/x1} = \frac{r_{yx2} - r_{yx1} \cdot r_{x2x1}}{\sqrt{(1-r_{yx1}^2)(1-r_{x2x1}^2)}} \\ r_{yx2/x1} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.95$$

Тіснота зв'язку сильна.

$$r_{yx2/x3} = \frac{r_{yx2} - r_{yx3} \cdot r_{x2x3}}{\sqrt{(1-r_{yx3}^2)(1-r_{x2x3}^2)}} \\ r_{yx2/x3} = \frac{0.99 - 0.99 \cdot 1}{\sqrt{(1-0.99^2)(1-1^2)}} = 0.0939$$

Тіснота зв'язку низька.

$$r_{yx3/x1} = \frac{r_{yx3} - r_{yx1} \cdot r_{x3x1}}{\sqrt{(1-r_{yx1}^2)(1-r_{x3x1}^2)}} \\ r_{yx3/x1} = \frac{0.99 - 0.99 \cdot 0.97}{\sqrt{(1-0.99^2)(1-0.97^2)}} = 0.95$$

Тіснота зв'язку сильна.

$$r_{yx3/x2} = \frac{r_{yx3} - r_{yx2} \cdot r_{x3x2}}{\sqrt{(1-r_{yx2}^2)(1-r_{x3x2}^2)}}$$

Тіснота зв'язку помірна.

$$r_{x1x2/y} = \frac{r_{x1x2} - r_{x1y} \cdot r_{x2y}}{\sqrt{(1-r_{x1y}^2)(1-r_{x2y}^2)}} \\ r_{x1x2/y} = \frac{0.97 - 0.99 \cdot 0.99}{\sqrt{(1-0.99^2)(1-0.99^2)}} = -0.79$$

Тіснота зв'язку сильна

$$r_{x1x2/x3} = \frac{r_{x1x2} - r_{x1x3} \cdot r_{x2x3}}{\sqrt{(1-r_{x1x3}^2)(1-r_{x2x3}^2)}} \\ r_{x1x2/x3} = \frac{0.97 - 0.97 \cdot 1}{\sqrt{(1-0.97^2)(1-1^2)}} = -0.13$$

Тіснота зв'язку низька.

$$r_{x1x3/y} = \frac{r_{x1x3} - r_{x1y} \cdot r_{x3y}}{\sqrt{(1-r_{x1y}^2)(1-r_{x3y}^2)}} \\ r_{x1x3/y} = \frac{0.97 - 0.99 \cdot 0.99}{\sqrt{(1-0.99^2)(1-0.99^2)}} = -0.75$$

Тіснота зв'язку сильна.

$$r_{x1x3/x2} = \frac{r_{x1x3} - r_{x1x2} \cdot r_{x3x2}}{\sqrt{(1-r_{x1x2}^2)(1-r_{x3x2}^2)}} \\ r_{x1x3/x2} = \frac{0.97 - 0.97 \cdot 1}{\sqrt{(1-0.97^2)(1-1^2)}} = 0.48$$

Тіснота зв'язку не сильна.

$$r_{x2x3/y} = \frac{r_{x2x3} - r_{x2y} \cdot r_{x3y}}{\sqrt{(1-r_{x2y}^2)(1-r_{x3y}^2)}} \\ r_{x2x3/y} = \frac{1 - 0.99 \cdot 0.99}{\sqrt{(1-0.99^2)(1-0.99^2)}} = 0.72$$

Тіснота зв'язку сильна.

$$r_{x2x3/x1} = \frac{r_{x2x3} - r_{x2x1} \cdot r_{x3x1}}{\sqrt{(1-r_{x2x1}^2)(1-r_{x3x1}^2)}} \\ r_{x2x3/x1} = \frac{1 - 0.97 \cdot 0.97}{\sqrt{(1-0.97^2)(1-0.97^2)}} = 0.93$$

Тісноту спільного впливу факторів на результат оцінює індекс множинної кореляції.

На відмінну від парного коефіцієнта кореляції, який може набувати від'ємних значень, він приймає значення від 0 до 1. Тому R не може бути використаний для інтерпретації напрямку зв'язку. Чим щільніше фактичні значення y_i розташовуються щодо лінії регресії, тим менша залишкова дисперсія й більша величина $R_{y(x_1, \dots, x_m)}$.

Таким чином, при значенні R близькому до 1, рівняння регресії краще описує фактичні дані й фактори сильніше впливають на результат. При значенні R близькому до 0 рівняння регресії погано описує фактичні дані й фактори слабо впливають на результат.

$$R = \sqrt{1 - \frac{s_e^2}{\sum (y_i - \bar{y})^2}} = \sqrt{1 - \frac{807.89}{655608.74}} = 1$$

Зв'язок між ознакою Y та факторами X сильна.

Значущість коефіцієнта кореляції.

$$T_{набл} = R \frac{\sqrt{n-m-1}}{\sqrt{1-R^2}} = 1 \frac{\sqrt{14}}{\sqrt{1-1}} = 107.95$$

За таблицею розподілу Стюдента знаходимо $T_{табл}$

$$T_{крит}(n-m-1; \alpha/2) = (14; 0.025) = 2.145$$

Оскільки $T_{набл} > T_{крит}$, то відхиляємо гіпотезу про рівність 0 коефіцієнта кореляції. Інакше кажучи, коефіцієнт кореляції статистично – значущий.

Інтервальна оцінка для коефіцієнта кореляції (довірчий інтервал).

$$T_{набл} = R \frac{\sqrt{n-m-1}}{\sqrt{1-R^2}} = 1 \frac{\sqrt{14}}{\sqrt{1-1}} = 107.95$$

Довірчий інтервал для коефіцієнта кореляції $t(1; 1)$.

3.6.13 Коефіцієнт детермінації.

Коефіцієнт детермінації дуже високий, це свідчить, що зміни x суттєво впливають на змінну y . Іншими словами - точність добору рівняння регресії – висока.

3.6.14. Оцінка значення результативної ознаки при заданих значеннях факторів.

$$Y(225,95,30) = -34.12 + 0.76 * 225 + 0.56 * 95 + 3.13 * 30 = 284.7$$

Довірчі інтервали з імовірністю 0.95 для середнього значення результативної ознаки $M(Y)$.

$$S^2 = X_0^T (X^T X)^{-1} X_0,$$

де

$$X_0^T = [1; 225; 95; 30]$$

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

1.35	-0.00428	0.0281	-0.0991
-0.00428	9.3E-5	3.9E-5	-0.000672
0.0281	3.9E-5	0.000948	-0.00419
-0.0991	-0.000672	-0.00419	0.0214

$$X_0 =$$

1
225
95
30

Множимо матриці, знаходимо $S^2 = 0.0576$

$$S_y = S \sqrt{X_0 (X^T X)^{-1} X_0} = 7.6 \sqrt{0.0576} = 1.82$$

$$(Y - t * S_y; Y + t * S_y) =$$

$$(284.7 - 2.145 * 1.82; 284.7 + 2.145 * 1.82) = (280.8; 288.6)$$

З імовірністю 0.95 значення Y при X_{0i} перебуває в зазначених межах.

Довірчі інтервали з імовірністю 0.95 для індивідуального значення результативної ознаки.

$$S_y = S \sqrt{1 + X_0 (X^T X)^{-1} X_0} = 7.6 \sqrt{1 + 0.0576} = 7.81$$

$$(284.7 - 2.145 * 7.81; 284.7 + 2.145 * 7.81)$$

$$(267.95; 301.45)$$

З імовірністю 0.95 індивідуальне значення Y при X_{0i} перебуває в зазначених межах.

3.6.15 Перевірка гіпотез щодо коефіцієнтів рівняння регресії (перевірка значущості параметрів множинного рівняння регресії).

$\nu = n - m - 1$ називається числом ступенів вільності. Уважається, що при оцінюванні множинної лінійної регресії для забезпечення статистичної надійності потрібно, щоб число спостережень, принаймні, в 3 рази перевершувало число оцінюваних параметрів.

1) t -статистика

$$T_{\text{табл}}(n-m-1; \alpha) = (14; 0.025) = 2.145,$$

$$t_i = \frac{b_i}{S_{b_i}}$$

Знаходимо стандартну помилку коефіцієнта регресії b_0 :

$$S_{b_0} = \sqrt{10.24} = 3.2,$$

$$t_0 = \frac{-34.12}{3.2} = 10.66 > 2.145$$

Статистична значущість коефіцієнта регресії b_0 підтверджується.
Знаходимо стандартну помилку коефіцієнта регресії b_1 :

$$S_{b_1} = \sqrt{0.000707} = 0.0266,$$
$$t_1 = \frac{0.76}{0.0266} = 28.64 > 2.145$$

Статистична значущість коефіцієнта регресії b_1 підтверджується.
Знаходимо стандартну помилку коефіцієнта регресії b_2 :

$$S_{b_2} = \sqrt{0.0072} = 0.0848,$$
$$t_2 = \frac{0.56}{0.0848} = 6.65 > 2.145$$

Статистична значущість коефіцієнта регресії b_2 підтверджується.
Знаходимо стандартну помилку коефіцієнта регресії b_3 :

$$S_{b_3} = \sqrt{0.16} = 0.4,$$

$$t_3 = \frac{3.13}{0.4} = 7.76 > 2.145.$$

Статистична значущість коефіцієнта регресії b_3 підтверджується.

Довірчий інтервал для коефіцієнтів рівняння регресії.

Довірчі інтервали коефіцієнтів регресії з надійністю 95% будуть наступними:

$$\begin{aligned} & (b_i - t_i S_{b_i}; b_i + t_i S_{b_i}), \\ & b_0: (-34.12 - 2.145 \cdot 3.2; -34.12 + 2.145 \cdot 3.2) = (-40.98; -27.25), \\ & b_1: (0.76 - 2.145 \cdot 0.0266; 0.76 + 2.145 \cdot 0.0266) = (0.7; 0.82), \\ & b_2: (0.56 - 2.145 \cdot 0.0848; 0.56 + 2.145 \cdot 0.0848) = (0.38; 0.75), \\ & b_3: (3.13 - 2.145 \cdot 0.4; 3.13 + 2.145 \cdot 0.4) = (2.26; 3.99). \end{aligned}$$

3.6.16 Перевірка загальної якості рівняння множинної регресії.

Оцінка значущості рівняння множинної регресії здійснюється шляхом перевірки гіпотези про рівність нулю коефіцієнт детермінації розрахованого за даними генеральної сукупності: R^2 або $b_1 = b_2 = \dots = b_m = 0$ (гіпотеза про незначущість рівняння регресії, розрахованого за даними генеральної сукупності).

Для її перевірки використовують F -Критерій Фішера.

При цьому обчислюють фактичне (спостережуване) значення F -критерію, через коефіцієнт детермінації R^2 , розрахований за даними конкретного спостереження.

За таблицями розподілу Фішера-Снедекора знаходять критичне значення F -критерію ($F_{кр}$). Для цього задаються рівнем значущості α (звичайно його беруть рівним 0,05) і двома числами ступенів волі $k_1 = m$ і $k_2 = n - m - 1$.

2) F -статистика. Критерій Фішера

$$R^2 = 1 - \frac{s_e^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{807.89}{655608.74} = 1$$

Чим ближче цей коефіцієнт до одиниці, тим більше рівняння регресії пояснює поведінку Y .

Більш об'єктивною оцінкою є скорегований коефіцієнт детермінації:

Введення в модель нових пояснюючих змінних здійснюється доти, доки росте скорегований коефіцієнт детермінації.

Перевіримо гіпотезу про загальну значущість - гіпотезу про одночасну рівність нулю всіх коефіцієнтів регресії при пояснюючих змінних:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0.$$

Перевірка цієї гіпотези здійснюється за допомогою F -статистики розподілу Фішера.

Якщо $F < F_{kp} = F_{\alpha; n-m-1}$, то немає підстав для відхилення гіпотези H_0 .

$$F = \frac{R^2}{1-R^2} \frac{(n-m-1)}{m} = \frac{1}{1-1} \frac{18-3-1}{3} = 3884.22$$

Табличне значення при ступенях волі $k_1 = 3$ і $k_2 = n-m-1 = 18 - 3 - 1 = 14$, $F_{kp}(3;14) = 3.34$.

Оскільки фактичне значення $F > F_{kp}$, те коефіцієнт детермінації статистично значущий і рівняння регресії статистично надійне.

Оцінка значущості додаткового включення фактору (частинний F -критерій).

Необхідність такої оцінки пов'язана з тим, що не кожний фактор, що ввійшов у модель, може суттєво збільшити частку поясненої варіації результативної ознаки. Це може бути пов'язане з послідовністю факторів, що вводяться (тому що існує кореляція між самими факторами). Заходом оцінки значущості поліпшення якості моделі, після включення в неї фактору x_j , служить частинний F -критерій F_{xj} :

$$F_{xj} = \frac{R^2 - R^2(x_1, x_n)}{1-R^2} (n - m - 1)$$

де m – число оцінюваних параметрів.

У чисельнику - приріст частки варіації в за рахунок додатково включеного в модель фактору x_j .

Якщо спостережуване значення F_{xj} більше F_{kp} , то додаткове введення фактору x_j у модель статистично виправдане.

Частинний F -критерій оцінює значущість коефіцієнтів «чистої» регресії (b_j). Існує взаємозв'язок між часткам F -критерієм - F_{xj} і t -критерієм, використовуваним для оцінки значущості коефіцієнта регресії при j -м факторі:

$$t(b_j=0) = \sqrt{F_{xj}}$$

3.6.17 Перевірка на наявність гетероскедастичності

1) Методом графічного аналізу залишків.

У цьому випадку по осі абсцис відкладаються значення пояснюючої змінної X_i , а по осі ординат квадрати відхилення u_i^2 .

Якщо є певний зв'язок між відхиленнями, то гетероскедастичність має місце. Відсутність залежності скоріш за все буде свідчити про відсутність гетероскедастичності.

2) За допомогою тесту рангової кореляції Спірмена.

Задаються два параметри: x_i , u .

Y	$y(x)$	$u = y - y(x)$	u^2
64.54	81.52	-16.98	288.29
87.52	83.06	4.46	19.89
94.58	96.29	-1.71	2.91
107.62	106.44	1.18	1.38
125.05	121.34	3.71	13.74
142.28	146.45	-4.17	17.41
156.55	155.26	1.29	1.66
179.28	176.46	2.82	7.94
209.3	203.27	6.03	36.36
245.52	245.86	-0.34	0.12
279.28	280.3	-1.02	1.04
329.3	317.91	11.39	129.64
385.52	375.02	10.5	110.35
430.05	437.64	-7.59	57.68
484.87	485.5	-0.63	0.4
522.03	526.14	-4.11	16.91
570.39	579.53	-9.14	83.45
741.82	737.49	4.33	18.72

Покажемо, наприклад, для x_1 .

X_i	u	ранг x_i , dx	ранг u , dy	$(dx - dy)^2$
95.98	-16.98	2	1	1
90.93	4.46	1	15	196
107.29	-1.71	3	6	9
116.18	1.18	4	10	36
128.9	3.71	5	13	64
147.09	-4.17	6	4	4
151.94	1.29	7	11	16
165.06	2.82	8	12	16
183.09	6.03	9	16	49
202.93	-0.34	10	9	1
228.06	-1.02	11	7	16
255.09	11.39	12	18	36
289.93	10.5	13	17	16
318.39	-7.59	14	3	121
350.9	-0.63	15	8	49
379.87	-4.11	16	5	121
404.32	-9.14	17	2	225
436.73	4.33	18	14	16
				992

$$T_{kp} = t(\alpha, k) \sqrt{\frac{1-p^2}{n-2}}.$$

3.6.18 Перевірка на наявність автокореляції залишків.

Важливою передумовою побудови якісної регресійної моделі по МНК є незалежність значень випадкових відхилень від значень відхилень в усі інших спостереженнях. Це гарантує відсутність корельованості між будь-якими відхиленнями й, зокрема, між сусідніми відхиленнями.

Автокореляція (послідовна кореляція) визначається як кореляція між спостережуваними показниками, упорядкованими за часом (часові ряди) або в просторі (перехресні ряди). Автокореляція залишків (відхилень) звичайно зустрічається в регресійному аналізі при використанні даних часових рядів і дуже рідко при використанні перехресних даних.

В економічних задачах значно частіше зустрічається **позитивна автокореляція**, ніж **негативна автокореляція**. У більшості випадків позитивна автокореляція викликається спрямованим постійним впливом деяких неврахованих у моделі факторів.

Негативна автокореляція фактично означає, що за позитивним відхиленням впливає негативне й навпаки. Така ситуація може мати місце, якщо ту ж залежність між попитом на прохолодні напої й доходами розглядати за сезонними даними (зима-літо).

Серед **основних причин, що викликають автокореляцію**, можна виділити наступні:

1. Помилки специфікації. Неврахування в моделі який-небудь важливої пояснюючої змінної або неправильний вибір форми залежності звичайно приводять до системних відхилень спостережень від лінії регресії, що може обумовити автокореляцію.
 2. Інерція. Багато економічних показників (інфляція, безробіття, ВВП і т.д.) мають певну циклічність, пов'язану з хвилеподібністю ділової активності. Тому зміна показників відбувається не миттєво, а має певну інертність.
 3. Ефект павутини. У багатьох виробничі й інших сферах економічні показники реагують на зміну економічних умов із запізнюванням (часовим лагом).
 4. Згладжування даних. Найчастіше дані деякого тривалому часового періоду отримуються усередненням даних по його інтервалах. Це може привести до певного згладжування коливань, які були усередині розглянутого періоду, що у свою чергу може служити причиною автокореляції.
- Наслідки автокореляції схожі з наслідками **гетероскедастичності**: висновки за t - і F -статистикам, що визначають значущість коефіцієнта регресії й коефіцієнта детермінації, можливо, будуть невірними.

3.6.19 Виявлення автокореляції

1. Графічний метод.

Є ряд варіантів графічного визначення автокореляції. Один з них **погоджує відхилення u_i з моментами їх одержання i** . При цьому по осі абсцис відкладають або час одержання статистичних даних, або порядковий номер спостереження, а по осі ординат – відхилення u_i (або оцінки відхилень).

Природно припустити, що якщо є певний зв'язок між відхиленнями, то автокореляція має місце. Відсутність залежності швидше за все буде свідчити про відсутність автокореляції.

Автокореляція стає більш наочною, якщо побудувати графік залежності u_i від u_{i-1} .

2. Коефіцієнт автокореляції.

Якщо коефіцієнт автокореляції $re_i < 0.5$, то є підстави затверджувати, що автокореляція відсутня.

3. Критерій Дарбіна-Уотсона.

Цей критерій є найбільш відомим для виявлення автокореляції.

При статистичному аналізі рівняння регресії на початковому етапі часто перевіряють виконання однієї передумови: умови статистичної незалежності відхилень між собою. При цьому перевіряється некорельованість сусідніх величин e_i .

y	$y(x)$	$ei = y - y(x)$	e^2	$(ei - ei-1)^2$
64.54	81.52	-16.98	288.29	0
87.52	83.06	4.46	19.89	459.63

94.58	96.29	-1.71	2.91	38.02
107.62	106.44	1.18	1.38	8.3
125.05	121.34	3.71	13.74	6.41
142.28	146.45	-4.17	17.41	62.08
156.55	155.26	1.29	1.66	29.83
179.28	176.46	2.82	7.94	2.34
209.3	203.27	6.03	36.36	10.32
245.52	245.86	-0.34	0.12	40.62
279.28	280.3	-1.02	1.04	0.46
329.3	317.91	11.39	129.64	153.89
385.52	375.02	10.5	110.35	0.78
430.05	437.64	-7.59	57.68	327.59
484.87	485.5	-0.63	0.4	48.45
522.03	526.14	-4.11	16.91	12.1
570.39	579.53	-9.14	83.45	25.23
741.82	737.49	4.33	18.72	181.23
			807.89	1407.26

Для аналізу коррельованості відхилень використовують *статистику Дарбіна-Уотсона*:

$$DW = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2},$$

Критичні значення d_1 і d_2 визначаються на основі спеціальних таблиць для необхідного рівня значущості α , числа спостережень $n = 18$ і кількості пояснюючих змінних $m=3$.

Автокореляція відсутня, якщо виконується наступна умова:

$$d_1 < DW \text{ і } d_2 < DW < 4 - d_2.$$

Не звертаючись до таблиць, можна користуватися приблизним правилом і вважати, що автокореляція залишків відсутня, якщо $1.5 < DW < 2.5$. Оскільки $1.5 < 1.74 < 2.5$, то автокореляція залишків відсутня.

Для більш надійного висновку доцільно звертатися до табличних значень.

За таблицею Дарбіна-Уотсона для $n=18$ і $k=3$ (рівень значущості 5%) знаходимо: $d_1 = 0.93$; $d_2 = 1.69$.

Оскільки $0.93 < 1.74$ і $1.69 < 1.74 < 4 - 1.69$, то автокореляція залишків відсутня.

3.6.20. Завдання на дослідження мультиколінеарності алгоритмом Феррара—Глобера

Дослідження мультиколінеарності на основі алгоритму Феррара-Глобера полягає в перевірці трьох видів критеріїв статистичної залежності:

- всього масиву незалежних змінних (χ^2 , хі-квадрат);
- кожної незалежної змінної зі всіма іншими незалежними змінними (F -критерій)
- кожної пари незалежних змінних (t -критерій).

В усіх трьох випадках виконується порівняння фактичних значень критеріїв з їх критичними значеннями. Ці порівняння дають право зробити конкретні висновки щодо наявності чи відсутності мультиколінеарності незалежних змінних.

Вихідні дані

На середньомісячну заробітну плату впливає ряд факторів. Виділимо серед них продуктивність праці, фондомісткість та коефіцієнт плинності робочої сили. Щоб побудувати економетричну модель заробітної плати від наведених чинників на основі методу найменших квадратів, треба переконатись, що продуктивність праці, фондомісткість та коефіцієнт плинності робочої сили як незалежні змінні — не мультиколінеарні.

Таблиця 3.1

Вихідні дані

№ п/п	Вантажооборот	Запаси	Трудомісткість
1	16.9	39.2	2.02
2	16.2	41	2.13
3	15.5	41.3	2.14
4	18.2	45.2	1.89
5	17.3	50.2	2.48
6	17.1	51.6	1.94
7	16.4	48	1.93
8	16.7	48.6	1.96
9	14.2	49.8	2.57
10	17.2	45	2.21

3.6.21 Порядок і рекомендації щодо виконання практикуму

Крок 1. Нормалізація змінних.

Позначимо вектори незалежних змінних — вантажообороту, запасів та трудомісткості — через відповідно X_1, X_2, X_3 . Елементи стандартизованих векторів розрахуємо за формулою:

$$X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{\sqrt{n\sigma_{X_k}^2}},$$

де n — кількість спостережень, $n = 10$;

m — число незалежних змінних, $m = 3$;

$\bar{X}_k$ — середня арифметична вектора X_k ;

$\sigma_{X_k}^2$ — дисперсія змінної X_k .

Середнє значення розрахуємо за формулою:

$$\bar{X}_k = \frac{\sum_{i=1}^n X_{ik}}{n};$$

$$\bar{X}_1 = 16,57;$$

$$\bar{X}_2 = 45,99;$$

$$\bar{X}_3 = 2,127.$$

Дисперсії кожної незалежної знаходимо за формулою:

$$\sigma_{X_k}^2 = \frac{\sum_{i=1}^n (X_{ik} - \bar{X}_k)^2}{n};$$

$$\sigma_{X_1}^2 = 1,092;$$

$$\sigma_{X_2}^2 = 16,917;$$

$$\sigma_{X_3}^2 = 0,05.$$

Провівши всі розрахунки по стандартизації незалежних змінних X_1, X_2, X_3 маємо:

Таблиця 3.2

	X_{i2}^*	X_{i3}^*
0,099858	-0,52205	-0,15174
-0,11196	-0,38365	0,004255
-0,32378	-0,36059	0,018436
0,493238	-0,06074	-0,33611
0,220898	0,323684	0,500616
0,160378	0,431323	-0,2652
-0,05144	0,154538	-0,27938
0,039338	0,200669	-0,23684
-0,71716	0,29293	0,628252
0,190638	-0,07612	0,117709

Матриця стандартизованих змінних матиме вигляд:

$$X^* = \begin{bmatrix} 0,099858 & -0,52205 & -0,15174 \\ -0,11196 & -0,38365 & 0,004255 \\ -0,32378 & -0,36059 & 0,018436 \\ 0,493238 & -0,06074 & -0,33611 \\ 0,220898 & 0,323684 & 0,500616 \\ 0,160378 & 0,431323 & -0,2652 \\ -0,05144 & 0,154538 & -0,27938 \\ 0,039338 & 0,200669 & -0,23684 \\ -0,71716 & 0,29293 & 0,628252 \\ 0,190638 & -0,07612 & 0,117709 \end{bmatrix}$$

Крок 2. Знаходження кореляційної матриці R :

$$R = X^{*T} \cdot X^*,$$

де X^{*T} — матриця транспонована до матриці X^* .

Ця матриця симетрична і має розмір 3 x 3. Для даної задачі:

$$R = \begin{bmatrix} 1 & -0,00635 & -0,54239 \\ -0,00635 & 1 & 0,223382 \\ -0,54239 & 0,223382 & 1 \end{bmatrix}$$

Кожен елемент цієї матриці характеризує тісноту зв'язку однієї незалежної змінної з іншою. Оскільки діагональні елементи характеризують

тісноту зв'язку кожної незалежної змінної з цією самою змінною, то вони дорівнюють одиниці.

Інші елементи матриці R трактуються так:

$$r_{X_1X_2} = -0,006,$$

$$r_{X_1X_3} = -0,54239,$$

$$r_{X_2X_3} = 0,223382 ,$$

тобто вони є парними коефіцієнтами кореляції незалежних змінних. На основі цих коефіцієнтів можна зробити висновок, що між змінними X_1, X_2, X_3 існує зв'язок. Але чи можна стверджувати, що цей зв'язок є явищем мультиколінеарності і він негативно впливатиме на оцінку регресійної моделі?

Щоб відповісти на це запитання, треба продовжити розв'язання на основі алгоритму Феррара—Глобера і в результаті знайти статистичні критерії оцінки мультиколінеарності.

Крок 3. Знайдемо детермінант кореляційної матриці R і критерій χ^2 :

$$\chi^2 = - \left[n - 1 - \frac{1}{6}(2m + 5) \right] \cdot \ln |R|;$$

$$|R| = 0,657;$$

$$\chi^2 = 3,006;$$

При ступені вільності

і рівні значущості $\alpha = 0,05$ $\chi^2_{\text{табл}} = 7,815$. Приймаючи $\chi^2_{\text{факт}} < \chi^2_{\text{табл}}$,

можемо зробити висновок, що в масиві змінних **мультиколінеарності не існує**.

Крок 4. Знайдемо матрицю C , обернену до матриці R :

$$C = \begin{bmatrix} 1,445212 & -0,17464 & 0,822879 \\ -0,17464 & 1,073623 & -0,33455 \\ 0,822879 & -0,33455 & 1,521054 \end{bmatrix}$$

Крок 5. Використовуючи діагональні елементи матриці C , розрахуємо

F -критерії:

$$F_1 = 1,56;$$

$$F_2 = 0,26;$$

$$F_3 = 1,82;$$

При рівні значущості $\alpha = 0,05$ і ступенях свободи $\gamma_1 = 7$ і $\gamma_2 = 2$ критичне (табличне) значення критерію $F_{\text{табл}} = 4,74$.

Через те, що

$$F_1 < F_{\text{табл}},$$

$$F_2 < F_{\text{табл}},$$

$$F_3 < F_{\text{табл}},$$

то жодна із незалежних змінних не мультиколінеарна з двома іншими.

Крок 6. Щоб визначити наявність попарної мультиколінеарності розрахуємо часткові коефіцієнти кореляції, використавши елементи матриці C :

$$r_{12.3} = \frac{-c_{12}}{\sqrt{c_{11} \cdot c_{22}}}, r_{12.3} = 0,14;$$

$$r_{13.2} = \frac{-c_{13}}{\sqrt{c_{11} \cdot c_{33}}}, r_{13.2} = -0,555;$$

$$r_{23.1} = \frac{-c_{23}}{\sqrt{c_{22} \cdot c_{33}}}, r_{23.1} = 0,262.$$

Часткові коефіцієнти кореляції характеризують тісноту зв'язку між двома змінними за умови, що третя не впливає на цей зв'язок.

Крок 7. Визначимо t -критерії на основі часткових коефіцієнтів кореляції:

$$t_{12} = \frac{r_{12 \cdot 3} \sqrt{n-m}}{\sqrt{1-r_{12 \cdot 3}^2}}, t_{12} = 0,375;$$

$$t_{13} = \frac{r_{13 \cdot 3} \sqrt{n-m}}{\sqrt{1-r_{13 \cdot 3}^2}}, t_{13} = -1,765;$$

$$t_{23} = 0,718.$$

Табличне значення t -критерію при $n - m = 7$ ступенях вільності і рівні значущості $\alpha = 0,05$ дорівнює 1,89. Всі числові значення t -критеріїв, знайдених для кожної пари змінних, менше за їх табличне значення. Звідси робимо висновок, що **всі пари незалежних змінних не є мультиколінеарними.**

3.6.22 Контрольні питання для підготовки до виконання практикуму

1. У чому суть методу найменших квадратів?
2. Які основні причини наявності в регресійній моделі випадкового відхилення?
3. Як розрахувати невідомі параметри лінійної моделі?
4. Пояснити сутність поняття "тіснота зв'язку".
5. Пояснити сутність поняття "значущість зв'язку".
6. За допомогою яких характеристик перевіряються тіснота зв'язку між змінними моделі?
7. За допомогою якої характеристики перевіряються значущість зв'язку між змінними моделі?
8. Що показує коефіцієнт детермінації і в яких межах він приймає значення?
9. Що показує коефіцієнт кореляції?
10. Запишіть формулу дисперсії залишків.
11. З якою ціллю розраховуються стандартні похибки оцінок параметрів?
12. За якими характеристиками вибирається табличне значення критерію Фішера?
13. Як визначити коефіцієнт детермінації у парній регресійній моделі?
14. Як визначити коефіцієнт кореляції у парній регресійній моделі?
15. У чому відмінність між точковим і інтервальним прогнозом?

Рекомендована література: [7, 8].

3.6.23 Варіанти вихідних даних завдання №3

Y – продуктивність праці, тис. грн/чол.; X_1 – втрати робочого часу, тис. людино-годин на рік; X_2 – коефіцієнт використання потужності, %.

Варіант 1		
Y	X_1	X_2
12,3	9,6	33,4
14,7	8,1	34,5
15,8	6,3	35,8
16,3	5,5	35,9
17,1	4,1	36,1
20,9	2,8	40,4
21,4	1,6	45,8
22,8	0,9	46,3
23,9	0,8	49,1
24,1	0,4	50,4
25,5	0,2	55,7
26,6	0,2	58,4
27,9	0,1	59,3
28,1	0,1	60,4
28,1	0,1	60,4

Варіант 2		
Y	X_1	X_2
16,1	12,2	23,8
16,3	10,0	24,9
17,8	8,1	25,5
17,6	8,0	25,6
18,5	7,6	27,8
18,9	7,4	29,3
20,1	6,1	30,4
24,3	5,3	31,7
25,8	2,7	32,8
26,7	2,6	33,9
27,8	1,5	35,5
30,1	0,5	36,9
32,3	0,4	37,7
33,9	0,1	38,1
33,9	0,1	38,1

Варіант 3		
Y	X_1	X_2
12,7	12,4	30,7
12,9	10,3	30,8
13,3	10,1	31,9
14,7	8,4	32,3
16,9	6,3	33,7
20,1	6,2	33,8
23,4	5,4	40
24,5	5,3	45,1
28,4	5,0	46,9
30,7	4,8	47,7
31,3	4,1	47,8
32,9	2,4	50,3
34,1	0,6	52,7
35,8	0,2	53,8
35,8	0,2	53,8

Варіант 4		
Y	X_1	X_2
15,8	12,3	30,7
19,4	12,0	31,3
20,3	6,1	32,9
27,8	6,2	33,4
29,3	5,1	33,5
24,4	4,3	36,7
31,3	3,8	37,9
32,8	3,0	38,4
32,9	2,7	39,4
34,7	2,5	40,6
41,3	2,0	42,9
44,9	1,9	47,8
45,6	1,6	48,1
46,1	1,5	49,3
46,1	1,5	49,3

Варіант 5		
Y	X_1	X_2
16,9	19,3	30,3
17,7	18,7	30,4
18,3	8,9	30,0
18,4	6,3	31,1
19,7	6,0	32,8
20,8	5,5	30,4
21,3	4,0	33,9
22,9	2,9	34,5
25,4	1,3	37,8
26,9	0,7	39,3
30,7	0,6	40,5
30,8	0,6	41,6
31,9	0,2	42,9
40,5	0,1	44,3
40,5	0,1	44,3

Варіант 6		
Y	X_1	X_2
10,9	12,4	30,1
11,3	10,1	33,4
12,8	8,4	33,9
13,8	6,0	34,8
14,9	7,9	36,3
15,6	5,4	38,9
18,1	6,1	39,4
19,4	3,1	39,8
19,0	3,0	45,0
21,3	2,1	46,4
22,8	2,0	47,3
24,1	1,5	48,8
26,7	0,5	49,1
28,3	0,4	50,4
28,3	0,4	50,4

Варіант 7		
Y	X_1	X_2
16,3	10,3	20,3
16,7	8,1	20,7
18,1	6,4	21,8
18,9	5,0	22,9
19,3	4,1	25,4
20,4	3,3	27,8
21,5	3,3	30,7
24,6	2,7	31,3
27,8	2,1	33,9
27,9	0,9	35,8
30,4	0,5	40,3
31,7	0,3	44,1
33,8	0,1	44,9
35,9	0,1	46,3
35,9	0,1	46,3

Варіант 8		
Y	X_1	X_2
16,7	15,1	20,7
17,3	14,0	21,8
18,4	12,4	30,4
19,5	9,3	31,7
20,3	8,6	38,9
21,4	5,1	40,3
25,5	3,3	44,1
26,9	3,2	44,7
27,8	4,0	45,8
30,1	2,1	46,3
32,3	1,9	47,9
33,9	0,8	50,3
34,5	0,6	51,8
36,6	0,1	52,1
36,6	0,1	52,1

Варіант 9		
Y	X_I	X_2
15,9	6,3	20,4
16,1	6,0	20,9
17,4	6,1	20,1
18,3	5,5	19,3
18,9	5,0	24,5
20,4	4,3	27,8
21,3	4,0	30,3
22,8	4,0	34,4
23,4	3,5	35,5
25,7	3,1	36,6
26,9	2,2	37,7
27,3	0,7	38,8
27,8	0,5	40,6
27,9	0,2	41,9
27,9	0,2	41,9

Варіант 10		
Y	X_I	X_2
12,8	6,3	40,5
12,9	5,1	4,9
13,1	5,0	41,7
14,4	4,2	42,8
15,8	3,3	43,9
16,3	2,1	4,0
17,9	2,0	45,6
18,1	1,9	47,8
19,3	1,7	48,1
20,4	1,1	50,3
21,8	0,8	51,9
21,9	0,8	58,8
24,3	0,6	60,1
25,7	0,2	61,3
25,7	0,2	61,3

Варіант 11		
Y	X_I	X_2
10,4	12,7	30,9
11,5	11,8	31,3
12,7	10,3	32,8
13,8	9,1	33,9
14,3	9,0	34,5
14,9	8,5	36,0
15,1	7,4	40,1
16,3	7,0	45,8
17,9	6,3	46,7
18,4	5,1	45,7
20,1	4,2	48,8
24,3	4,0	49,3
25,5	2,1	50,1
26,9	1,5	52,4
26,9	1,5	52,4

Варіант 12		
Y	X_I	X_2
10,7	9,1	45,8
11,8	8,6	46,3
13,8	8,6	47,4
14,9	8,0	47,8
15,1	7,4	50,1
16,3	7,1	51,3
17,1	6,9	51,9
18,7	6,2	53,8
18,9	5,3	59,1
19,3	4,2	60,3
20,1	3,7	67,8
21,7	2,8	69,1
22,8	1,5	69,4
23,1	0,5	69,7
23,1	0,5	69,7

Варіант 13		
Y	X_I	X_2
15,6	12,3	40,5
15,8	10,1	41,9
16,3	8,7	42,3
16,9	6,5	44,9
20,1	3,1	50,1
21,3	2,0	53,9
24,9	2,1	55,8
25,8	1,5	56,6
26,9	1,4	57,8
27,3	0,9	59,1
30,4	0,8	60,4
31,3	0,4	61,3
33,9	0,4	62,2
34,5	0,2	64,0
34,5	0,2	64,0

Варіант 14		
Y	X_I	X_2
12,3	12,3	31,7
12,5	10,1	32,8
13,7	8,8	33,0
13,9	7,3	33,1
14,7	6,9	35,4
15,1	5,1	36,9
16,3	4,0	40,1
17,4	4,1	45,5
18,1	3,2	46,9
18,9	3,1	47,7
19,3	2,4	48,1
20,1	0,9	49,3
21,7	0,4	50,4
22,8	0,2	51,7
22,8	0,2	51,7

Варіант 15		
Y	X_I	X_2
12,3	6,3	40,4
12,9	6,0	41,9
13,4	6,0	43,0
14,7	5,8	44,7
15,1	5,3	49,8
16,9	5,1	50,3
17,3	5,0	51,4
18,7	4,1	55,6
19,8	3,2	57,8
20,1	2,7	60,3
24,4	1,9	61,4
25,9	1,3	60,0
26,1	0,8	61,9
28,8	0,6	62,1
28,8	0,6	62,1

Варіант 16		
Y	X_I	X_2
10,3	6,7	40,9
10,9	6,5	41,7
11,7	6,0	42,8
12,8	5,9	43,9
13,9	4,1	45,8
14,7	4,0	45,0
15,1	3,8	49,0
16,9	2,7	50,1
21,9	1,9	52,3
18,7	0,8	52,4
20,3	0,6	55,0
24,5	0,6	57,0
26,6	0,4	58,1
27,8	0,4	59,3
27,8	0,4	59,3

Варіант 17		
Y	X_I	X_2
15,1	10,7	30,1
15,0	8,1	31,7
15,6	8,0	32,8
15,9	6,7	33,4
16,1	6,6	34,9
16,3	7,0	35,8
17,8	6,5	36,9
17,0	6,0	37,8
19,0	5,1	40,1
18,4	4,3	42,8
18,0	2,8	43,9
18,9	1,9	45,5
19,3	0,7	46,0
20,1	0,6	47,9
20,1	0,6	47,9

Варіант 18		
Y	X_I	X_2
12,7	10,0	30,9
12,8	9,1	31,6
12,0	8,7	30,7
13,9	8,8	31,8
14,1	8,6	33,9
15,8	7,1	34,5
16,1	6,3	35,7
17,3	5,1	40,1
18,9	4,2	42,3
18,9	2,8	44,9
20,5	0,9	46,8
24,1	0,8	50,4
25,8	0,7	51,3
26,3	0,8	52,8
26,3	0,8	52,8

Варіант 19		
Y	X_I	X_2
20,8	5,1	30,7
21,9	5,0	31,8
21,9	4,9	32,8
23,4	4,9	32,0
25,7	3,1	32,9
26,8	3,0	34,5
27,1	2,8	36,6
28,1	2,1	38,9
30,3	1,8	40,5
31,4	0,9	42,4
32,9	0,9	47,0
34,5	0,7	47,1
37,8	0,6	47,9
40,4	0,6	52,1
40,4	0,6	52,1

Варіант 20		
Y	X_I	X_2
6,9	10,1	30,9
10,3	9,4	30,9
12,5	6,5	31,3
12,4	5,1	32,8
13,7	4,0	34,6
16,8	4,2	35,8
19,3	3,1	36,9
20,1	2,8	40,1
21,7	2,0	42,3
25,8	1,9	44,1
30,1	1,8	45,8
32,9	0,8	46,1
34,5	0,7	46,4
36,6	0,6	50,3
36,6	0,6	50,3

Варіант 21		
Y	X_I	X_2
19,8	10,5	40,9
20,3	9,4	41,3
21,4	8,3	44,9
22,9	6,1	45,6
24,6	5,4	46,9
27,8	2,2	47,3
29,3	2,0	48,8
30,1	1,9	49,5
33,4	0,7	50,1
35,8	0,6	52,4
36,6	0,4	53,9
37,7	0,4	54,6
38,8	0,2	54,9
38,1	0,2	56,1
38,1	0,2	56,1

Варіант 22		
Y	X_I	X_2
5,6	7,9	30,1
5,6	7,8	33,0
5,8	7,6	31,7
5,9	6,1	32,8
6,7	6,0	33,4
7,9	5,8	33,9
8,3	5,5	33,9
8,7	5,4	34,1
10,1	4,1	35,7
15,4	4,0	36,8
16,3	3,1	37,7
17,8	3,7	40,1
19,1	3,2	40,3
20,3	2,5	40,3
20,3	2,5	40,3

Варіант 23		
Y	X_I	X_2
15,6	9,1	30,7
15,8	9,0	31,8
16,1	7,1	32,7
18,4	7,0	33,9
18,0	6,8	34,5
19,1	6,3	36,7
20,3	6,3	40,1
21,7	6,0	42,3
29,8	6,1	44,8
30,3	5,8	45,9
30,1	5,4	47,8
30,4	3,1	50,3
31,2	2,0	51,4
32,3	1,9	55,8
32,3	1,9	55,8

Варіант 24		
Y	X_I	X_2
10,7	6,1	20,7
10,0	6,0	20,1
10,9	7,0	27,8
11,5	6,3	27,9
11,9	5,8	30,1
11,0	5,7	35,5
13,0	6,0	36,1
13,7	5,4	36,7
14,1	5,1	37,8
14,8	5,0	40,1
15,1	4,5	40,0
16,3	3,0	40,5
17,7	2,1	41,9
18,1	1,7	43,7
18,1	1,7	43,7

Варіант 25		
Y	X_1	X_2
10,1	6,7	50,1
10,0	5,1	51,1
10,7	6,9	51,3
11,3	6,9	50,4
11,8	5,4	51,7
12,4	7,1	52,3
15,7	6,3	53,8
15,8	6,5	54,1
16,1	6,1	55,0
16,0	5,8	55,0
17,1	5,7	56,1
17,8	5,6	57,8
17,8	5,1	57,9
19,0	5,0	58,0
19,0	5,0	58,0

Варіант 26		
Y	X_1	X_2
12,9	3,3	40,4
12,0	3,0	45,6
12,7	3,1	47,8
13,4	3,1	47,0
14,7	2,8	48,1
15,1	2,5	49,4
15,8	2,3	49,0
16,3	2,1	49,5
17,4	2,0	50,1
18,1	1,5	50,0
19,3	1,4	51,3
20,7	1,2	54,0
21,9	1,0	54,3
23,4	0,9	55,0
23,4	0,9	55,0

Варіант 27		
Y	X_1	X_2
15,4	7,7	60,7
15,6	7,0	60,6
16,1	7,9	61,3
17,3	7,1	62,7
17,5	6,8	64,1
18,1	5,3	65,1
18,6	5,1	66,3
18,9	4,2	37,8
20,3	2,0	69,1
21,4	1,9	70,5
25,4	1,7	70,6
24,3	1,4	71,8
24,7	1,2	72,3
25,8	1,2	74,1
25,8	1,2	74,1

Варіант 28		
Y	X_1	X_2
19,1	7,3	48,7
19,3	7,0	47,6
19,0	7,1	49,1
19,4	6,9	49,8
20,5	6,5	50,3
21,6	5,5	51,7
22,8	5,7	52,8
23,7	5,0	53,4
24,1	4,1	54,1
25,6	3,1	55,0
27,1	3,3	56,1
28,3	2,1	57,3
29,1	2,1	58,4
30,4	1,9	58,1
30,4	1,9	58,1

Варіант 29		
Y	X_I	X_2
18.1	4.5	20.5
18.0	4.0	21.6
17.6	4.1	21.8
18.4	4.6	21.3
19.1	3.7	22.8
19.9	3.1	23
20.1	3.0	24.0
20.3	2.8	25.6
21.4	2.5	25.0
22.5	2.0	27.8
23.4	1.9	29.1
25.1	0.9	30.5
26.3	0.7	31.3

Варіант 30		
Y	X_I	X_2
13.1	8.3	55.8
13.4	8.0	56.1
13.0	8.7	57.1
12.8	8.1	60.3
14.5	7.5	61.4
14.8	6.5	62.5
15.1	6.0	70.4
15.4	5.9	71.7
15.9	5.4	73.7
16.3	5.1	73.7
17.4	4.3	75.8
18.1	2.1	76.7
19.0	2.0	77.9

Розділ 4. **ПОБУДОВА ТА ДОСЛІДЖЕННЯ** **НЕОКЛАСИЧНИХ ВИРОБНИЧИХ ФУНКЦІЙ**

4.1.Короткі теоретичні відомості

Виробнича функція – це економетрична модель, яка кількісно описує зв'язок основних результативних показників виробничо-господарської діяльності з факторами, що визначають ці показники.

4.2. Економічний аналіз неокласичних систем

Розглянемо мультиплікативну виробничу функцію у загальному вигляді:

$$Y = A \cdot K^{\alpha} \cdot L^{\beta}.$$

Коефіцієнти α та β називаються коефіцієнтами еластичності.

Економічний аналіз моделі проводиться на базі коефіцієнтів еластичності, а також сумісного їх впливу на $Y : (\alpha + \beta)$.

Якщо коефіцієнт еластичності має знак “–”, це свідчить про існування оберненого зв'язку.

Коефіцієнт еластичності α показує на скільки збільшиться у процентах виробнича функція Y , якщо фактор K збільшиться на 1% при незмінному факторі L .

Коефіцієнт еластичності β показує на скільки зменшиться у відсотках виробнича функція Y , якщо фактор L збільшиться на 1% при незмінному факторі K .

Якщо $(\alpha + \beta) > 1$, в цьому випадку можна говорити про зростаючу ефективність факторів, тобто темпи зростання Y будуть вище ніж темпи зростання виробничих ресурсів (K та L).

Якщо $(\alpha + \beta) < 1$, то Y зростає нижчими темпами, ніж темпи зростання ресурсів (K та L).

Якщо $(\alpha + \beta) = 1$, то темпи зростання Y дорівнює темпам зростання K і L (ефективність факторів постійна).

4.3. Мета та основні завдання практикуму №4

Побудувати виробничу функцію виду $Y = A \cdot K^{\alpha} \cdot L^{\beta}$

де Y – обсяг випущеної продукції, тис. грн.; K – вартість основних засобів, млн. грн.; L – чисельність працюючих, чол.

Знайти числові параметри цієї функції A, α, β .

Провести економічний аналіз впливу факторів K та L на Y , використовуючи як окремі коефіцієнти еластичності α і β так і об'єднані $(\alpha + \beta)$.

У таблиці наведено дані вашого варіанту про випуск продукції, затратах праці і витратах виробничих фондів за 12 років. Використовуючи ці дані, потрібно побудувати виробничу функцію Кобба-Дугласа та розрахувати характеристики:

1. Середню продуктивність праці;
2. Середню фондівдачу;
3. Граничну продуктивність праці;
4. Граничну фондівдачу;
5. Еластичність випуску продукції за витратами праці;
6. Еластичність випуску продукції по виробничих фондах;
7. Потребу в ресурсах праці;
8. Потребу в виробничих фондах;
9. Фондоозброєність праці;
10. Граничну норму заміщення витрат праці виробничими фондами; еластичність заміщення ресурсів;
11. Знайти прогноз випуску для заданих значень.

Мультиплікативна виробнича функція

$$\alpha_1 := 0.53 \quad \alpha_2 := 1 - 0.3$$

90

$$F(K, L) := 2 \cdot K^{\alpha_1} \cdot L^{\alpha_2}$$

Обсяг виробництва: $X(K, L) := F(K, L)$

Якщо коефіцієнти еластичності задовольняють умові

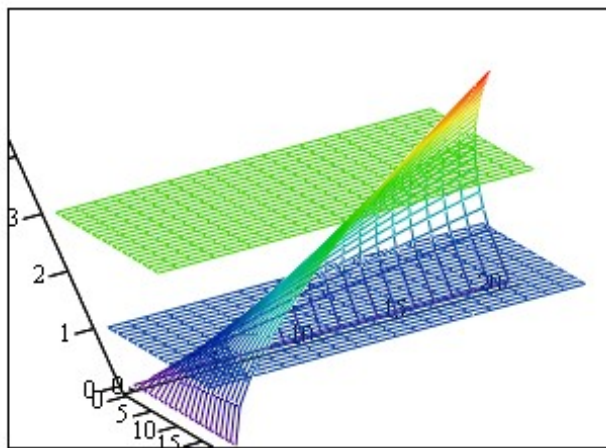
$$\alpha_1 + \alpha_2 > 1,$$

то темп росту основного (результуючого) показника більший за темп росту будь-якого з факторів !

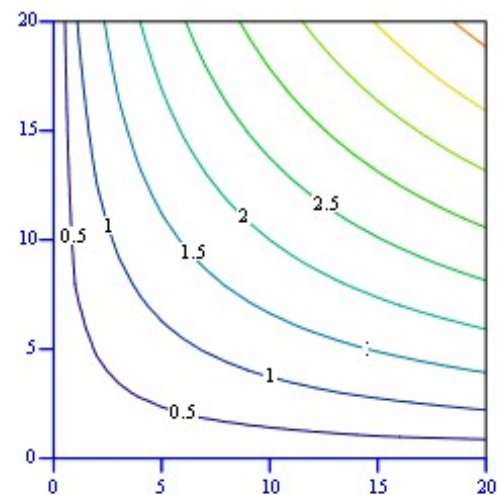
В нас ця властивість, очевидно, виконується, бо:

$$\alpha_1 = 0.53 \quad \alpha_2 = 0.7$$

$$i := 0..20 \quad j := 0..20 \quad k_i := i \cdot 0.1 \quad m_j := j \cdot 0.1 \quad z_{i,j} := F(k_i, m_j) \quad z1_{i,j} := 1 \quad z2_{i,j} := 3 \quad z3_{i,j} := 5$$



$z, z1, z2$



z

Граничний продукт основних виробничих фондів (капіталу):

$$F_k(K, L) := \frac{d}{dK} F(K, L) \quad F_k(K, L) \rightarrow \frac{1.06 \cdot L^{0.7}}{K^{0.47}}$$

$$F_k(K, L) = \alpha_1 \cdot \frac{X(K, L)}{K} \quad \text{Наприклад:} \quad F_k(7, 8) = 1.821 \quad \alpha_1 \cdot \frac{X(7, 8)}{7} = 1.821$$

Рис.4.1. Копія екрану з результатами роботи програми до практикуму №4.

4.4. Порядок і рекомендації щодо виконання практикуму

+

Граничний продукт праці: $\frac{d}{dL} F(K, L)$

$$Fl(K, L) := \frac{d}{dL} F(K, L) \quad Fl(K, L) \rightarrow \frac{1.4 \cdot K^{0.53}}{L^{0.3}}$$

$$\frac{d}{dL} F(K, L) = \alpha_2 \cdot \frac{X(K, L)}{L} \quad \text{Наприклад:} \quad Fl(3, 5) = 1.546 \quad \alpha_2 \cdot \frac{X(3, 5)}{5} = 1.546$$

$$\underline{\underline{R}}(K, L) := \frac{Fl(K, L)}{F(K, L)} \quad R(K, L) \rightarrow \frac{1.3207547169811320755 \cdot K^{1.0}}{L^{1.0}} \quad R(3, 5) = 0.792$$

$$\frac{\alpha_2}{\alpha_1} = 1.321 \quad +$$

Еластичності по ресурсах:

$$Ek(K, L) := \frac{\frac{d}{dK} X(K, L)}{\frac{X(K, L)}{K}} \quad Ek(K, L) \rightarrow 0.53$$

$$\alpha_1 = 0.53$$

$$El(K, L) := \frac{\frac{d}{dL} X(K, L)}{\frac{X(K, L)}{L}} \quad El(K, L) \rightarrow \frac{0.7}{L^{1.2621774483536188887e-29}}$$

$$\alpha_2 = 0.7$$

Сума еластичностей:

$$E(K, L) := Ek(K, L) + El(K, L)$$

Рис. 4.2. Копія екрану з результатами роботи програми до практикуму №4.

4.5.Індивідуальні вихідні дані варіантів практикуму до розділу

Варіант1			Варіант2			Варіант 3			Варіант4			Варіант5		
<i>Y</i>	<i>K</i>	<i>L</i>	<i>Y</i>	<i>K</i>	<i>L</i>	<i>Y</i>	<i>K</i>	<i>L</i>	<i>Y</i>	<i>K</i>	<i>L</i>	<i>Y</i>	<i>K</i>	<i>L</i>
58.9	107.2	49.4	30	30.5	29.2	38.6	40.2	32.9	58.3	105	48.1	47.8	109.9	37.3
46.7	104.2	43.5	41	57.8	36.5	49.4	107.7	46.4	43.5	59.4	40	47	104.2	44
33.5	35.4	33.2	33.2	35.1	32.7	35.4	38.3	36.7	33.8	35.1	34	43.5	59.7	40.2
41.9	53.5	37.8	41.9	53.5	37.8	30	30.5	29.2	41	52.7	37	41.3	58.3	37.3
27.3	27	27.5	56.2	85.1	49.1	38.6	40.5	35.4	28.6	29.2	27.5	33.5	35.4	33.2
54.5	79.7	46.2	55.4	80.5	47.3	51.6	77	44.6	53.7	75.3	44.8	27.3	27	27.5
30.2	30.8	29.7	43.5	59.4	40	31.1	30.8	30.2	30.2	30.8	29.7	38.6	40.2	33.8
28.4	28.9	26.7	27	27	27.3	27.8	28.9	27.3	56.2	85.1	49.1	50.8	71.6	41.3
38.6	40.2	33.8	38.1	40	32.9	58.9	108	49.7	37.5	39.4	32.7	38.6	40.5	35.4
61.3	90.5	50.2	48.1	113.7	37.3	62.4	90.5	50.5	61.3	90.5	50.2	41.3	47.5	37
40.8	47.5	37.3	44.6	59.9	40.5	51	72.4	41.9	40.2	46.2	35.9	58.9	107.2	49.4
41.3	58.3	37.3	46.4	104	43.5	41.6	58.9	37.5	40.5	58.3	36.7	49.1	63.5	42.1
51	71.8	41.6	38.6	40.5	35.4	41	47.8	36.5	38.6	40.5	35.4	61.3	90.5	50.2
49.7	63.7	41.6	49.4	63.5	41.9	51.6	65.6	44	48.6	61.8	41.6	30	30.5	29.2
48.3	112.6	37	61.3	89.9	50	56.2	85.1	49.1	44.6	59.9	40.5	27.5	28.4	27

4.6.Контрольні питання для підготовки до виконання практикуму

1. Неокласичні виробничі функції.
2. Мультиплікативна виробнича функції. Виробнича функція Кобба-Дугласа.
3. Умови неокласичності мультиплікативної виробничої функції.
4. Економетрична побудова неокласичної моделі.
5. Розрахунки граничних показників по побудованій економетричній моделі виробничої функції.
6. Аналіз економічної системи з неокласичною виробничою функцією.

Рекомендована література: [7,8].

Розділ 5. ОДНОЧАСНІ СТРУКТУРНІ РІВНЯНЯ

5.1. Короткі теоретичні відомості

Система одночасних структурних рівнянь в матричному вигляді має такий вигляд:

$$Y = AY + BX + u.$$

Якщо кожне рівняння системи розв'язати відносно Y , то одержимо приведену форму моделі, яка має вигляд:

$$Y = RX + v,$$

де залишки v є лінійною комбінацією залишків u .

Зв'язок між коефіцієнтами структурної і приведенної форми моделі визначиться:

$$\begin{aligned} R &= (E - A)^{-1} B, \\ \text{або } R &= -A^{-1} B, \\ \text{або } AR + B &= 0. \end{aligned}$$

Оцінка параметрів моделі на основі одночасних рівнянь методом 1МНК буде давати зміщення, яке буде дорівнювати:

$$\frac{(1 - a_1) \sigma^2 / \bar{m}_{zz}}{1 + \sigma^2 / \bar{m}_{zz}},$$

Де $\bar{m}_{zz}$ - момент другого порядку залежної змінної, який прямує до деякої константи.

Чисельна оцінка параметрів моделі на основі одночасних структурних рівнянь пов'язана з проблемою ідентифікації.

Необхідна умова ідентифікації системи — справедливості нерівності для кожного рівняння:

$$k_s - 1 \leq m - m_s,$$

де k_s — кількість ендогенних змінних, які входять в s -те рівняння структурної форми;

m — загальна кількість екзогенних змінних моделі;

m_s — кількість екзогенних змінних, які не входять в S -те рівняння структурної форми моделі.

Якщо записане вище співвідношення виконується як рівність, то відповідне рівняння є строго ідентифікованим, а коли — як нерівність, то відповідне рівняння є надідентифікованим.

Якщо в структурній формі моделі $Y = AY + BX + u$ матриця A є трикутною, а залишки характеризуються діагональною матрицею виду:

$$\Sigma = M(uu^T) = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix},$$

то така система рівнянь називається рекурсивною і для оцінки параметрів такої моделі можна застосувати 1МНК.

Якщо кожне рівняння моделі є строго ідентифікованим, то для оцінки параметрів моделі можна застосувати непрямий метод найменших квадратів (НМНК). Алгоритм цього методу складається з чотирьох кроків:

Крок 1. Перевіряється умова ідентифікованості для кожного рівняння. Якщо кожне рівняння точно ідентифіковане, то виконується перехід до кроку 2.

Крок 2. Перехід від структурної форми моделі до приведеної.

Крок 3. Оцінка параметрів кожного рівняння приведеної форми моделі 1МНК.

Крок 4. Розрахунок оцінок параметрів рівнянь структурної форми на основі співвідношення:

$$AR = -B,$$

де A і B — параметри структурних рівнянь, а R — матриця оцінок параметрів приведеної форми моделі.

Якщо рівняння структурної форми моделі надідентифіковані, то для оцінки параметрів моделі застосовується двокроковий метод найменших квадратів (2МНК). Система рівнянь для обчислення оцінок двокроковим методом найменших квадратів запишеться так:

$$\begin{pmatrix} Y_1^T X(X^T X)^{-1} X^T Y_1 & Y_1^T X_1 \\ X_1^T Y_1 & X_1^T X_1 \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} Y_1^T X(X^T X)^{-1} X^T Y \\ X_1^T Y \end{pmatrix},$$

де Y — вектор залежної або ендогенної змінної;

Y_1 — матриця поточних ендогенних змінних, які входять у праву частину рівняння;

X — матриця всіх пояснюючих або екзогенних змінних;

X_1 — матриця пояснюючих або екзогенних змінних даного рівняння;

$\hat{a}$ — вектор структурних параметрів, які відносяться до змінних матриці Y_1 ;

$\hat{b}$ — вектор структурних параметрів, які відносяться до змінних матриці X_1 .

Оператор оцінювання 2МНК:

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} Y_1^T X (X^T X)^{-1} X^T Y_1 & Y_1^T X_1 \\ X^T Y_1 & X^T X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_1^T X (X^T X)^{-1} X^T Y \\ X^T Y \end{pmatrix}$$

Дисперсія залишків для кожного рівняння:

$$\sigma_{u_s}^2 = \frac{1}{n-k-r+1} u_s^T u_s.$$

Матриця коваріацій параметрів кожного рівняння визначається на основі співвідношення:

$$asy \text{ var} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \sigma_{u_s}^2 \begin{pmatrix} Y_1^T X (X^T X)^{-1} X^T Y_1 & Y_1^T X_1 \\ X^T Y_1 & X^T X_1 \end{pmatrix}^{-1}$$

5.2. Алгоритм вибору методу оцінювання



Рис.5.1

5.2.1. Мета та основні завдання комп'ютерного практикуму

Визначіть параметри найпростішої мультиплікативної моделі споживання Кейнса для певного регіону на підставі статистики за 12 років:

t	$C(t)$	$Y(t)$	$I(t)$
1	58,8	7,3	9,22
2	67,4	9,56	13,82
3	68,9	11,1	15,02
4	80,1	12,04	17,08
5	70,45	13,34	18,94
6	84,35	13,26	20,36
7	77,25	15,4	21,56
8	81,4	13,98	22,2
9	73,35	16,86	27,56
10	77,95	15,88	30,36
11	77,65	18,98	28,14
12	82,35	17,18	31,46

Де $C(t)$ – споживання; $Y(t)$ – дохід; $I(t)$ – інвестиції (всі дані у млн. грн).

5.2.2. Порядок і рекомендації щодо виконання

Модель Кейнса є системою одночасних структурних рівнянь:

$$\begin{aligned}C(t) &= a_0 + a_1 Y(t) + u(t), \\ Y(t) &= C(t) + I(t),\end{aligned}$$

де $C(t)$ – споживання; $Y(t)$ – дохід; $I(t)$ – інвестиції; $u(t)$ – стохастична похибка.

Легко перевірити, що модель є строго ідентифікованою.

Робимо припущення, що споживання $C(t)$ добре описується законом лінійної парної регресії:

$$C(t) = a_0 + a_1 Y(t).$$

$$\bar{C}(t) = 952.3958, \bar{Y}(t) = 26.3694,$$

оцінюємо коефіцієнти:

$$a_1 = \left(\sum_{t=1}^n C(t)Y(t) - n\bar{C}(t)\bar{Y}(t) \right) / \left(\sum_{t=1}^n Y^2(t) - n\bar{Y}^2(t) \right) = 32.1427$$

$$a_0 = \bar{C}(t) - a_1\bar{Y}(t) = 104.7858,$$

Відповідь:

$$C(t) = 104.7858 + 32.1427Y(t) + u(t)$$

$$Y(t) = C(t) + I(t)$$

Вихідні дані варіантів індивідуальних завдань №5

Варіант 1			Варіант 2			Варіант 3		
59.71	7.41	49.43	68.3	7.53	50.22	77.2	7.65	51.04
68.44	9.71	74.1	79.54	9.86	75.28	91.03	10.02	76.51
69.96	11.27	80.53	82.7	11.45	81.81	95.88	11.64	83.15
81.34	12.23	91.58	95.24	12.42	93.03	109.62	12.62	94.55
71.54	13.55	101.55	86.65	13.76	103.16	102.28	13.99	104.85
85.65	13.46	109.16	100.91	13.68	110.9	116.68	13.9	112.71
78.44	15.64	115.6	95.82	15.89	117.43	113.79	16.15	119.35
82.66	14.2	119.03	98.62	14.42	120.92	115.12	14.66	122.9
74.48	17.12	147.77	93.33	17.39	150.11	112.81	17.68	152.57
79.15	16.13	162.78	97.05	16.38	165.37	115.55	16.65	168.07
78.85	19.27	150.87	99.99	19.58	153.27	121.84	19.9	155.78
83.62	17.45	168.68	102.95	17.72	171.36	122.93	18.01	174.16

Варіант 4			Варіант 5			Варіант 6		
86.4	7.78	51.9	95.94	7.92	52.79	105.84	8.06	53.73
102.91	10.19	77.79	115.23	10.37	79.13	128	10.55	80.53
109.52	11.83	84.55	123.65	12.04	86	138.3	12.25	87.53
124.51	12.84	96.14	139.94	13.06	97.8	155.93	13.29	99.53
118.45	14.22	106.61	135.2	14.47	108.45	152.57	14.72	110.37
133.01	14.14	114.6	149.92	14.38	116.58	167.46	14.63	118.64
132.39	16.42	121.36	151.66	16.7	123.45	171.63	17	125.64
132.2	14.9	124.96	149.9	15.16	127.12	168.25	15.43	129.37
132.98	17.97	155.13	153.86	18.28	157.81	175.51	18.61	160.6
134.7	16.93	170.89	154.53	17.22	173.84	175.09	17.53	176.92
144.45	20.23	158.4	167.87	20.58	161.13	192.15	20.95	163.98
143.61	18.32	177.08	165.03	18.63	180.14	187.24	18.96	183.33

Варіант 7			Варіант 8			Варіант 9		
116.11	8.2	54.7	126.77	8.36	55.72	137.86	8.51	56.78
141.26	10.74	82	155.03	10.94	83.52	169.34	11.15	85.11
153.5	12.47	89.12	169.29	12.7	90.77	185.7	12.95	92.5
172.53	13.53	101.34	189.77	13.78	103.22	207.69	14.04	105.19
170.6	14.99	112.37	189.31	15.27	114.46	208.77	15.56	116.64
185.67	14.9	120.8	204.57	15.18	123.04	224.22	15.47	125.39
192.36	17.3	127.92	213.88	17.63	130.3	236.25	17.96	132.78
187.29	15.71	131.71	207.06	16	134.16	227.62	16.31	136.72
197.98	18.95	163.52	221.31	19.3	166.56	245.56	19.67	169.73
196.43	17.84	180.13	218.59	18.18	183.48	241.62	18.52	186.97
217.34	21.33	166.96	243.5	21.72	170.06	270.69	22.14	173.3
210.29	19.31	186.66	234.22	19.66	190.13	259.09	20.04	193.75

Варіант 10			Варіант 11			Варіант 12		
149.4	8.68	57.89	161.4	8.85	59.04	173.91	9.03	60.25
184.23	11.37	86.77	199.72	11.59	88.5	215.87	11.83	90.31
202.78	13.2	94.31	220.55	13.46	96.19	239.06	13.74	98.15
226.33	14.32	107.24	245.73	14.6	109.38	265.94	14.9	111.61
229.01	15.86	118.92	250.07	16.18	121.29	272.01	16.51	123.76
244.66	15.77	127.83	265.94	16.08	130.38	288.11	16.41	133.04
259.52	18.31	135.37	283.74	18.68	138.07	308.97	19.06	140.89
249	16.62	139.39	271.25	16.96	142.17	294.43	17.3	145.07
270.78	20.05	173.04	297.03	20.45	176.49	324.37	20.87	180.09
265.58	18.88	190.62	290.51	19.26	194.42	316.48	19.65	198.39
298.97	22.57	176.68	328.4	23.02	180.21	359.06	23.49	183.88
284.96	20.43	197.53	311.89	20.84	201.47	339.94	21.26	205.58

Варіант 13			Варіант 14			Варіант 15		
186.95	9.22	61.5	200.55	9.42	62.82	214.74	9.62	64.18
232.69	12.08	92.19	250.24	12.34	94.15	268.55	12.6	96.2
258.35	14.02	100.2	278.48	14.32	102.33	299.48	14.63	104.56
287.01	15.21	113.94	308.98	15.54	116.36	331.91	15.87	118.9
294.88	16.85	126.35	318.73	17.21	129.04	343.63	17.59	131.84
311.21	16.75	135.82	335.3	17.11	138.71	360.45	17.48	141.73
335.26	19.46	143.82	362.69	19.87	146.89	391.31	20.3	150.08
318.59	17.66	148.09	343.79	18.04	151.25	370.09	18.43	154.54
352.87	21.3	183.85	382.59	21.75	187.76	413.61	22.23	191.85
343.55	20.06	202.53	371.79	20.49	206.84	401.26	20.94	211.34
391.02	23.98	187.72	424.35	24.49	191.72	459.13	25.02	195.89
369.18	21.71	209.86	399.67	22.17	214.33	431.49	22.65	219

Варіант16			Варіант 17			Варіант18		
239.61	9.84	65.61	255.33	10.06	67.09	271.76	10.29	68.64
300.85	12.88	98.34	321.14	13.18	100.57	342.35	13.48	102.89
336.7	14.96	106.88	359.97	15.3	109.3	384.3	15.65	111.82
372.45	16.23	121.53	397.85	16.59	124.29	424.41	16.98	127.16
388	17.98	134.77	415.59	18.38	137.82	444.43	18.81	141.01
404.98	17.87	144.87	432.84	18.27	148.16	461.96	18.7	151.58
442.42	20.75	153.41	474.15	21.22	156.89	507.31	21.71	160.51
416.82	18.84	157.97	445.96	19.27	161.54	476.43	19.71	165.28
469.24	22.72	196.11	503.63	23.24	200.55	539.58	23.77	205.18
453.91	21.4	216.03	486.57	21.89	220.92	520.72	22.39	226.03
521.61	25.58	200.23	560.17	26.16	204.77	600.49	26.76	209.5
488.4	23.15	223.86	523.67	23.68	228.93	560.55	24.22	234.22

Варіант19			Варіант 20			Варіант 21		
288.94	10.54	70.26	306.92	10.79	71.94	325.75	11.05	73.7
364.53	13.8	105.31	387.75	14.13	107.84	412.06	14.47	110.48
409.75	16.02	114.46	436.38	16.4	117.2	464.27	16.81	120.07
452.19	17.38	130.15	481.26	17.79	133.28	511.71	18.23	136.54
474.61	19.25	144.33	506.19	19.71	147.79	539.25	20.2	151.4
492.43	19.14	155.15	524.32	19.6	158.87	557.7	20.08	162.76
542.01	22.23	164.29	578.32	22.76	168.23	616.35	23.32	172.35
508.3	20.18	169.17	541.66	20.66	173.23	576.58	21.17	177.46
577.2	24.33	210.01	616.56	24.92	215.05	657.79	25.53	220.31
556.44	22.92	231.35	593.83	23.47	236.9	632.97	24.04	242.69
642.67	27.39	214.43	686.82	28.05	219.58	733.05	28.74	224.95
599.13	24.79	239.73	639.51	25.39	245.49	681.79	26.01	251.49

Варіант22			Варіант 23			Варіант24		
345.47	11.33	75.54	366.14	11.61	77.45	387.8	11.91	79.45
437.52	14.83	113.23	464.21	15.21	116.1	492.19	15.6	119.09
493.48	17.22	123.06	524.1	17.66	126.18	556.2	18.12	129.43
543.59	18.68	139.94	577.02	19.16	143.48	612.06	19.65	147.18
573.89	20.7	155.17	610.21	21.22	159.11	648.28	21.77	163.21
592.68	20.58	166.81	629.33	21.1	171.04	667.77	21.64	175.45
656.19	23.9	176.64	697.94	24.5	181.12	741.73	25.13	185.79
613.17	21.69	181.88	651.52	22.24	186.49	691.74	22.82	191.31
700.98	26.16	225.8	746.26	26.82	231.52	793.74	27.52	237.49
673.99	24.64	248.74	716.98	25.27	255.04	762.07	25.92	261.62
781.49	29.45	230.55	832.26	30.2	236.39	885.51	30.98	242.49
726.09	26.66	257.75	772.52	27.33	264.28	821.22	28.04	271.1

Варіант25			Варіант 26			Варіант27		
410.53	12.23	81.54	434.38	12.55	83.71	459.42	12.89	85.99
521.54	16.01	122.22	552.34	16.44	125.48	584.69	16.89	128.89
589.88	18.59	132.83	625.23	19.09	136.38	662.34	19.61	140.08
648.82	20.17	151.05	687.4	20.7	155.08	727.91	21.27	159.29
688.23	22.34	167.5	730.16	22.94	171.97	774.18	23.56	176.64
708.09	22.21	180.05	750.41	22.8	184.86	794.84	23.42	189.88
787.68	25.79	190.67	835.9	26.48	195.76	886.53	27.2	201.07
733.93	23.42	196.33	778.21	24.04	201.57	824.7	24.69	207.04
843.55	28.24	243.73	895.84	28.99	250.24	950.75	29.78	257.03
809.37	26.6	268.49	859.01	27.31	275.66	911.15	28.05	283.14
941.38	31.79	248.86	1000.02	32.64	255.5	1061.61	33.52	262.44
872.31	28.78	278.22	925.93	29.54	285.65	982.24	30.35	293.4

Варіант28			Варіант 29			Варіант30		
485.71	13.25	88.36	513.34	13.62	90.84	542.38	14.01	93.43
618.65	17.35	132.45	654.35	17.84	136.16	691.87	18.35	140.04
701.33	20.15	143.95	742.29	20.71	147.98	785.36	21.3	152.2
770.46	21.85	163.69	815.17	22.47	168.28	862.16	23.11	173.08
820.43	24.21	181.51	869.04	24.89	186.6	920.14	25.6	191.92
841.51	24.07	195.12	890.55	24.74	200.6	942.1	25.45	206.31
939.72	27.95	206.62	995.62	28.74	212.42	1054.4	29.56	218.47
873.53	25.37	212.76	924.86	26.09	218.72	978.81	26.83	224.96
1008.44	30.6	264.12	1069.07	31.46	271.53	1132.82	32.36	279.27
965.91	28.82	290.96	1023.47	29.63	299.12	1083.98	30.48	307.64
1126.31	34.45	269.68	1194.31	35.42	277.25	1265.81	36.43	285.15
1041.39	31.18	301.5	1103.56	32.06	309.96	1168.92	32.97	318.79

Контрольні питання для підготовки до виконання практикуму

1. Запишіть в загальному вигляді структурну форму моделі на основі одночасних рівнянь.
2. Що означає зведена форма моделі? Як її одержати?
3. Дайте визначення рекурсивних систем і запишіть модель на основі рекурсивної системи.
4. Яка система рівнянь називається точно ідентифікованою?
5. Яка система рівнянь називається надідентифікованою?
6. Запишіть умову ідентифікованості системи рівнянь.
7. На основі якого методу можна оцінити параметри моделі, якщо вона складається із системи рекурсивних рівнянь?
8. Який метод оцінки параметрів можна застосувати, коли всі рівняння моделі є точно ідентифікованими?

9. На основі якого методу можна оцінити параметри моделі, якщо вона має надідентифіковані рівняння?

Рекомендована література: [8].

Розділ 6. МЕТОДИКА БОКСА-ДЖЕНКІНСА ПОБУДОВИ МАТЕМАТИЧНОЇ МОДЕЛІ

Вище ми розглянули узагальнений алгоритм побудови математичної моделі, який може бути застосований до будь-якого класу процесів. Тепер розглянемо методику Бокса-Дженкінса [9-10], яка більше орієнтована на побудову моделей фінансово-економічних, соціальних та екологічних процесів, для яких, як правило, набагато складніше поставити експеримент та отримати інформативні експериментальні дані в достатньому об'ємі. Хоча в наведеному вигляді ця методика також може бути успішно застосована до побудови моделей динаміки технічних систем та технологічних процесів.

Основи методики побудови моделей часових рядів запропоновані Боксом і Дженкінсом. Для ознайомлення з методом рекомендуємо [9]. Модифікована методика побудови математичної моделі процесу, в основу якої покладено ідеї Бокса і Дженкінса, складається з таких кроків:

- Виконати аналіз процесу, для якого будується модель, на основі спеціальних літературних джерел, експертних оцінок протікання процесу, візуального дослідження вимірів вхідних і вихідних змінних, представлених часовими рядами, та іншої доступної інформації.
 - Попередня обробка експериментальних даних.
 - Аналіз часових рядів на можливу наявність нелінійностей за допомогою множини статистичних критеріїв.
 - Вибрати структури моделей-кандидатів, для чого необхідно виконати такі дії: (1) обчислити та виконати аналіз кореляційної матриці для часових рядів залежної та незалежних змінних з метою визначення тих екзогенних змінних, які необхідно включити в модель; (2) обчислити автокореляційну (АКФ) та часткову автокореляційну функції (ЧАКФ) залежної змінної з метою визначення оцінки порядку авторегресійної частини моделі; (3) оцінити характеристики інших елементів структури математичної моделі.
 - Вибрати метод (методи) оцінювання параметрів математичних моделей вибраних структур. Найчастіше це метод найменших квадратів (МНК), метод максимальної правдоподібності (ММП) та їх модифікації.
 - Вибрати кращу з оцінених моделей-кандидатів за допомогою множини статистичних критеріїв адекватності (якості) моделі.
- Тепер розглянемо докладно кожний з етапів побудови моделі.

6.1. Аналіз процесу

Аналіз процесу – це надзвичайно важливий етап, коректне виконання якого потребує досвіду дослідження реальних процесів різної природи. Ігнорування цього етапу призводить до неможливості побудови моделі

високого ступеня адекватності процесу. Як і при використанні узагальненого алгоритму побудови моделі, аналіз процесу спрямовується на виконання таких задач:

- визначення числа входів і виходів, тобто визначення вимірності процесу;
- встановлення логічних зв'язків між змінними та аналіз можливостей їх математичного описання (коректного об'єднання в одному математичному виразі);
- визначення кількості зовнішніх збурень та їх типу (детерміноване чи стохастичне);
- встановлення можливості декомпозиції процесу на окремі підпроцеси, які є простішими як з точки зору їх функціонування, так і з точки зору математичного описання; декомпозиція – це досить складний процес, який ґрунтується на спеціальних математичних методах;
- якщо процес має ієрархічну структуру (верхній та нижній рівень функціонування), то необхідно чітко розмежувати ці рівні, визначити функції кожного з них і встановити які типи зв'язків існують між ними; наприклад, технологічні процеси часто можна розмежувати на два та більше рівнів;
- використання знань із спеціальної літератури щодо особливостей функціонування процесу, відомих законів та закономірностей його протікання, виявлення існуючих моделей процесу та досвіду його теоретичного чи експериментального дослідження;
- при наявності розроблених моделей досліджуваного процесу необхідно встановити їх недоліки та переваги, а також визначити можливість подальшого використання (модифікації); аналіз і використання існуючих моделей надає можливість суттєво скоротити час та інші витрати на побудову та використання моделі.

Отриману інформацію максимально використовують для попереднього оцінювання структури моделі або декількох моделей-кандидатів, параметри яких оцінюють за допомогою експериментальних даних. В процесі виконання аналізу функціонування досліджуваного процесу доцільно використовувати та порівнювати інформацію з різних джерел. Це особливо стосується фінансово-економічних процесів, щодо яких може надходити інформація з протиріччями.

6.2.Попередня обробка даних

Процес попередньої обробки експериментальних (статистичних) даних, як правило, включає такі операції:

- нормування та візуальну перевірку даних і, при необхідності, їх корегування; нормування даних означає їх логарифмування або приведення до зручного діапазону їх зміни, наприклад, від 0 до 1; від -1 до $+1$; від $+10$ до -10 і т.ін.;
- корегування даних полягає у заповненні пропусків та зменшенні викидів (екстремальних значень), що виходять за основний діапазон значень змінних.
- формування перших або різниць вищих порядків, які необхідні для аналізу відповідних складових часового ряду.

Поширеним методом нормування даних є їх логарифмування з наступним формуванням додаткових часових рядів з перших чи других різниць. Нагадаємо, що перші різниці представляють собою наближений дискретний аналог першої похідної, а другі різниці – другої похідної. Використання різниць дає можливість будувати моделі для швидкості та прискорення основної змінної. Часто із значень ряду віднімають його середнє для того щоб отримати можливість працювати з відхиленнями, а не повними значеннями змінних. Такий підхід застосовують, наприклад, при побудові моделей у просторі станів. Застосування того чи іншого методу підготовки даних для моделювання визначається в кожному випадку по-своєму.

6.3.Обробка екстремальних (аномальних) значень

Хоча виявлення та обробка екстремальних значень – це велика окрема тема для дослідження, розглянемо деякі можливості щодо розв’язання цієї проблеми. В подальшому будемо вважати дані *аномальними*, якщо вони виникли внаслідок впливу значних похибок вимірів або похибок, пов’язаних з некоректним збором статистичних даних. Якщо можна встановити факт наявності аномальних даних, то їх просто видаляють з розподілу.

Екстремальні значення – це правильно виміряні (зібрані) дані, які характеризують фактичні раптові (стрибоподібні) зміни процесу. Підхід до розв’язання задачі дослідження екстремальних значень спостережень залежить від поставленої мети. Якщо дослідника цікавить тільки факт наявності таких значень (наприклад, з метою виявлення умов, що приводять до появи екстремальних значень), то досить мати надійний критерій для виявлення таких спостережень.

Якщо ж ставиться задача виявлення і виключення екстремальних значень (наприклад, з метою покращення оцінок статистичних параметрів і моделей), то виникає задача – як правильно виконати обробку даних. Так, спираючись на критерій для визначення екстремальних значень, можна визначити величину зміщення оцінок параметрів.

Критерії аналізу екстремальних значень застосовують з метою:

- вирівняти спостереження перед аналізом (як правило, суттєво зменшити) великі значення;

- переконатись, що дані містять аномальні значення, що свідчить про необхідність перегляду процедури отримання даних;
- виділити спостереження, які є цікавими з точки зору їх аномальності та, по можливості, описати встановлений ефект математично.

Класичний підхід до виявлення аномальних спостережень полягає в тому, що вибіркові спостереження розглядають як випадкові, нормально розподілені величини. При цьому для аналізу (виявлення екстремальних значень) створюється статистика (статистичний тест), яка є чутливою до різких відхилень такого типу. Необхідно встановити розподіл цієї статистики при нульовій гіпотезі, що всі спостереження належать нормальній сукупності, а потім відхилити цю гіпотезу, якщо виявиться, що обчислена статистика їй протирічить.

Розглянемо можливий критерій відкидання екстремальних значень. Нехай дана деяка вибірка $\{x_1, x_2, \dots, x_N\}$, $N \geq 3$, яка, за припущенням, є випадковою для випадкової змінної X з нормальним розподілом: $\{X\} \sim (\mu_x, \sigma_x^2)$. Позначимо відхилення від середнього через

$$\tilde{x}_i = x_i - \bar{x}, \quad i = 1, 2, \dots, N, \quad \text{де} \quad \bar{x} = \frac{1}{N} \sum_{k=1}^N x(k).$$

Якщо виділити одне значення із спостережень, то вибіркове середнє для спостережень, що залишились, визначається як

$$\sum_{\substack{k=1 \\ k \neq i}}^N \frac{x_k}{N-1} = \bar{x} - \frac{\tilde{x}_i}{N-1}. \quad (6.2.1)$$

Якщо виділити декілька значень $x_1, x_2, \dots, x_r$, то вибіркове середнє дорівнює:

$$\bar{x} - \frac{\tilde{x}_1 + \tilde{x}_2 + \dots + \tilde{x}_r}{N-r}. \quad (6.2.2)$$

Позначимо максимальне відхилення через $\tilde{x}_m = x_m - \bar{x}$. Тепер правило визначення екстремального значення можна сформулювати так: *при заданому значенні c спостереження x_m відкидається, якщо $|\tilde{x}_m| > c S_x$, де S_x – вибіркове стандартне відхилення змінної X .*

Якщо вибірка має досить великий об'єм, то значення x_m видаляється і аналіз продовжується. Величина константи c може змінюватись із зміною довжини вибірки; вона зв'язана неявно з t – статистикою:

$$\sqrt{\frac{N c^2 (v + v_0 - 1)}{v(v + v_0 - \frac{N c^2}{v})}} \approx t_{1-\alpha/2}^{v_0+v-1}, \quad (6.2.3)$$

де $v = N - 1$; α – рівень значущості; v_0 – будь-яке інше число додаткових ступенів свободи, яке зв'язане з оцінюванням σ_x^2 за вибіркою, об'єм якої не дорівнює N ($v_0 = 0$, якщо такої інформації немає). Також існує наближений вираз для c через розподіл F :

$$c \approx \left(\frac{v}{N} \right)^{1/2} \left(\frac{3 F_{1-q}}{1 + [3 F_{1-q} - 1] / (v + v_0)} \right)^{1/2}, \quad (6.2.4)$$

де $q = \Delta \sigma_x^2 \frac{v}{N}$; $\Delta \sigma_x^2$ – очікуваний приріст дисперсії внаслідок появи екстремальних значень. При використанні (3.2.4) значення c визначається додатнім значенням квадратного кореня.

Виразом (3.2.4) можна скористатись наступним чином: якщо з ряду значення не видалялись, то допустимий (очікуваний) відносний приріст дисперсії (“премію”) $\Delta \hat{\sigma}_x^2$ необхідно помножити на v/N і, таким чином, отримаємо q . За його допомогою знайдемо відповідну верхню процентну точку для відношення дисперсій F_{1-q} при трьох і $v + v_0 - 1$ ступенях свободи. За виразом (3.2.4) обчислимо значення c і застосуємо критерій до x_m . Очікуваний відносний приріст дисперсії (“премію”) залежить від того, наскільки ймовірно є поява екстремальних значень, наприклад, можна прийняти невеликий відносний приріст $\Delta \sigma_x^2 = 0,01 \div 0,03$.

Наприклад, якщо $N = 4$, $v = 3$ і $v/N = 0,75$, то при $\Delta \sigma_x^2 = 0,02$ маємо: $q = 0,02 \cdot 0,75 = 0,05$. При 3-х степенях свободи $F_{1-q} = F_{1-0,05} = 9,28$. Тепер знайдемо значення c :

$$c = (0,75)^2 \left(\frac{3 F_{0,95}}{1 + (3 F_{0,95} - 1) / 3} \right)^{1/2} = 0,831.$$

Спостереження x_m необхідно видалити, якщо $|\tilde{x}_m| > 0,831 S_x$. Можливі інші підходи до аналізу екстремальних значень.

Приклад 2.1. Обчислення критерію для виявлення екстремального значення. Є ряд значень: $X = \{23,2 \ 23,4 \ 23,5 \ 24,1 \ 25,5\}$. Встановити, чи можна вважати значення 25,5 екстремальним і чи необхідно його видалити з вибірки?

Розв'язок. Обчислимо: $\bar{x} = 23,9$; $\tilde{x}_5 = 25,5 - 23,9 = 1,6$; $S_x = 0,77$. Для $\alpha = 0,05$; $\nu = 4$ і $N = 5$ за виразом (6.2.3) маємо:

$$\left(\frac{15c^2}{9 - 4c^2} \right)^{1/2} = 2,776^3.$$

Звідси, за методом проб та помилок, $c = 1,49$. Згідно з критерієм

$$|1,6| > 1,49 \cdot 0,77 = 1,05,$$

тобто спостереження x_5 видаляється.

6.4. Аналіз наявності нелінійностей

Однією з проблем при визначенні структури моделі є встановлення факту наявності нелінійностей в досліджуваному процесі та їх типу. Для розв'язання цієї проблеми обов'язково використовують *візуальний аналіз* даних та формальні тести на наявність нелінійностей. Досвідченому фахівцю з моделювання візуальний аналіз дозволяє оперативно виявити наявність участків з лінійним або нелінійним трендом, в якійсь мірі наявність гетероскедастичності та значних викидів (імпульсів), які можуть суттєво впливати на якість моделі. Необхідно зазначити, що існують окремі навчальні курси з візуального аналізу даних. Це свідчить про те, що не варто нехтувати такою доступною, але ефективною можливістю дослідження даних. Психологи встановили, що один інформативний рисунок може замінити до двох тисяч слів.

Існує також ряд формальних тестів на наявність нелінійності. Розглянемо простий тест для визначення наявності нелінійності. Цей тест можна застосувати у випадку, коли можна набрати кілька груп (вибірок) спостережень для одного і того ж процесу:

$$\hat{F} = \frac{\frac{1}{m-2} \sum_{i=1}^m n_i (\bar{y}_i - \hat{y}_i)^2}{\frac{1}{n-m} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2},$$

де $\bar{y}_i$ – середнє значення для i -ї групи (вибірки або групи) даних; $\hat{y}_i$ – середнє для лінійної апроксимації даних; m – число груп даних; n_i – число

вимірів в i -й групі; n – загальне число вимірів. Фактично, наведена статистика представляє собою таке відношення:

$$\hat{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{Відхилення значень } y(k) \text{ від групових середніх}}.$$

Якщо статистика $\hat{F}$ з $\nu_1 = m - 2$ та $\nu_2 = n - m$ степенями свободи досягає або перевищує рівень значущості, то гіпотезу щодо лінійності необхідно відхилити. Недоліком даного підходу є те, що для його застосування необхідно мати декілька (не менше трьох) груп даних для одного і того ж процесу, які можна отримати в результаті виконання повторних експериментів. Очевидно, що це не завжди можливо.

Наявність нелінійності можна встановити також за допомогою вибірових *нелінійних кореляційних функцій* (НКФ), тобто, кореляційних функцій, розрахованих за вибірками експериментальних (статистичних) даних. Наприклад, якщо дискретна НКФ

$$r_{yx^2}(s) = r_{y(k)x^2(k-s)} = \frac{1}{N} \frac{\sum_{k=s+1}^N \{ [y(k) - \bar{y}] [x(k-s) - \bar{x}]^2 \}}{\sigma_y \sigma_x^2}, \quad s = 0, 1, 2, 3, \dots,$$

містить значення, які суттєво відрізняються від нуля в статистичному сенсі, то процес містить квадратичну нелінійність відносно регресора x .

Наявність нелінійного детермінованого тренду в процесі можна визначити шляхом оцінювання рівняння:

$$y(k) = a_0 + c_1 k + c_2 k^2 + \dots + c_m k^m,$$

яке представляє собою поліном порядку m відносно часу. Якщо хоча б один із коефіцієнтів c_i , $i = 1, \dots, m$ є статистично значимим, то гіпотеза щодо відсутності тренду відхиляється. Якщо тренд відносно швидко змінює свій напрям розвитку і для нього трудно знайти адекватне функціональне описання, то застосовують моделі випадкових трендів, які ґрунтуються на комбінаціях випадкових величин.

Автоматично оцінює структуру математичної моделі *метод групового врахування аргументів* (МГВА), запропонований академіком Івахненком О.Г. (Київський інститут кібернетики). Цей метод вже багаторазово застосовано до широкого класу процесів; його успішно застосовують і сьогодні до моделювання процесів різної природи з нелінійностями та нестационарностями. Подальшим розвитком даного методу є нечіткий МГВА,

який ґрунтується на нечіткому представленні параметрів оцінюваної моделі (розділ 15).

6.5. Формування інших елементів структури моделі

На третьому етапі необхідно вибрати структури моделей кандидатів. Поняття структури моделі було розглянуто вище, але нагадаємо, що воно включає *порядок моделі* (найвищий порядок рівнянь, що його утворюють); *вимірність* (число рівнянь моделі); час *запізнення* по входу (лаг) та його оцінка; *можливі нелінійності* та їх тип; *зовнішні збурення* та їх тип (детерміновані або випадкові; адитивні та мультиплікативні).

Для того щоб визначити які незалежні змінні (регресори) необхідно включити в праву частину рівняння, обчислюють коефіцієнт кореляції між залежною та відповідною незалежною змінною.

Коефіцієнт кореляції, а в загальному випадку кореляційна функція, дозволяє встановити факт існування зв'язку між змінними. Кореляція може бути лінійною або нелінійною в залежності від типу функціональної залежності, яка фактично має місце між змінними. В більшості практичних випадків розглядають лінійну кореляцію (взаємозв'язок) між змінними, але більш глибокий аналіз потребує використання нелінійних залежностей. Складну нелінійну залежність можна спростити, але знати про її існування необхідно для того щоб побудувати, при необхідності, (складнішу за структурою) модель процесу з вищим ступенем адекватності.

Кореляційна матриця дає можливість встановити існування зв'язку між залежною (ендогенною) змінною та незалежними (екзогенними) змінними в правій частині. Розглянемо кореляційну матрицю **R** вимірності 3×3 , яка будується для трьох змінних x, y, z :

..

$$\mathbf{R} = \begin{bmatrix} r_{yy} & r_{xy} & r_{zy} \\ r_{yx} & r_{xx} & r_{zx} \\ r_{yz} & r_{xz} & r_{zz} \end{bmatrix}, \quad \text{де } r_{yx} = r_{xy}, \quad r_{yz} = r_{zy}, \quad r_{xz} = r_{zx}.$$

Нехай y – показник якості технологічного процесу; x, z – технологічні параметри, які, за припущенням, впливають на показник якості. Тобто, ставиться задача встановлення існування залежності вигляду:

$$y = f(x, z),$$

яка може бути представлена у формі регресії змінної y на незалежні змінні x, z :

$$y(k) = a_0 + a_1 x(k) + a_2 z(k) + \varepsilon(k),$$

де k – дискретний час (наприклад, в долях секунди, секундах, хвилинах, годинах, днях, тижнях, місяцях і т.і.); $\varepsilon(k)$ – випадкова змінна, введення якої в модель пояснюється наступними причинами:

- часто буває неможливо встановити всі незалежні змінні, які впливають на залежну змінну, а тому наведене рівняння описує процес з похибкою;
- можуть існувати такі незалежні змінні, які неможливо виміряти і включити в модель, а тому їх розглядають як збурення і вважають, що їх спільний вплив на залежну змінну описується випадковою змінною $\varepsilon(k)$;
- в наведене вище регресійне рівняння можуть бути введені пояснюючі змінні, які є формально корельованими із залежною змінною, але фактично вони не впливають на неї;
- для будь-якого методу оцінювання параметрів рівнянь властиві методичні похибки, які повинні бути, по можливості, враховані в моделі.

Вважається, що сукупний вплив всіх вказаних факторів можна описати в деякій мірі за допомогою випадкової змінної $\varepsilon(k)$. Оскільки вона не вимірюється, то оцінити її значення (*похибку моделі або залишок*) можна тільки після оцінювання коефіцієнтів моделі, тобто

$$\hat{\varepsilon}(k) = e(k) = y(k) - \hat{y}(k),$$

де $\hat{y}(k)$ – оцінка змінної $y(k)$, отримана за допомогою моделі; $y(k)$ – фактичний вимір.

Для обчислення елементів матриці $\mathbf{R}$ необхідно мати синхронізовані в часі вибірки значень всіх трьох змінних y, x, z . Формула для розрахунку коефіцієнтів кореляції має вигляд:

$$r_{yx} = \frac{\frac{1}{N-1} \sum_{k=1}^N \{[x(k) - \bar{x}][y(k) - \bar{y}]\}}{\sigma_x \sigma_y},$$

де $\bar{x}, \bar{y}$ – вибіркові середні значення змінних x, y ; σ_x, σ_y – стандартні відхилення цих змінних, тобто корені квадратні з їх дисперсії. Наприклад,

$$\sigma_y = \sqrt{\sigma_y^2} = \left[\frac{1}{N-1} \sum_{k=1}^N [y(k) - \bar{y}]^2 \right]^{1/2},$$

де N – число вимірів змінної y ; $\bar{y}$ – вибіркове середнє значення ряду $\{y(k)\}$, яке обчислюється за відомою формулою:

$$\bar{y} = \frac{1}{N} \sum_{k=1}^N y(k).$$

Очевидно, що перед формальним обчисленням коефіцієнтів кореляції, необхідно зробити аналіз процесу і визначити наявність (або відсутність) логічних зв'язків між змінними. Це дозволяє ввести до розгляду тільки ті змінні, які дійсно впливають на залежну змінну, наприклад, на показник якості. Очевидно, що для правильного вибору незалежних (екзогенних) змінних необхідно глибоко знати технологічний або інший процес, що моделюється.

На основі значень коефіцієнтів кореляції приймається рішення про включення їх у рівняння регресії:

$$y(k) = a_0 + b_1 x(k) + b_2 z(k) + \varepsilon(k),$$

яке може бути представлено в загальному вигляді як

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + a_3 x_3(k) + \dots + a_{p-1} x_{p-1}(k) + \varepsilon(k).$$

Можна показати, що між коефіцієнтами регресії b_1, b_2 і коефіцієнтами кореляції r_{yx}, r_{yz} існує однозначний взаємозв'язок.

Останнє рівняння представляє собою *лінійну регресію* p – го порядку, але досить часто необхідно застосовувати більш складні нелінійні моделі. Характерним представником нелінійної відносно змінних регресії є поліноміальна регресія порядку $p - 1$:

$$y(k) = a_0 + a_1 x(k) + a_2 x^2(k) + a_3 x^3(k) + \dots + a_{p-1} x^{p-1}(k) + \varepsilon(k).$$

Хоча в це рівняння включено тільки одну незалежну змінну $x(k)$, очевидно, що воно може бути розширене будь-якими іншими змінними.

Для визначення необхідності введення в рівняння регресії авторегресійної складової необхідно обчислити і дослідити вибірккову *автокореляційну функцію* змінної $y(k)$. Рівняння з авторегресійною складовою має вигляд:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + b_1 x(k) + b_2 z(k) + \varepsilon(k),$$

тобто, в рівняння регресії введено авторегресійну (АР) складову другого порядку. Порядок авторегресії визначається за допомогою автокореляційної функції. Число коефіцієнтів автокореляційної функції, які відмінні від нуля в статистичному сенсі, і буде складати оцінку порядку авторегресії.

Природа авторегресії пояснюється існуванням так званої “пам’яті” процесу, яка проявляється в тому, що його поточний стан в значній мірі визначається попередніми станами. Наприклад, стан людини вранці залежить від того, яким було самопочуття ввечері та в попередні дні. На формування ринкових цін суттєво впливають значення цін в попередні періоди часу. Поточний стан технічної системи або технологічного процесу також залежить від її стану в попередні моменти часу.

АКФ та ЧАКФ використовують для визначення попередньої оцінки порядку авторегресійної частини моделі, тобто скільки затриманих в часі значень необхідно брати для описання процесу. При цьому необхідно врахувати, що АКФ дає менш «чітку» оцінку порядку процесу ніж ЧАКФ. Наприклад, для процесу АР(1) значення основної змінної $y(k)$ та $y(k-2)$ будуть корельованими, незважаючи на те, що $y(k-2)$ не присутнє в моделі. Кореляція між $y(k)$ і $y(k-2)$, тобто ρ_2 , дорівнює коефіцієнту кореляції між значеннями $y(k)$ і $y(k-1)$, помноженому на коефіцієнт кореляції між $y(k-1)$ і $y(k-2)$ або $\rho_2 = \rho_1 \rho_1 = \rho_1^2$. Подібні «непрямі» кореляції наявні в АКФ будь-якого процесу авторегресії.

Вибіркова АКФ обчислюється за виразом:

$$r_y(s) = r_{y(k)y(k-s)} = \frac{1}{N-1} \frac{\sum_{k=s+1}^N \{[y(k) - \bar{y}][y(k-s) - \bar{y}]\}}{\sigma_y^2}, \quad s = 1, 2, 3, \dots,$$

де σ_y^2 – вибіркова дисперсія змінної $y(k)$; $\bar{y}$ – середнє значення вибірки даних.

Число коефіцієнтів АКФ, відмінних від нуля в статистичному сенсі, вказує на порядок авторегресійної частини моделі. Для стаціонарного процесу (це процес із постійними середнім значенням, дисперсією та коваріацією) коефіцієнти $r_y(s)$ мають нормальний розподіл та нульове середнє.

На відміну від АКФ, часткова АКФ між значеннями $y(k)$ та $y(k-s)$ виключає вплив величин $y(k-1) \dots y(k-s+1)$, а це означає, що коефіцієнти ЧАКФ чіткіше відображають зв’язок між окремими значеннями основної змінної. Так, для процесу АР(1) ЧАКФ між $y(k)$ та $y(k-2)$ дорівнює нулю за визначенням, що підтверджується обчисленими значеннями ЧАКФ. Для

того, щоб знайти попередню оцінку порядку моделі, вибіркові коефіцієнти ЧАКФ (тобто, коефіцієнти, знайдені за вибіркою даних) можна обчислити також за допомогою простого методу, який полягає в наступному.

а) Формують додатковий часовий ряд з відхилень основної змінної:

$$\{y'(k)\} = \{y(k)\} - \mu,$$

де μ – середнє значення ряду.

б) Формують рівняння першого порядку

$$y'(k) = \Phi_{11} y'(k-1) + e(k),$$

де $e(k)$ – похибка моделі. В такому рівнянні Φ_{11} відіграє роль коефіцієнта АКФ та ЧАКФ між $y(k)$ та $y(k-1)$. Для оцінювання двох коефіцієнтів можна сформулювати рівняння другого порядку:

$$y'(k) = \Phi_{11} y'(k-1) + \Phi_{22} y'(k-2) + e(k),$$

де Φ_{22} – коефіцієнт ЧАКФ між $y(k)$ та $y(k-2)$.

Коефіцієнти ЧАКФ можна обчислити також за допомогою коефіцієнтів АКФ, використовуючи наступні вирази:

$$\Phi_{11} = r(1), \quad \Phi_{22} = \frac{r_2 - r_1^2}{1 - r_1^2};$$

$$\Phi_{ss} = \frac{r_s - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_{s-j}}{1 - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_j}.$$

В загальному випадку коефіцієнти ЧАКФ стаціонарного процесу $АРКС(p,q)$ повинні збігатися до нуля, починаючи із p -го значення. АКФ процесу $АРКС(p,q)$ починається збігатися до нуля при значеннях зміщення $s \geq q$.

Коли ми говоримо, що значення коефіцієнтів автокореляційної функції повинні бути відмінними від нуля в статистичному сенсі, це означає, що існує вираз (формула), який дозволяє підтвердити або спростувати цей факт. Одним із загальноприйнятих підходів до встановлення того факту, що коефіцієнти АКФ суттєво відмінні від нуля в статистичному сенсі, є обчислення та аналіз значущості статистичного параметра (або просто статистики) Льюнга-Бокса $Q(r_k)$ за формулою :

$$Q(r_k) = N(N+2) \sum_{k=1}^s r_k^2 / (N-k),$$

де N – довжина вибірки даних змінної, для якої обчислено значення автокореляційної функції r_k ; s – число коефіцієнтів АКФ, які досліджуються на суттєву відмінність від нуля. Якщо дані згенеровані процесом АР чи АРКС [3], то значення $Q(r_k)$ асимптотично мають розподіл χ^2 з s ступенями свободи, а тому для перевірки їх значущості необхідно користуватись відповідними статистичними таблицями. Очевидно, що більші значення вибіркової автокореляційної функції приводять до більших значень $Q(r_k)$.

Так, для ідеального процесу білого шуму $Q(r_k) = 0$. Якщо значення $Q(r_k)$, обчислене за наведеним виразом, перевищує критичне значення з розподілу χ^2 з s ступенями свободи, то існує щонайменше одне значення $r(k)$, яке є відмінним від нуля в статистичному сенсі. Статистику Льюнга-Бокса можна застосовувати також для встановлення близькості залишків моделі до білого шуму. Однак, необхідно пам'ятати, що при обчисленні s значень кореляційної функції число ступенів вободи зменшується на число коефіцієнтів моделі. Таким чином, при аналізі залишків моделі АРКС(p, q) статистика $Q(r_k)$ має розподіл χ^2 з $s - p - q$ ступенями свободи, а із врахуванням константи $s - p - q - 1$.

Цей етап закінчується формуванням структур декількох моделей-кандидатів з векторами параметрів $\theta_1, \dots, \theta_m$, де m – число кандидатів. Кандидатів може бути декілька, оскільки встановити структуру точно за один раз, як правило, неможливо. Загалом побудова моделі високого ступеня адекватності – це трудомісткий ітераційний процес, який вимагає значних зусиль. На наступному етапі оцінюють параметри моделей-кандидатів.

6.6. Оцінювання коефіцієнтів моделей-кандидатів

На *четвертому етапі* оцінюють коефіцієнти (параметри) рівняння, використовуючи принцип економії або збереження. Цей принцип означає, що *кількість коефіцієнтів, що оцінюються, не повинна перевищувати їх необхідне число* (“необхідність” можна визначити, наприклад, як необхідність збереження в моделі основних статистичних характеристик процесу).

При моделюванні процесів будь-якої природи необхідно пам’ятати, що поведінку процесу необхідно *апроксимувати* за допомогою рівнянь, а не старатися описати її до найменших дрібниць. Необхідно враховувати також, що різні за структурою моделі можуть мати однакові властивості. Наприклад, рівняння авторегресії першого порядку

$$y(k) = 0,5y(k-1) + \varepsilon(k)$$

еквівалентне ковзному середньому у вигляді

$$y(k) = \varepsilon(k) + 0,5\varepsilon(k-1) + 0,25\varepsilon(k-2) + 0,125\varepsilon(k-3) + \dots$$

Модель, що оцінюється, повинна задовольняти принципу інверсії, тобто щоб за допомогою отриманого рівняння можна було б згенерувати початковий ряд, на основі якого оцінювались коефіцієнти. Це означає, що хоча модель і спрощена, вона повинна співпадати з досліджуваним процесом за такими основними характеристиками як середнє, дисперсія та коваріація.

В процедурі оцінювання часто використовують не абсолютні значення змінних, а їх відхилення від середнього, тобто

$$y(k) = Y(k) - \mu_y,$$

де $Y(k)$ – значення виміру, μ_y – середнє значення ряду. Якщо для оцінювання параметрів використовується рекурсивна процедура, то поточне середнє можна обчислювати за формулою:

$$\mu_y(k) = \mu_y(k-1) + \frac{1}{k}[y(k) - \mu_y(k-1)].$$

Найбільш поширеними методами оцінювання параметрів моделі є такі:

- *метод найменших квадратів* (МНК);
- *метод максимальної правдоподібності* (ММП);
- *метод допоміжної (інструментальної) змінної* (МДП);
- *нелінійний метод найменших квадратів* (НМНК)

та їх рекурсивні версії (РМНК, РММП, РМДП). Деякі методи будуть розглянуті в розділі, присвяченому методам оцінювання. Оцінки (звичайного) МНК обчислюють за допомогою наступного виразу:

$$\hat{\theta} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y},$$

де $\theta [p]$ – вектор оцінок параметрів вимірності p ; $\mathbf{X}[N \times p]$ – матриця вимірів; $\mathbf{y}[N]$ – вектор вимірів залежної змінної. В квадратних дужках вказана вимірність векторів і матриці. Елементи матриці вимірів обчислюються по-своєму для кожної конкретної моделі. Так, для моделі

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + a_3 x_3(k) + \varepsilon(k)$$

матриця вимірів має вигляд:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1(1) & x_2(1) & x_3(1) \\ 1 & x_1(2) & x_2(2) & x_3(2) \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_1(N) & x_2(N) & x_3(N) \end{bmatrix}.$$

Одиниці в першому стовпчику матриці $\mathbf{X}$ означають, що вимір при коефіцієнті a_0 завжди дорівнює одиниці.

Елементи матриці вимірів дещо ускладнюються у випадку використання поліноміальної моделі, але її також можна оцінювати за допомогою лінійних методів. Безпосереднє застосування методу мінімізації суми квадратів похибок до поліноміальної моделі порядку p приводить до формування такої матриці вимірів:

$$\mathbf{X}' = \begin{bmatrix} N & \sum_{k=1}^N x(k) & \sum_{k=1}^N x^2(k) & \dots & \sum_{k=1}^N x^p(k) \\ \sum_{k=1}^N x(k) & \sum_{k=1}^N x^2(k) & \sum_{k=1}^N x^3(k) & \dots & \sum_{k=1}^N x^{p+1}(k) \\ \cdot & \cdot & \cdot & \dots & \dots \\ \sum_{k=1}^N x^p(k) & \sum_{k=1}^N x^{p+1}(k) & \sum_{k=1}^N x^{p+2}(k) & \dots & \sum_{k=1}^N x^{2p}(k) \end{bmatrix}.$$

При такому представленні векторно-матричне рівняння для N вимірів залежної та незалежної змінних можна записати так: $\mathbf{y}' = \mathbf{X}' \theta$; звідси

$$\hat{\theta} = [\mathbf{X}']^{-1} \mathbf{y}', \text{ де } \mathbf{y}' = \begin{bmatrix} \sum_1^N y(k) & \sum_1^N x(k) y(k) & \dots & \sum_1^N x^p(k) y(k) \end{bmatrix}^T;$$

$\hat{\theta}$ – вектор оцінок параметрів моделі. Тобто оцінку вектора параметрів можна знайти шляхом знаходження розв'язку системи лінійних (нормальних) рівнянь.

Для отримання незміщених, консистентних та ефективних оцінок вектора параметрів θ лінійної регресійної математичної моделі, наприклад, моделі змішаної регресії:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + b_1 x(k) + b_2 z(k) + \varepsilon(k)$$

за допомогою методу найменших квадратів необхідно задовольнити наступні умови:

а) $\varepsilon(k)$ – некорельована послідовність випадкових чисел з нульовим середнім, тобто,

$$E[\varepsilon(k)] = 0, \quad \text{cov}[\varepsilon(k)] = E[\varepsilon(k)\varepsilon(j)] = \begin{cases} \sigma_\varepsilon^2, & k = j; \\ 0, & k \neq j. \end{cases}$$

б) послідовності $\varepsilon(k)$ і $y(k)$ не повинні бути корельовані між собою.

Зазначимо, що перевірити виконання наведених умов ми можемо тільки після оцінювання коефіцієнтів моделі, а до оцінювання можна тільки постулювати їх виконання. Тобто, після оцінювання моделі оцінка значень випадкового процесу визначається похибками моделі:

$$\hat{\varepsilon}(k) = e(k) = y(k) - \hat{y}(k),$$

що дає можливість виконати аналіз характеристик випадкового процесу $\{\varepsilon(k)\}$.

6.7. Діагностика моделей – вибір кращої з множини оцінених кандидатів

На *п'ятому етапі* аналізується якість моделі, тобто виконується перевірка оцінених кандидатів на адекватність процесу. Діагностика складається з таких кроків:

а) *Візуальне дослідження графіка похибок* моделі $e(k) = y(k) - \hat{y}(k)$, де $\hat{y}(k)$ – оцінка змінної, отримана за допомогою побудованої моделі. На графіку не повинно бути значних викидів та довгих інтервалів, на яких похибка приймає великі значення (тобто довгих інтервалів суттєвої неадекватності). У випадку

застосування рекурсивних методів оцінювання (наприклад, РМНК) найбільші похибки будуть в перехідному процесі, коли інформаційна матриця ще не містить достатньо інформації про процес.

б) *Похибки моделі не повинні бути корельовані між собою.* Для аналізу наявності кореляції між значеннями похибок необхідно обчислити АКФ та ЧАКФ для ряду $\{e(k)\}$ і за допомогою Q –статистики визначити ступінь корельованості (наприклад, Q –статистика вважається несуттєвою до рівня 10%).

Крім того, корельованість похибок визначають за допомогою статистики Дарбіна-Уотсона (DW), яка розраховується за формулою:

$$DW = 2 - 2\rho,$$

де $\rho = E[e(k)e(k-1)]/\sigma_e^2$ – коефіцієнт кореляції між сусідніми значеннями похибки; σ_e^2 – дисперсія послідовності похибок $\{e(k)\}$. Таким чином, при повній відсутності кореляції між похибками $DW = 2$ – це ідеальне значення. Граничними значеннями для DW є 0 (при $\rho = 1$) та +4 (при $\rho = -1$).

Отримати формулу $DW = 2 - 2\rho$ можна досить просто. Автори цієї статистики (Durbin і Watson) запропонували скористатись для перевірки корельованості похибок моделі таким виразом:

$$DW = \frac{\sum_{k=2}^N [e(k) - e(k-1)]^2}{\sum_{k=1}^N e^2(k)} = \frac{\sum_{k=2}^N [e(k) - e(k-1)][e(k) - e(k-1)]}{\sum_{k=1}^N e^2(k)},$$

тобто, DW можна, в деякій мірі, трактувати як коефіцієнт автокореляції для (перших різниць) приростів похибок.

Розкриваючи квадрат різниці в чисельнику, отримаємо:

$$DW = \frac{\sum_{k=2}^N e^2(k)}{\sum_{k=1}^N e^2(k)} + \frac{\sum_{k=2}^N e^2(k-1)}{\sum_{k=1}^N e^2(k)} - 2 \frac{\sum_{k=2}^N e(k)e(k-1)}{\sum_{k=1}^N e^2(k)},$$

$$\text{де } \frac{\sum_{k=2}^N e^2(k)}{\sum_{k=1}^N e^2(k)} \approx 1; \quad \frac{\sum_{k=2}^N e^2(k-1)}{\sum_{k=1}^N e^2(k-1)} \approx 1; \quad \text{а } \frac{\sum_{k=2}^N e(k)e(k-1)}{\sum_{k=1}^N e^2(k-1)} = \rho.$$

Тому можна записати, що $DW = 2 - 2\rho$.

в) Для лінійної моделі 2-3 порядку оцінки параметрів повинні збігатися до ustalених значень після 30-40 (не більше) ітерацій алгоритму оцінювання. Якщо кількість ітерацій набагато перевищує вказані числа, то це свідчить про те, що процес може бути нестационарним.

г) Перевірка значущості параметрів моделі. *Статистика Стьюдента або t -статистика* (випадкова величина, що має t -розподіл), яка використовується для визначення значущості оцінки кожного коефіцієнта в статистичному смислі, визначається за виразом:

$$t = \frac{\hat{a} - a^0}{SE_{\hat{a}}},$$

де $\hat{a}$ – оцінка коефіцієнта моделі; a^0 – нуль-гіпотеза (початкова гіпотеза) щодо цієї оцінки; $SE_{\hat{a}}$ – стандартна похибка оцінки. За нуль-гіпотезу щодо значущості оцінки можна висувати будь-яку: що коефіцієнт значимий, тобто, $(H_0 : a^0 \neq 0)$ або незначимий $(H_0 : a^0 = 0)$. Статистична теорія перевірки гіпотез пропонує висувати нуль-гіпотезу, яка є протилежною бажаному результату. В даному випадку бажаним результатом є значимість коефіцієнтів математичної моделі. Таким чином, необхідно висувати нульову гіпотезу, що коефіцієнт незначимий. Це дає можливість коректно підійти до визначення значущості оцінок коефіцієнтів та дещо спростити розрахунки.

Для того щоб встановити, чи є оцінка коефіцієнта значимою, необхідно знати довжину вибірки даних N (потужність вибірки); число ступенів свободи $f = N - n$, де n – число коефіцієнтів моделі, які оцінюються на основі ряду даних, і вибрати рівень значущості $\alpha = 1\%$ або $\alpha = 5\%$ або $\alpha = 10\%$ (для цих значень існують розраховані таблиці для критичних значень t -статистики). Фактично, рівень значущості означає ймовірність припуститись помилки першого роду при перевірці гіпотези. Згадаємо, що

$$\alpha = p\{X \in G/\omega | H_0\} = \int_{n-m(G/\omega)} L_{H_0}(X) dx,$$

де $X = [x_1, \dots, x_n] \in R^n$ – вся вибірка, яка розбивається на дві множини, що перетинаються: ω і G/ω (ω – область прийняття нуль-гіпотези); G/ω – критична область: якщо $X \in G/\omega$, то H_0 відхиляється; $L_{H_0}(X)$ – закон розподілу X . Помилка першого роду означає відхилення вірної гіпотези.

Користуючись значеннями N, f і α , з таблиць для t – розподілу знаходять критичне значення t – статистики, тобто $t_{кр}$. Для перевірки правильності висунутої гіпотези розраховане значення t порівнюють з критичним $t_{кр}$. Якщо

$$-t_{кр} < t < t_{кр} \quad \text{або} \quad |t| < |t_{кр}|,$$

то нуль-гіпотеза щодо незначущості коефіцієнта приймається (його можна не враховувати в регресії). Звідси випливає, що чим більшим є значення t – статистики для оцінки коефіцієнта, тим імовірніше, що цей коефіцієнт є значимим.

Загалом послідовність дій при перевірці значущості оцінок коефіцієнтів побудованої моделі можна сформулювати так:

- сформулювати нуль-гіпотезу щодо значущості коефіцієнта;
- обчислити значення t – статистики для кожного коефіцієнта регресії (це робить кожний пакет для математичного моделювання);
- за допомогою значень N, f і α знайти із таблиць для t – статистики її критичне значення;
- перевірити нуль-гіпотезу за наведеним вище простим правилом (аналіз виконання нерівності $-t_{кр} < t < t_{кр}$).

д) Коефіцієнт множинної детермінації R^2 , який обчислюється так:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST},$$

де $\text{var}(\hat{y})$ – дисперсія залежної змінної, оціненої за допомогою побудованої моделі; $\text{var}(y)$ – дисперсія вимірів залежної змінної; $SSE = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2$ – сума квадратів похибок (залишків) моделі (*sum of squared errors*); $SST = \sum_{k=1}^N [y(k) - \bar{y}]^2$ – загальна сума квадратів (*total sum of squares*); $\bar{y}$ – середнє значення; $SST = SSE + SSR$, де $SSR = \sum_{k=1}^N [\hat{y}(k) - \bar{y}]^2$ – загальна сума квадратів для регресії (*sum of squares for regression*).

Очевидно, що найкращим значенням є $R^2 = 1$, тобто, коли дисперсії вимірів змінної, та цієї ж змінної, оціненої за рівнянням, збігаються. Цей параметр можна трактувати, також, як міру інформативності моделі, якщо вибрати за міру інформативності дисперсію. Таким чином, R^2 показує рівень

інформативності моделі по відношенню до інформативності вибірки даних, за допомогою якої вона була оцінена.

е) Сума квадратів похибок для вибраної моделі повинна бути мінімальною, тобто,

$$\sum_{k=1}^N e^2(k) = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \rightarrow \min_{\hat{\theta}}$$

у порівнянні з усіма іншими моделями.

є) Для оцінки адекватності моделі також використовують *інформаційний критерій Акайке*

$$AIC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + 2n$$

та критерій Байєса-Шварца

$$BSC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + n \ln(N),$$

де $n = p + q + 1$ – число параметрів моделі, які оцінюються за допомогою статистичних даних (p – число параметрів авторегресійної частини моделі; q – число параметрів ковзного середнього; 1 з'являється тоді, коли оцінюється зміщення (або *перетин*), тобто a_0).

Критерії Акайке і Байєса-Шварца містять в правій частині суму квадратів похибок, а тому за цими критеріями вибирають ту модель, для якої критерії приймають найменші значення. Введення нового регресора приводить до збільшення критерію (при цьому збільшується n), але, разом з тим, зменшується сума квадратів похибок і критерій в цілому зменшується. Якщо регресор не покращує модель, то критерій збільшується. Необхідно також зазначити, що асимптотичні властивості для довгих виборок кращі у критерія Байєса-Шварца, тобто, його рекомендують застосовувати при відносно великих значеннях N ($N > 100$).

ж) Окрім згаданих параметрів, для визначення адекватності моделі в цілому використовують F – статистику Фішера, яка пропорційна відношенню:

$$F \sim \frac{R^2}{1 - R^2},$$

а для множинної (багатофакторної) регресії вона визначається за виразом

$$F = \frac{R^2}{1 - R^2} \cdot \frac{(N - p - 1)}{p},$$

де, як і раніше, N – число значень ряду; p – число параметрів моделі без врахування перетину (константи).

Таким чином, якщо $R^2 \rightarrow 1$, то $F \rightarrow \infty$. Порядок застосування F –статистики такий же, як і t –статистики. Нуль-гіпотезою є в даному випадку припущення про те, що модель неадекватна в цілому, тобто,

$$H_0 : a_1 = a_2 = \dots = a_p = 0$$

проти альтернативної гіпотези

H_1 : хоча б одне значення a_i відмінне від нуля в статистичному сенсі.

Значення $F_{\text{крит}}$ знаходять із таблиць для F – розподілу. Послідовність застосування цієї статистики можна представити наступним чином:

1. Сформулювати нуль-гіпотезу щодо адекватності моделі в цілому.
Наприклад, H_0 : модель неадекватна в цілому (або $H_0 : a_1 = a_2 = \dots = a_p = 0$).
2. Розрахувати значення F для оціненої моделі (як правило, воно розраховується всіма пакетами статистичної обробки даних).
3. Задати рівень значущості $\alpha = 1\%$ або $\alpha = 5\%$ або $\alpha = 10\%$.
4. Користуючись значеннями N , f і α , знайти критичне значення $F_{\text{крит}}$ знаходять із таблиць для F – розподілу при $(p, N - p - 1)$ степенях свободи.
5. Перевірити нуль-гіпотезу:
якщо $F > F_{\text{крит}}$, то нуль-гіпотеза щодо неадекватності моделі в цілому відкидається на вибраному рівні значущості.

Коректне застосування методики Бокса-Дженкінса забезпечує побудову адекватної математичної моделі процесу, якщо експериментальні дані відповідають *вимогам представництва та інформативності*. Перша вимога означає, що вибірка даних повинна охоплювати досить довгий проміжок часу, щоб повністю відображати поведінку того режиму функціонування процесу, для яких будується модель. Вимога *інформативності* означає, що вибірка повинна містити в собі об'єм інформації, достатній для оцінювання коефіцієнтів моделі. Наприклад, якщо моделюється процес другого порядку, то вибірка повинна забезпечувати коректне обчислення першої та другої

похідної. Іноді формально інформативність оцінюють за допомогою величини дисперсії процесу, а також за кількістю гармонічних складових, які містяться в процесі. Чим більше гармонік містить вибірка, тим вищою є її інформативність.

Умову інформативності даних пов'язують з умовою достатнього збудження процесу. Достатнє збудження означає, що вхідний сигнал повинен охоплювати всю смугу частот, які може пропускати на вихід процес (об'єкт). Тобто вхідний сигнал повинен охоплювати всю амплітудно-частотну характеристику процесу. Ця вимога залишається однаковою для процесів будь-якої природи.

6.8. Приклади побудови математичних моделей за допомогою експериментальних даних

Приклад 2.2. Побудуємо математичні моделі динаміки для наступних макроекономічних процесів України: формування внутрішнього валового продукту (ВВП), індекс споживчих цін (ІСЦ) і грошовий агрегат МЗ. Для побудови використано фактичні щомісячні дані з січня 1996 по січень 2005 року, всього 109 значень. Кореляційна матриця для цих змінних має такий вигляд:

ІСЦ	ВВП	МЗ
1	-0.3043610	-0.2491123
-0.3043610	1	0.9317529
-0.2491123	0.9317529	1

Корельованість ІСЦ з ВВП та агрегатом МЗ незначна; в подальшому ця інформація буде використана при побудові альтернативних варіантів математичних моделей процесів, що розглядаються. Корельованість між ВВП і агрегатом МЗ складає 0,931 – велике значення, яке може негативно вплинути на якість оцінок моделі при використанні цих змінних у правій частині рівняння.

6.8.1. Регресійна модель індексу споживчих цін

Авторегресійна модель індексу споживчих цін

Спочатку розглянемо можливість описання індексу споживчих цін за допомогою моделі авторегресії з ковзним середнім. Авторегресійні моделі самі прості за своєю структурою, але досить часто вони мають високу ступінь адекватності досліджуваному процесу, прийнятну для подальшого

використання. Автокореляційна функція процесу наведена в таблиці 2.1. При побудові моделі індекс споживчих цін позначимо як *isc*.

Таблиця 2.1. Автокореляційна функція процесу формування споживчих цін (ІСЦ)

Часові дані: 1996:01 2005:01

Всього спостережень: 109

АКФ	Часткова АКФ	АКФ	ЧАКФ	Q-стат.	Ймов	
.****	.****	1	0.570	0.570	36.373	0.000
.**	.*	2	0.230	-0.140	42.378	0.000
.*	.*	3	0.129	0.089	44.284	0.000
.*	.	4	0.076	-0.022	44.957	0.000
.	.	5	0.064	0.043	45.433	0.000
.*	.*	6	0.120	0.101	47.137	0.000
.*	.*	7	0.180	0.088	50.960	0.000
.*	.*	8	0.068	-0.136	51.515	0.000
.	.*	9	0.052	0.105	51.840	0.000
.*	.	10	0.072	-0.001	52.468	0.000

В таблиці 2.1 скорочення АКФ – автокореляційна функція, а ЧАКФ – часткова АКФ.

Як видно з АКФ, при побудові моделей необхідно починати з моделей нижчих порядків, які часто мають прийнятну адекватність процесу і забезпечують високу якість прогнозу. Результати оцінювання моделі авторегресії першого порядку за методом найменших квадратів наведені в таблиці 2.2.

Таблиця 2.2. Результати оцінювання моделі АР(1) для ІСЦ

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1996:02 2005:01

Використано спостережень: 108 після корегування

Моделі АР(1): $I=C(1)+C(2)*I(-1)$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	43.29386	6.624742	6.535178	0.0000
C(2)	0.569826	0.065703	8.672790	0.0000
R-квадрат	0.415067	Середнє залежної змінної		100.7407
Скорегований R-квадрат	0.409549	Станд. відхил. зал. змінної		1.506189
Станд. похибка регресії	1.157368	Інформ. критерій Акайке		3.148519
Сума квадратів похибок	141.9871	Критерій Шварца		3.198188
		Стат. Дарбіна-Уотсона		1.931805

Отримана модель АР(1):

$$isc(k) = a_0 + a_1 isc(k-1) + \varepsilon(k) = \\ = 43,294 + 0,57 isc(k-1) + e(k),$$

де $e(k)$ – залишки (похибки) моделі, значення яких можна знайти у відповідному файлі пакету програм, що застосовується для побудови моделі. Статистичні характеристики моделі:

$$R^2 = 0,415; \quad J = CKП = 141,99; \quad DW = 1,931.$$

Коефіцієнт детермінації має низьке значення (0,415), сума квадратів похибок досить високе (141,99), а статистика Дарбіна-Уотсона (1,931) наближається до найкращого значення. Таким чином, загалом адекватність моделі $AP(1)$ досить низька, а тому структура моделі потребує уточнення. Характеристики якості (історичного – на навчальній вибірці) однокрокового (історичного) прогнозу:

$$CeКП = 1,36; \quad САП = 1,02; \quad САПП = 1,008; \quad U = 0,0067,$$

тобто середньоквадратична похибка ($CeКП$), середня абсолютна похибка ($САП$), середня абсолютна похибка в процентах ($САПП$) і коефіцієнт Тейла, який свідчить про загальну придатність моделі для прогнозування (ідеальне значення – нуль).

В таблиці 2.3 наведені результати оцінювання авторегресійної моделі $AP(3)$. Всі три коефіцієнти моделі значимі.

Таблиця 2.3. Результати оцінювання моделі $AP(3)$

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1996:04 2005:01

Всього спостережень після корегування крайніх значень: 106

Модель: $I=C(1)+C(2)*I(-1)+C(3)*I(-2)+C(4)*I(-3)$

	Оцінки коеф-в	Станд. похиб.	t-статистика	Ймов.
C(1)	46.96556	9.330171	5.033729	0.0000
C(2)	0.613458	0.098563	6.223995	0.0000
C(3)	-0.157670	0.113613	-1.387787	0.1682
C(4)	0.077517	0.086493	0.896229	0.3722
R-квадрат	0.316655	Середнє залежної змінної		100.6604
Скорегований R-квадрат	0.296556	Станд. відхил. зал. змінної		1.372428
Станд. похибка регресії	1.151076	Інформ. критерій Акайке		3.156277
Сума квадратів похибок	135.1476	Критерій Шварца		3.256785
		Стат. Дарбіна-Уотсона		1.992389

Таким чином, можна записати наступну модель:

$$\begin{aligned} isc(k) &= a_0 + a_1 isc(k-1) + a_2 isc(k-2) + a_3 isc(k-3) + \varepsilon(k) = \\ &= 46,96 + 0,61isc(k-1) - 0,15isc(k-2) + 0,08isc(k-3) + e(k), \end{aligned}$$

Для цієї моделі спостерігається зменшення коефіцієнта детермінації від 0,415 до 0,317 (деяке погіршення); зменшення суми квадратів похибок від 141,99 до 135,148 і деяке покращення статистики Дарбіна-Уотсона: від 1,931 до 1,992 (похибки моделі можна вважати практично некорельованими). Тобто отримані значення характеристик моделі:

$$R^2 = 0,317; \quad J = СКП = 135,148; \quad DW = 1,992.$$

Характеристики однокрокового прогнозу для даної моделі:

$$CeKP = 1,36; \quad CAП = 1,02; \quad CAПП = 1,01; \quad U = 0,0068,$$

тобто модель загалом придатна для прогнозування (за коефіцієнтом Тейла, який наближається до ідеального значення), а три інших показники свідчать про високу точність прогнозу. Необхідно зазначити, що показники якості прогнозу для моделей AP(1) і AP(3) є практично однаковими.

Розглянемо характеристики моделі вищого порядку. В таблиці 2.4 наведені результати оцінювання моделі AP(7).

Таблиця 2.4. Результати оцінювання моделі AP(7) для ІСЦ

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1996:08 2005:01

Всього використано спостережень після корегування крайніх значень: 102

Модель: $I = C(1) + C(2)*I(-1) + C(3)*I(-2) + C(4)*I(-3) + C(5)*I(-5) + C(6)*I(-6) + C(7)*I(-7)$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	33.13446	12.79240	2.590168	0.0111
C(2)	0.603211	0.101625	5.935625	0.0000
C(3)	-0.184907	0.117804	-1.569623	0.1198
C(4)	0.125878	0.101891	1.235415	0.2197
C(5)	-0.031530	0.102954	-0.306254	0.7601
C(6)	0.093246	0.117601	0.792904	0.4298
C(7)	0.064850	0.089527	0.724365	0.4706
R-квадрат	0.346348	Середнє залежної змінної		100.6667
Скорегований R-квадрат	0.305064	Станд. відхил. зал. змінної		1.388306
Станд. похибка регресії	1.157330	Інформ. критерій Акайке		3.196268
Сума квадратів похибок	127.2443	Критерій Шварца		3.376413
		Стат. Дарбіна-Уотсона		1.811821

Порівнюючи моделі AP(3) і AP(7), можна сказати, що коефіцієнт детермінації збільшився (від 0,317 до 0,346; сума квадратів похибок моделі зменшилась від 135,15 до 127,24, а статистика Дарбіна-Уотсона зменшилась від 1,992 до 1,811). Характеристики однокрокового прогнозу для AP(7):

$$CeKP = 1,36; \quad CAП = 1,012; \quad CAПП = 1,002; \quad U = 0,0067,$$

тобто на 0,008 зменшились середня абсолютна похибка і середня абсолютна похибка в процентах в процентах .

Результати оцінювання моделі 12-го порядку наведені в таблиці 2.5.

Таблиця 2.5. Результати оцінювання моделі AP(12) для ІСЦ

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1997:01 2005:01

Всього використано спостережень після корегування крайніх значень: 97

Модель: $I = C(1) + C(2)*I(-1) + C(3)*I(-2) + C(4)*I(-3) + C(5)*I(-6) + C(6)*I(-7) + C(7)*I(-8) + C(8)*I(-9) + C(9)*I(-11) + C(10)*I(-12)$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	26.23923	14.28763	1.836500	0.0697
C(2)	0.666410	0.102198	6.520741	0.0000
C(3)	-0.247996	0.122488	-2.024654	0.0460
C(4)	0.179378	0.103248	1.737345	0.0859
C(5)	-0.038240	0.096715	-0.395392	0.6935
C(6)	0.016650	0.112406	0.148120	0.8826
C(7)	-0.037042	0.111025	-0.333639	0.7395
C(8)	0.030055	0.095744	0.313915	0.7543
C(9)	0.020445	0.093392	0.218911	0.8272
C(10)	0.149194	0.081977	1.819964	0.0722
R-квадрат	0.435002	Середнє залежної змінної		100.6082
Скорегований R-квадрат	0.376553	Станд. відхил. зал. змінної		1.342857
Станд. похибка регресії	1.060301	Інформ. критерій Акайке		3.052366
Сума квадратів похибок	97.80880	Критерій Шварца		3.317800
		Статистика Дарбіна-Уотсона		1.940130

Модель AP(12) має кращі характеристики ніж попередні моделі:

$$R^2 = 0,435; \quad J = СКП = 97,80; \quad DW = 1,94.$$

Значно зменшилась сума квадратів похибок, підвищилось значення R^2 , а значення статистики Дарбіна-Уотсона майже таке ж, як для моделі AP(3).

Характеристики однокрокового прогнозу для цієї моделі:

$$CeKP = 1,337; \quad CAП = 1,02; \quad CAПП = 1,013; \quad U = 0,0066.$$

Таким чином, характеристики однокрокового прогнозу також найкращі для моделі AP(7). Можна зробити висновок, що процес формування індексу оптових цін може бути описаний моделлю авторегресії AP(7) з високим ступенем адекватності. Ця модель забезпечує також отримання кращого однокрокового прогнозу.

6.8.2. Модель авторегресії для відхилень ІСЦ від середнього

Якщо із вихідних (фактичних) значень ряду ІСЦ відняти середнє ($isc(k) - 100,66$), $k = 1, \dots, 109$, то отримаємо ряд відхилень від середнього. Автокореляційна функція залишається фактично незмінною. Результати оцінювання моделі $AR(1)$ наведені в таблиці 2.6.

Таблиця 2.6. Результати оцінювання моделі $AR(1)$ для відхилень ІСЦ

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1996:02 2005:01

Всього використано спостережень після корегування крайніх значень: 109

Модель: $IV = C(1) + C(2) * IV(-1)$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	-0.007477	0.111831	-0.066857	0.9468
C(2)	0.569826	0.065703	8.672790	0.0000
R-квадрат	0.415067	Середнє залежної змінної		0.080741
Скорегований R-квадрат	0.409549	Станд. відхил. зал. змінної		1.506189
Станд. похибка регресії	1.157368	Інформ. критерій Акайке		3.148519
Сума квадратів похибок	141.9871	Критерій Шварца		3.198188
		Стат. Дарбіна-Уотсона		1.931805

Три вибрані статистичні характеристики адекватності цієї моделі:

$$R^2 = 0,415; \quad J = СКП = 141,99; \quad DW = 1,931.$$

Характеристики якості однокрокового прогнозу:

$$СеКП = 1,36; \quad САП = 1,02; \quad САПП = 100,40; \quad U = 0,662.$$

У порівнянні з моделлю $AR(1)$ для повних значень (без віднімання середнього) середньоквадратична похибка і середня абсолютна похибка не змінилися, але в 100 разів збільшилась $САПП$ і коефіцієнт Тейла: від 0,0067 до 0,662. Таким чином, модель $AR(1)$ для відхилень не придатна для прогнозування. Це можна пояснити тим, що відхилення мають різні знаки, що утруднює побудову моделі.

Авторегресія 13-го порядку для відхилень ІСЦ від середнього:

Таблиця 2.7. Результати оцінювання моделі $AR(13)$ для відхилень ІСЦ.

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1997:02 2005:01

Всього використано спостережень після корегування крайніх значень: 96

Модель: $IV = C(1) + C(2) * IV(-1) + C(3) * IV(-2) + C(4) * IV(-3) + C(5) * IV(-5) + C(6) * IV(-6) + C(7) * IV(-7) + C(8) * IV(-8) + C(9) * IV(-9) + C(10) * IV(-11) + C(11) * IV(-12) + C(12) * IV(-13)$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	-0.045818	0.110205	-0.415756	0.6786
C(2)	0.706976	0.107583	6.571429	0.0000
C(3)	-0.261301	0.123875	-2.109395	0.0379
C(4)	0.194879	0.105046	1.855172	0.0671
C(5)	-0.110121	0.107924	-1.020352	0.3105

C(6)	0.043907	0.120182	0.365334	0.7158
C(7)	0.002351	0.115561	0.020347	0.9838
C(8)	-0.023657	0.114078	-0.207375	0.8362
C(9)	0.033030	0.096947	0.340705	0.7342
C(10)	0.002612	0.096331	0.027112	0.9784
C(11)	0.187169	0.109248	1.713242	0.0904
C(12)	-0.074932	0.085883	-0.872494	0.3854
R-квадрат	0.440922	Середнє зал. Змінної		-0.066250
Скорегований R-квадрат	0.367709	Станд. відхил. зал. змінної		1.342254
Станд. похибка регресії	1.067316	Інформ. крит. Акайке		3.084639
Сума квадратів похибок	95.68969	Критерій Шварца		3.405183
Логарифм правдоподібн.	-136.0627	Статистика Дарбіна-Уотсона		1.965193

Три вибрані статистичні характеристики адекватності цієї моделі:

$$R^2 = 0,440; \quad J = СКП = 95,69; \quad DW = 1,965.$$

Характеристики якості однокрокового прогнозу:

$$CeKP = 1,346; \quad CAП = 1,02; \quad CAПП = 109,09; \quad U = 0,848.$$

Спостерігається погіршення характеристик прогнозу, особливо погіршилось значення $CAПП$ і коефіцієнта Тейла: від 0,662 для моделі AP(1) до 0,848 для моделі AP(13). За цим параметром модель є непридатною для прогнозування. Коефіцієнти C(6) – C(10) є незначущими, а тому їх можна видалити з моделі.

Після вилучення цих коефіцієнтів із моделі та відповідних їм складових процесу отримаємо модель, параметри якої наведені в таблиці 2.8.

Таблиця 2.8. Результати оцінювання моделі AP(13) для відхилень

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1997:02 2005:01

Всього використано спостережень після корегування крайніх значень: 96

Модель: $IV = C(1) + C(2) \cdot IV(-1) + C(3) \cdot IV(-2) + C(4) \cdot IV(-3) + C(5) \cdot IV(-5) + C(6) \cdot IV(-12) + C(7) \cdot IV(-13)$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	-0.046129	0.107257	-0.430077	0.6682
C(2)	0.700062	0.103640	6.754728	0.0000
C(3)	-0.256660	0.120078	-2.137440	0.0353
C(4)	0.195599	0.101703	1.923247	0.0576
C(5)	-0.081877	0.083177	-0.984368	0.3276
C(6)	0.190417	0.089745	2.121760	0.0366
C(7)	-0.067455	0.081058	-0.832180	0.4075
R-квадрат	0.438765	Середнє залеж. змінної		-0.066250
Скорегований R-квадрат	0.400929	Станд. відхил. залеж. змінної		1.342254
Станд. похибка регресії	1.038900	Інформ. критерій Акайке		2.984323
Сума квадратів похибок	96.05881	Критерій Шварца		3.171307
Логарифм правдоподібн.	-136.2475	Статистика Дарбіна-Уотсона		1.959533

Вибрані статистичні характеристики адекватності цієї моделі:

$$R^2 = 0,439; \quad J = СКП = 96,05; \quad DW = 1,96.$$

Характеристики якості однокрокового прогнозу:

$$СеКП = 1,346; \quad САП = 1,017; \quad САПП = 108,2; \quad U = 0,854.$$

Таким чином, після видалення з моделі незначимих коефіцієнтів її характеристики залишились практично незмінними. В цілому можна зробити висновок, що побудовані авторегресійні моделі для відхилень ІСЦ від середнього не придатні для прогнозування. Це свідчить про те, що побудувати модель для процесу, який має різні знаки вимірів в різні моменти часу (тобто розвиток відбувається в двох квадрантах), складніше ніж для процесу, який змінюється в межах одного квадранту.

6.8.3. Авторегресія з ковзним середнім для ІСЦ

Розглянемо можливість описання ІСЦ за допомогою моделі АРКС. Характеристики моделі АРКС(1,1) наведені в таблиці 2.9.

Таблиця 2.9. Результати оцінювання моделі АРКС(1, 1) для ІСЦ

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1996:02 2005:01

Всього використано спостережень після корегування крайніх значень: 108

Змінна	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C	100.6529	0.263645	381.7742	0.0000
AR(1)	0.546826	0.093191	5.867793	0.0000
KC(1)	0.060696	0.133153	0.455838	0.6494
R-квадрат	0.415932	Середнє залеж. змінної		100.7407
Скорегований R-квадрат	0.404807	Станд. відхил. залеж. змінної		1.506189
Станд. похибка регресії	1.162006	Інформ. критерій Акайке		3.165558
Сума квадратів похибок	141.7772	Критерій Шварца		3.240062
Логарифм правдоподібн.	-167.9401	F-статистика		37.38676
Статист. Дарбіна-Уотсона	1.996202	Ймовірність (F-стат.)		0.000000
Інвертовані АР корені	.55			
Інвертовані КС корені	-.06			

Вибрані статистичні характеристики адекватності цієї моделі:

$$R^2 = 0,416; \quad J = СКП = 141,78; \quad DW = 1,996.$$

Характеристики якості однокрокового прогнозу:

$CeKP = 1,362$; $CAП = 1,016$; $CAПП = 1,005$; $U = 0,0067$.

Моделі $AP(1)$ і $APKC(1,1)$ мають практично однакові характеристики адекватності та якості однокрокового прогнозу, а тому перевагу (при виборі з цих двох моделей) можна надати моделі $AP(1)$, яка є простішою. Нижче наведена порівняльна таблиця для всіх побудованих моделей.

6.8.4. Врахування впливу на ІСЦ агрегату МЗ

Коефіцієнт кореляції між ІСЦ та агрегатом МЗ від'ємний: $-0,249$, тобто формальний взаємозв'язок між цими змінними незначний, але цікаво розглянути вплив МЗ на ІСЦ за допомогою моделі. Врахування регресора може покращити деякі характеристики моделі, а також врахувати причинний зв'язок між вибраними змінними. Характеристики змішаної моделі: авторегресія $AP(1)$ + парна регресія наведені в таблиці 2.10.

Таблиця 2.10. Результати оцінювання моделі ІСЦ: $AP(1)$ + регресор МЗ

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1996:02 2005:01

Всього використано спостережень після корегування крайніх значень: 108

Модель: $I = C(1) + C(2) * I(-1) + C(3) * M$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	44.97869	6.920057	6.499757	0.0000
C(2)	0.554264	0.068261	8.119764	0.0000
C(3)	-2.82E-06	3.30E-06	-0.854463	0.3948
R-квадрат	0.419106	Середнє залеж. змінної		100.7407
Скорегований R-квадрат	0.408041	Станд. відхил. залеж. змінної		1.506189
Станд. похибка регресії	1.158844	Інформ. критерій Акайке		3.160108
Сума квадратів похибок	141.0066	Критерій Шварца		3.234612
Логарифм правдоподібн.	-167.6458	Стат. Дарбіна-Уотсона		1.919823

Спостерігається незначне покращення характеристик моделі та якості прогнозу у порівнянні з $AP(1)$, але коефіцієнт при МЗ дуже малий ($-2,82E-06$). Однак, формально він є значимим. Таким чином, на розглянутому часовому інтервалі вплив агрегату МЗ на індекс споживчих цін незначний і ним можна знехтувати.

Характеристики прогнозу:

$CeKP = 1,340$; $CAП = 1,004$; $CAПП = 0,994$; $U = 0,0066$.

Можна припустити, що в обороті було недостатньо грошової маси для того, щоб її вплив на ІСЦ був істотним. З іншого боку, недостатній об'єм грошової маси в національній валюті компенсувався (і продовжує

компенсуватись) “твердою” іноземною валютою, зокрема доларами США і, в деякій мірі, євро. Таким чином, слабка українська економіка позитивно впливає на курс долара США завдяки фактичному введенню його в частковий оборот на відносно великій території України. Встановити фактичне співвідношення між об’ємами національної та іноземної валют в обороті можливо, але для цього необхідно отримати додаткові статистичні дані і виконати спеціальне дослідження. Зокрема, необхідно мати дані щодо об’ємів реалізації торговельних операцій у валюті підприємствами всіх форм власності. Очевидно, що отримані такі дані для тіньового обороту не просто і пов’язане з деяким ризиком.

6.9 Визначення впливу ВВП на ІСЦ

Як було показано на початку цього параграфа, коефіцієнт кореляції між ІСЦ та ВВП складає $-0,304$, тобто формально це невелике значення. Також логічно припустити, що зростання ВВП має приводити до зменшення ІСЦ (про це свідчить також знак коефіцієнта кореляції між цими змінними).

Характеристики змішаної моделі: авторегресія $AR(1)$ + парна регресія наведені в таблиці 2.11.

Таблиця 2.11. Результати оцінювання змішаної регресії для ІСЦ і ВВП

Метод оцінювання: метод найменших квадратів

Скорегована часова вибірка даних: 1996:02 2005:01

Всього використано спостережень після корегування крайніх значень: 108

Модель: $I = C(1) + C(2) \cdot I(-1) + C(3) \cdot V$

	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	45.53125	7.156626	6.362112	0.0000
C(2)	0.549611	0.070124	7.837741	0.0000
C(3)	-1.37E-05	1.64E-05	-0.833627	0.4064
R-квадрат	0.418913	Середнє залеж. змінної		100.7407
Скорегований R-квадрат	0.407844	Станд. відхил. залеж. змінної		1.506189
Станд. похибка регресії	1.159037	Інформ. критерій Акайке		3.160441
Сума квадратів похибок	141.0536	Критерій Шварца		3.234945
Логарифм правдоподібн.	-167.6638	Стат. Дарбіна-Уотсона		1.916289

Отримано рівняння:

$$isc(k) = 45,53 + 0,55 \, isc(k-1) - (1,37E - 05) \, vvp(k) + e(k)$$

з характеристиками:

$$R^2 = 0,419; \quad J = СКП = 141,05; \quad DW = 1,92.$$

Характеристики якості однокрокового прогнозу:

$$CeKP = 1,335; \quad CАП = 1,004; \quad CАПП = 0,993; \quad U = 0,0066.$$

Коефіцієнт при змінній *ВВП* невеликий і знаходиться на межі статистичної значущості, тобто вплив *ВВП* на *ІСЦ* незначний. Причиною такого незначного впливу може бути некоректний розподіл *ВВП*, який призводить до того, що більшість населення проживає на межі або навіть нижче межі бідності. Одночасно інша (менша) частина привласнює більшість благ і користується ними, але, очевидно, що це користування не приводить до позитивного впливу на *ІСЦ*. Тому і маємо, що *ВВП* зростає начебто значними темпами, а добробут населення сильно від нього відстає. Таким чином, можна зробити висновок, що на часовому періоді, який розглядається в даному прикладі, відбувався несправедливий розподіл суспільних благ, що проявилось, в даному випадку, в незначному впливі *ВВП* на споживчі ціни для більшості населення України. Побудована математична модель є формальним (об'єктивним) підтвердженням відомого факту нерівномірного розподілу матеріальних благ, вироблених в Україні.

В таблиці 2.12 зведені характеристики математичних моделей, побудованих для індексу споживчих цін, і характеристики однокрокових прогнозів, обчислених на основі цих моделей. Ця таблиця дає можливість оперативно порівняти результати моделювання та прогнозування, отримані за допомогою методики Бокса-Дженкінса, а також встановити можливість практичного використання результатів.

Таблиця 2.12. Результати моделювання і однокрокового прогнозування індексу оптових цін

Тип моделі	Характеристики моделі			Характеристики однокрокового прогнозу			
	R^2	$\sum e^2(k)$	DW	$CeKP$	$CАП$	$CАПП$	Коеф-т Тейла
AP(1)	0,415	141,99	1,931	1,360	1,020	1,008	0,0067
AP(3)	0,317	135,148	1,992	1,360	1,020	1,011	0,0068
AP(7)	0,346	127,244	1,811	1,360	1,012	1,002	0,0067
AP(12)	0,435	97,80	1,941	1,337	1,020	1,013	0,0066
АРКС(1,1)	0,416	141,78	1,996	1,362	1,016	1,005	0,0067
AP(1)+МЗ	0,419	141,007	1,919	1,340	1,004	0,994	0,0066
AP(1)+ВВП	0,419	141,054	1,916	1,335	1,004	0,993	0,0066
Моделі для відхилень ІСЦ від середнього							
AP(1)	0,415	141,99	1,931	1,360	1,020	100,40	0,662
AP(13)	0,440	95,69	1,965	1,346	1,020	109,09	0,848

Результати моделювання, наведені в таблиці 2.12, свідчать про те, що практично всі моделі, побудовані для ІСЦ (перші сім моделей), є придатними для прогнозування, оскільки коефіцієнт Тейла вимірюється тисячними

долями. Характеристики однокрокового прогнозу для моделей різного порядку відрізняються несуттєво. Найкращі характеристики щодо прогнозування має модель $AR(1)+VBP$. Для моделі $AR(12)$ отримано найменше значення суми квадратів похибок моделі, але характеристики прогнозів, отриманих за цією моделлю, не кращі від інших.

Дві моделі, які побудовані для відхилень ІСЦ від середнього, мало придатні для прогнозування. Про це свідчать високі значення коефіцієнта Тейла та середньої абсолютної похибки в процентах. Цей факт можна пояснити тим, що описати за допомогою АРКС процес, динаміка якого спостерігається у двох квадрантах, трудніше, ніж процес, який спостерігається в одному квадранті. Тому для прогнозування відхилень від середнього необхідно знайти іншу структуру математичної моделі, придатну для описання різнознакових величин часового ряду.

6.10. Запитання і вправи до розділу №6

1. Назвіть етапи побудови математичних моделей за методикою Бокса-Дженкінса. Що забезпечує коректне використання цієї методики на практиці?
2. Яка мета поглибленого аналізу функціонування процесу? Які елементи структури моделі ми встановлюємо в процесі аналізу процесу?
3. Яка мета попередньої обробки даних? Назвіть основні операції, які виконують в процесі попередньої обробки даних? До чого призводить визначення значень змінних у великому діапазоні?
4. Які два основних типи нелінійностей зустрічаються в аналізі часових рядів? Поясніть на прикладах.
5. Яким чином можна встановити наявність нелінійностей у процесі?
6. Який метод дає можливість автоматизувати процес визначення та врахування нелінійностей процесу?
7. Яку інформацію можна отримати на основі візуального аналізу даних? Як можна нею скористатись?
8. Яким чином можна знайти оцінку порядку авторегресійної частини моделі?
9. В чому полягає відмінність між автокореляційною та частковою автокореляційною функціями процесу? Чи існує необхідність розрахунку обох функцій в процесі аналізу даних?
10. На чому ґрунтується відбір незалежних змінних (регресорів, екзогенних змінних) для включення в праву частину математичної моделі?
11. Назвіть три умови коректного застосування методу найменших квадратів до оцінювання параметрів математичної моделі?
12. Для чого призначена статистика Льюнга-Бокса і як вона обчислюється? Скористайтесь наявним пакетом програм для

статистичного аналізу даних для обчислення цього статистичного параметру і поясніть його значущість (чи незначущість) на прикладі.

13. В чому полягає принципова різниця між методом найменших квадратів (МНК), призначеним для оцінювання лінійних моделей, та МНК для оцінювання моделей, нелінійних відносно параметрів? Які критерії якості мінімізують ці методи?
14. Виведіть формулу МНК для поліноміальної моделі?
15. Сформууйте матрицю вимірів для математичної моделі вигляду:
$$y(k) = a_0 + a_1 y(k-1) + b_1 x_1(k) + b_2 x_2(k) + \varepsilon(k).$$
16. Про що свідчить корельованість похибок моделі між собою? За допомогою якого статистичного параметра якості моделі можна визначити ступінь корельованості похибок? Яке значення приймає ця статистика в ідеальному випадку?
17. Що означає значущість оцінки параметра (коефіцієнта) моделі в статистичному сенсі? За допомогою якої статистики можна встановити значущість оцінки параметра моделі?
18. Що означають помилки першого і другого роду при перевірці статистичних гіпотез?
19. Поясніть фізичну суть коефіцієнта (множинної) детермінації, яке його ідеальне значення? Для чого корегують R^2 із врахуванням числа ступенів свободи?
20. В чому полягає різниця між критеріями Акайке та Байєса-Шварца? Чи можна обмежитись використанням тільки одного з цих критеріїв?
21. Яка мета розрахунку статистики Фішера? Наведіть послідовність розрахунку та аналізу цієї статистики.
22. Сформууйте правила перевірки гіпотез для t – статистики Стюдента і F – статистики Фішера. Яким чином можна визначити критичні (порогові) значення цих статистик?
23. Поясніть фізичну суть умови достатнього збудження процесу?
24. Яким чином досягають достатнього збудження процесів (об'єктів) на практиці? Які сигнали використовують для досягнення цієї мети? Які властивості цих сигналів використовують в даному випадку?
25. Нарисуйте наближений графік перетворення Фур'є для білого шуму.
26. Які інші ортогональні перетворення даних ви знаєте крім перетворення Фур'є? Чому на практиці використовують декілька різних ортогональних перетворень, а не одне?

ДОДАТКИ

Додаток Д.1

Приклади задач експрес-тестування та модульної контрольної роботи

Головна проблема проведення експрес-тестувань без комп'ютера – неможливість виконати обчислення за 10-15 хвилин. На експрес тестуваннях (без комп'ютера) потрібно зменшити до мінімуму обчислення, що дає можливість перевіряти рівень володіння суттєвими поняттями і алгоритмами.

Задачі 1 - 12.

Преамбулою до кожної з задач 1-12 є вже обчислені дані.

Нехай в результаті попередньої обробки даних по $n=10$ незалежних вимірах пари економічних показників Y_i, X_i , які задовольняють умовам побудови звичайної лінійної регресії $Y = a_0 + a_1X + u$, вже обчислені суми:*

$$\begin{aligned} \sum_{i=1}^n Y_i &= 542; \quad \sum_{i=1}^n X_i = 1008; \quad \sum_{i=1}^n X_i^2 = 104\,784; \quad \sum_{i=1}^n X_i Y_i = 56221; \\ \sum_{i=1}^n (X_i - \bar{X})^2 &= 3376; \quad \sum_{i=1}^n ((X_i - \bar{X})(Y_i - \bar{Y})) = 1587,5; \\ \sum_{i=1}^n (Y_i - \bar{Y})^2 &= 819,6; \quad \sum_{i=1}^n u_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = 26,6. \end{aligned} \quad (1)$$

За даними (1) розв'язати задачі 1- 12 :

1. Знайти оцінку математичного сподівання MY .
2. Знайти варіацію X ($\text{var}(X)$).
3. Знайти варіацію Y ($\text{var}(Y)$).
4. Знайти стандартний ухил (Standard deviation) X ($\text{stdev}(X)$).
5. Знайти стандартний ухил Y (Standard deviation, ($\text{stdev}(Y)$)).
6. Знайти коваріацію (covariance) X та Y ($\text{covar}(X, Y)$).
7. Знайти коефіцієнт кореляції (correlation coefficient) X, Y ($\text{corr}(X, Y)$).

8. Записати систему нормальних рівнянь 1МНК та знайти оцінки коефіцієнтів лінійної регресії $\hat{a}_0, \hat{a}_1$
9. Знайти оцінки коефіцієнтів лінійної регресії $\hat{a}_0, \hat{a}_1$ альтернативним способом (використовуючи властивості середніх ухилів).
10. Підрахуйте коефіцієнт детермінації R^2 .
11. Яким буде граничне збільшення показника Y при збільшенні X на одиницю?
12. Визначити коефіцієнт еластичності $E_{\frac{Y}{X}}$ показника Y щодо X ?

* Прийнятий шлях побудови задач 1-12 орієнтований на експрес перевірку знань взаємозв'язків між базовими поняттями без використання комп'ютера. Наявність деяких даних надлишкова.

Дайте коротке обґрунтування та відповідь в задачах 13-15 :

13. Нехай параметр моделі A оцінюється двома методами:

- 1) $\hat{A}$ - незміщена оцінка A з дисперсією $\sigma_{\hat{A}}^2$
 - 2) $\tilde{A}$ - незміщена оцінка A з дисперсією $\sigma_{\tilde{A}}^2$
- Яка з оцінок ефективніша, якщо $\sigma_{\hat{A}}^2 > \sigma_{\tilde{A}}^2$?

У відповіді обрати потрібне. Необґрунтовані відповіді не зараховуються.

14. На рис. 1 зображено довірчий коридор, побудований за t-розподілом. Видно, що виміри виходять за межі довірчого коридору.

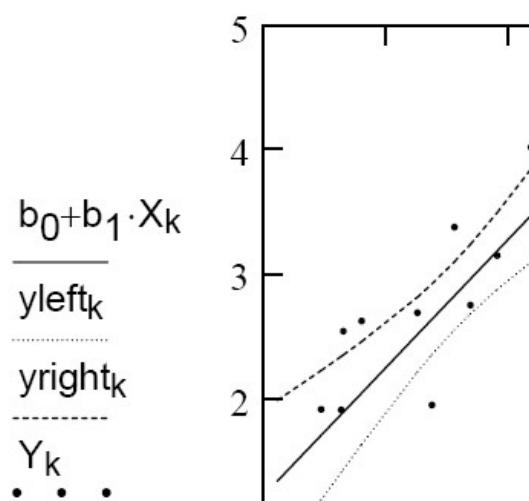


Рис.1.

Чи означає це, що наша обробка вимірів обов'язково неправильна?

Коротко поясніть та дайте відповідь ТАК або НІ. (Необґрунтовані відповіді не зараховуються).

15. Теорія порівняння різних груп даних була побудована Фішером.

Про які дві групи даних ми говоримо, коли мова йдеться про використання розподілу Фішера в звичайній лінійній регресії (для побудови довірчої області):

- а) дві групи вимірів, відмінність яких стала очевидною в результаті деякого аналізу;
- б) групу реальних вимірів і групу модельно побудованих даних ?
(обґрунтовано оберіть потрібне).

16. Годинникар записав в таблицю показання 500 зламаних годинників.

k - номер ($k=1,2,\dots,12$) проміжків шкали циферблата годинника від $k-1$ до k години, в яких відбулись зупинки годинників, а n_k -кількість годинників, покази яких належать k -му проміжку. Наводимо таблицю:

k	1	2	3	4	5	6	7	8	9	10	11	12	всього
n_k	41	34	54	39	49	45	41	33	37	41	47	39	500

Чи вірна гіпотеза, зроблена за вказаними спостереженнями, про те, що покази годинників рівномірно розподілені в інтервалі $(0, 12)$ з рівнем значущості $\alpha = 0.05$?

Обґрунтувати відповідь, використавши критерій Пірсона (χ^2) перевірки гіпотези про вказаний вид розподілу на 12 інтервалах: $\mu < \delta$, де

$$\mu = \sum_{k=1}^{12} \frac{(nP_k - n_k)^2}{nP_k}, \text{ а } \delta = 19.7 \text{ - взята з таблиці для } \chi^2\text{-розподілу}$$

величина за рівнем значущості $\alpha = 0.05$ і 11 ступенями вільності.

(У відповіді вказати „Так”, або „Ні”).

17. Відомо, що перевірка на наявність автокореляції за критерієм Дарбіна – Уотсона полягає в підрахунку статистики

$$DW = \frac{\sum_{k=2}^n (U_k - U_{k-1})^2}{\sum_{k=1}^n (U_k)^2}, \text{ де } U_i = Y_i - r(X_i), \quad i = 1..n, \quad r(X_i) \text{ - функція}$$

кореляції та порівнянні з табличними значеннями d_L , d_U , які залежать від кількості спостережень n .

Висновки про автокореляцію формалізує вказана нижче порівняльна таблиця перевірки гіпотези **Н₀** про відсутність автокореляції.

В третю колонку таблиці впишіть рішення:

- 1 -гіпотеза H_0 (про відсутність автокореляції) відхиляється;
- 2 - має місце невизначеність;
- 3 - гіпотеза H_0 приймається

У відповіді вкажіть послідовно шість обраних цифр без розділових знаків.

19.В результаті обрахунку методом множинної лінійної регресії на комп'ютері залежності $Y = f(X_1, \dots, X_m, u)$, де $m=2$ - кількість незалежних факторів,

вже обраховані матриці:

Таблиця Прийняття чи відхилення гіпотези H_0 про відсутність автокореляції		
N пп	Інтервали	Рішення про прийняття чи відхилення
1	2	3
1	$4 - d_L < DW < 4$	
2	$4 - d_U < DW < 4 - d_L$	
3	$2 < DW < 4 - d_U$	
4	$d_U < DW < 2$	
5	$d_L < DW < d_U$	
6	$0 < DW < d_L$	

$$\left(X^T X \right)^{-1} = \begin{pmatrix} 0.308 & -4.419 \times 10^{-5} \\ -4.419 \times 10^{-5} & 3.551 \times 10^{-5} \end{pmatrix}, \quad X^T Y = \begin{pmatrix} 1.464 \\ 4.987 \end{pmatrix}.$$

Знайдіть з точністю до 0.01 другий елемент вектора шуканих коефіцієнтів множинної лінійної регресії.

Д.2. Державна служба статистики України

<http://www.ukrstat.gov.ua/>

<http://www.me.gov.ua>

Д.3. Національні статистичні служби Європи як першоджерела реальних даних для засвоєння і використання вивчених методів прикладної статистики

Австрія: Statistics Austria <http://www.statistik.at/>

Албанія: Institute of Statistics <http://www.instat.gov.al/>

Андора: Departament d'Estadística <http://www.estadistica.ad/>

Білорусь: Ministry of Statistics and Analysis <http://belstat.gov.by/>

Бельгія: Statistics Belgium <http://statbel.fgov.be/>

Боснія та Герцеговина: Agency for Statistics <http://www.bhas.ba/>

Болгарія: National Statistical Institute <http://www.nsi.bg/>

Німеччина: Federal Statistical Office <http://www.nsi.bg/>

Гібралтар: Statistics Office <http://www.gibraltar.gov.gi/statistics>

Гренландія: Statistics Greenland <http://www.stat.gl/>

Греція: National Statistical Service <http://www.statistics.gr/>

Данія: Statistics Denmark <http://www.dst.dk/>

Естонія: Statistics Estonia <http://www.stat.ee/>

Ірландія: Central Statistics Office <http://www.cso.ie/>

Ісландія: Statistics Iceland <http://www.hagstofa.is/>

Іспанія: Instituto Nacional de Estadística <http://www.ine.es/>

Італія: National Statistical Institute <http://www.istat.it/>

Кіпр: Statistical

Service http://www.mof.gov.cy/mof/cystat/statistics.nsf/index_en/index_en?OpenDocument

Косово: Statistical Office <http://www.ks-gov.net/ESK/eng>

Латвія: Central Statistical Bureau <http://www.csb.lv/>

Литва: Statistics Lithuania <http://www.stat.gov.lt/>

Ліхтенштейн: Amt für Statistik <http://www.as.llv.li/>

Люксембург: Service Central de la Statistique et des Études Économiques <http://www.statec.public.lu/>

Македонія, Колишня югославська республіка: State Statistical Office <http://www.stat.gov.mk/>

Мальта: National Statistics Office <http://www.nso.gov.mt/>

Молдова: National Bureau of Statistics <http://www.statistica.md/>

Нідерланди: Statistics Netherlands <http://www.cbs.nl/>

Норвегія: Statistics Norway <http://www.ssb.no/>

Об'єднане королівство:

Office for National Statistics <http://www.statistics.gov.uk/default.asp>

UK Statistics Authority <http://www.statisticsauthority.gov.uk/>

Польща: Central Statistical Office <http://www.stat.gov.pl/>

Португалія: Statistics Portugal <http://www.ine.pt/>

Румунія: National Institute of Statistics <http://www.insse.ro/>

Російська федерація: ----

Сан Маріно: Ufficio Programmazione Economica, Centro Elaborazione Dati e Statistica <http://www.statistica.sm/>

Словакія: Statistical Office <http://portal.statistics.sk/>

Словенія: Statistical Office <http://www.stat.si/>

Сербія: Statistical Office <http://webrzs.statserb.sr.gov.yu/axd/en/index.php>

Турція: Turkish Statistical Institute <http://www.turkstat.gov.tr/>

Україна: State Statistics Committee <http://www.ukrstat.gov.ua/>

Угорщина: Central Statistical Office<http://portal.ksh.hu/>

Фінляндія: Statistics Finland<http://www.stat.fi/>

Франція: National Institute of Statistics and Economic Studies<http://www.insee.fr/>

Хорватія: Central Bureau of Statistics<http://www.dzs.hr/>

Чорногорія: Statistical Office<http://www.monstat.org/>

Чеська республіка: Czech Statistical Office<http://www.czso.cz/>

Швейцарія: Swiss Federal Statistical Office<http://www.bfs.admin.ch/>

Швеція: Statistics Sweden<http://www.scb.se/>

Д.4. Програми і середовища прикладної статистики

[Amos](http://www.spss.com/amos/) <http://www.spss.com/amos/>

[AREMOS](http://www.globalinsight.com/ProductsServices/ProductDetail1015.htm) <http://www.globalinsight.com/ProductsServices/ProductDetail1015.htm>

[Autobox](http://www.autobox.com/) <http://www.autobox.com/>

[BMDP](http://www.statsol.ie/html/bmdp/bmdp_home.html) http://www.statsol.ie/html/bmdp/bmdp_home.html

[DataDesk](http://www.geoafrica.co.za/reddog/datadesk/datadesk.htm) <http://www.geoafrica.co.za/reddog/datadesk/datadesk.htm>

[Dataplore](http://www.ixellence.com/dataplore) <http://www.ixellence.com/dataplore>

[EasyReg](http://econ.la.psu.edu/~hbierens/EASYREG.HTM) <http://econ.la.psu.edu/~hbierens/EASYREG.HTM>

[EasyPlot](http://www.adeptsience.co.uk/products/dataanal/easyplot/) <http://www.adeptsience.co.uk/products/dataanal/easyplot/>

[ECTM](http://russell-davidson.arts.mcgill.ca/ects4/) <http://russell-davidson.arts.mcgill.ca/ects4/>

[Egwald](http://www.egwald.com/statistics/statistics.php3) [Statistics](http://www.egwald.com/statistics/statistics.php3) - [Multiple](http://www.egwald.com/statistics/statistics.php3) [Regression](http://www.egwald.com/statistics/statistics.php3)

[Epicure](http://www.hirosoft.com/) <http://www.hirosoft.com/>

[EQS - Structural Equation Modeling Software](http://www.mvsoft.com/) <http://www.mvsoft.com/>

[EViews](http://www.eviews.com/) <http://www.eviews.com/>

[Fair-Parke program](http://fairmodel.econ.yale.edu/fp/fp.htm) <http://fairmodel.econ.yale.edu/fp/fp.htm>

[First Bayes](http://www.tonyohagan.co.uk/1b/) <http://www.tonyohagan.co.uk/1b/>

[Frontier](http://www.uq.edu.au/economics/cepa/frontier.htm) <http://www.uq.edu.au/economics/cepa/frontier.htm>

[GAMS](http://www.gams.com/) <http://www.gams.com/>

[Gauss](http://www.aptech.com/) <http://www.aptech.com/>

[Gaussx](http://www.econotron.com/gaussx/gaussx.html) <http://www.econotron.com/gaussx/gaussx.html>

[GB-Stat](http://www.gbstat.com/) <http://www.gbstat.com/>

[G*Power](http://www.psych.uni-duesseldorf.de/aap/projects/gpower/) <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>

[GQOPT](http://www.quandt.com/req.html) <http://www.quandt.com/req.html>

[Gretl](http://gretl.sourceforge.net/) <http://gretl.sourceforge.net/>

[Grocer](http://dubois.ensae.net/grocer.html) <http://dubois.ensae.net/grocer.html>

[JMP](http://www.jmp.com/) <http://www.jmp.com/>

[LIMDEP](http://www.limdep.com/) <http://www.limdep.com/>

[LISREL](http://www.ssicentral.com/lisrel/mainlis.htm) <http://www.ssicentral.com/lisrel/mainlis.htm>
[MacAnova](http://www.stat.umn.edu/macanova/) <http://www.stat.umn.edu/macanova/>
[Macsyma](http://www.symbolics.com/) <http://www.symbolics.com/>
[Maple](http://www.maplesoft.com/) <http://www.maplesoft.com/>
[Mathematica](http://www.wolfram.com/) <http://www.wolfram.com/>
[MathCad](http://www.mathsoft.com/) <http://www.mathsoft.com/>
[Matlab](http://www.mathworks.com/) <http://www.mathworks.com/>
[Matrixer](http://matrixer.narod.ru/) <http://matrixer.narod.ru/>
[Maxima](http://maxima.sourceforge.net/) <http://maxima.sourceforge.net/>
[Microfit](http://www.econ.cam.ac.uk/microfit/) <http://www.econ.cam.ac.uk/microfit/>
[Minitab](http://www.minitab.com/) <http://www.minitab.com/>
[MLEQuick](http://www.magma.ca/~cahill/) <http://www.magma.ca/~cahill/>
[Mx](http://views.vcu.edu/mx) <http://views.vcu.edu/mx>
[NAG](http://www.nag.com/) <http://www.nag.com/>
[NLOGIT](http://www.limdep.com/) <http://www.limdep.com/>
[O-Matrix](http://www.omatrix.com/) <http://www.omatrix.com/>
[PcGive](http://www.pcgive.com/index.html) <http://www.pcgive.com/index.html>
[R](http://www.r-project.org/) <http://www.r-project.org/>
[RATS and CATS](http://www.estima.com/) <http://www.estima.com/>
[Resampling Stats](http://www.resample.com/content/software/download.shtml) <http://www.resample.com/content/software/download.shtml>
[S Plus](http://www.insightful.com/products/splus/default.asp) <http://www.insightful.com/products/splus/default.asp>
[SAS](http://www.sas.com/) <http://www.sas.com/>
[Shazam](http://shazam.econ.ubc.ca/) <http://shazam.econ.ubc.ca/>
[Soritec](http://www.fisisoft.com/) <http://www.fisisoft.com/>
[Spatial Statistics Software, Spatial Data, and Articles](http://www.spatial-statistics.com/) <http://www.spatial-statistics.com/>
[STAMP](http://stamp-software.com/) <http://stamp-software.com/>
[StatsDirect](http://www.statsdirect.com/info.htm) <http://www.statsdirect.com/info.htm>
[SPSS](http://www.spss.com/) <http://www.spss.com/>
[SST](http://elsa.berkeley.edu/sst/) <http://elsa.berkeley.edu/sst/>
[Stata](http://www.stata.com/) <http://www.stata.com/>
[StatTransfer](http://www.stattransfer.com/) <http://www.stattransfer.com/>
[StatView](http://www.statview.com/index.html) <http://www.statview.com/index.html>
[Stibox](http://www.maths.lth.se/matstat/stibox/) <http://www.maths.lth.se/matstat/stibox/>
[SYSTAT](http://www.systat.com/) <http://www.systat.com/>
[Time Series Modelling \(TSM\)](http://www.timeseriesmodelling.com/) <http://www.timeseriesmodelling.com/>
[TSP International](http://www.tspintl.com/) <http://www.tspintl.com/>
[WinSolve](http://www.econ.surrey.ac.uk/winsolve/) <http://www.econ.surrey.ac.uk/winsolve/>
[Xlisp-Stat](http://lib.stat.cmu.edu/xlispstat/) <http://lib.stat.cmu.edu/xlispstat/>
[XploRe](http://www.xplore-stat.de/index_js.html) http://www.xplore-stat.de/index_js.html
[XTREMES](http://www.xtremes.de/xtremes/xtremes/index_xtremes.htm) http://www.xtremes.de/xtremes/xtremes/index_xtremes.htm

[Yorick](http://yorick.sourceforge.net/index.php) <http://yorick.sourceforge.net/index.php>

Д.5. Деякі джерела загальнодоступних вихідних кодів програм

[Econometrics Laboratory Software Archive](http://elsa.berkeley.edu/) <http://elsa.berkeley.edu/>

[Free C, C++ for Numerical Computation](ftp://ftp.math.psu.edu/pub/FAQ/numcomp-free-c)
<ftp://ftp.math.psu.edu/pub/FAQ/numcomp-free-c>

[Guide to Available Mathematical Software](http://gams.nist.gov/) <http://gams.nist.gov/>

[GNU gcc compiler for C and C++LCC C compiler](http://gcc.gnu.org/) <http://gcc.gnu.org/>

[GnuplotMathtools.Net](http://www.mathtools.net/) <http://www.mathtools.net/>

[MLE++](http://www.magma.ca/~cahill/) <http://www.magma.ca/~cahill/>

[Software for Graphics and Data AnalysisNAG](http://stommel.tamu.edu/~baum/ocean_graphics.html)
http://stommel.tamu.edu/~baum/ocean_graphics.html

[Netlib](http://netlib2.cs.utk.edu/) <http://netlib2.cs.utk.edu/>

[NetlibNumerical Recipes](http://www.nr.com/) <http://www.nr.com/>

[Numerical Recipes Books online](http://library.lanl.gov/numerical/index.html) <http://library.lanl.gov/numerical/index.html>

[Numerical Recipes in C](http://library.lanl.gov/numerical/bookcpdf.html) <http://library.lanl.gov/numerical/bookcpdf.html>

[Numerical Recipes in Fortran 77](http://library.lanl.gov/numerical/bookfpdf.html)
<http://library.lanl.gov/numerical/bookfpdf.html>

[Numerical Recipes in Fortran 90](http://library.lanl.gov/numerical/bookf90pdf.html)
<http://library.lanl.gov/numerical/bookf90pdf.html>

[QM&RBC \(Quantitative Macroeconomic and Real Business Cycle\)](http://dge.repec.org/codes.html)

[CodesStatCodes](http://dge.repec.org/codes.html) <http://dge.repec.org/codes.html>

[Statlib](http://lib.stat.cmu.edu/) <http://lib.stat.cmu.edu/>

[StatMath](http://www.indiana.edu/~statmath) <http://www.indiana.edu/~statmath>

[Symbolic Mathematical Computation Center](http://www.symbolicnet.org/) <http://www.symbolicnet.org/>

[UICSTAT](http://www.uic.edu/depts/adn/cso/docs/stats/intro.html) <http://www.uic.edu/depts/adn/cso/docs/stats/intro.html>

[Linux Resources](http://www.linux.org/) <http://www.linux.org/>Додаток Д.6

Д.6. Англо-український словник з прикладної статистики

aposteriori апостеріорний	canonical канонічний аналіз;
apriori апіорний	cluster кластерний аналіз;
accuracy точність	factor факторний аналіз;
accept приймати	correspondence аналіз відповідностей;
actual дійсний, фактичний	data аналіз даних;
addition додавання	discriminant дискримінантний аналіз;
adaptive адаптивний	dynamic динамічний аналіз;
additive адитивний	multivariate багатовимірний аналіз;
additivity адитивність	spectral спектральний аналіз
adequacy адекватність, відповідність	annual річний, щорічний
of forecast адекватність прогнозування;	anticipate передбачати
model адекватність моделі	antithetic протилежний, додатковий, що доповнює
adjust виправляти, приводити відповідність, вирівнювати	approach підхід;
adjustment коректування, виправлення (див. тж. correction);	Bayesian байєсівський підхід
bias виправлення зсуву	approximation апроксимація, наближення, оцінка;
partial часткове виправлення;	asymptotic асимптотична апроксимація;
instantaneous миттєве виправлення;	stochastic стохастична апроксимація
seasonal виправлення сезонності	area площа
admissible припустимий	arrangement розташування, конфігурація
aggregate сукупний, агрегований сукупна величина, агрегована величина, агрегат	association зв'язок, залежність, асоціація
aggregation агрегування	assumption припущення, допущення;
algebra алгебра;	tentative попереднє, пробне припущення;
matrix матрична алгебра	asymptotic(al) асимптотичний
algorithm алгоритм;	asymptotically асимптотично
Cochran-Orcutt алгоритм Кочрена-Оркатта;	augment збільшувати, додавати
conjugate directions алгоритм спряжених напрямків	augmentation збільшення, приріст
alternative альтернатива альтернативний;	autocorrelation автокореляція,
local локальна альтернатива	partial частинна автокореляція;
analyse(analyze) аналізувати, досліджувати	residual автокореляція залишків;
analysis аналіз;	spatial просторова автокореляція
of covariance коваріаційний аналіз;	autocorrelogram автокорелограма
of variance дисперсійний аналіз;	autocovariance автоковаріація
	autonomous автономний
	autoregression авторегресія
	of n-th order авторегресія n-го порядку;
	order of порядок авторегресії;
	vector векторна авторегресія
	autoregressive авторегресійний

auxiliary допоміжний, додатковий
average середнє, середній;
arithmetic середнє арифметичне;
autoregressive moving ковзне середнє авторегресії;
geometric середнє геометричне
moving ковзне середнє;
simple простої середнє (середнеарифметичне, незважене)
weighted зважене середнє

В

backward зворотний; назад, у зворотному напрямку
badness of fit (=lack of fit, poor fit) поганий підбір, неадекватність
band смуга, смуга частот
error довірчий інтервал похибки;
bandwidth ширина інтервалу, ширина вікна
base базисний, базовий
behavio(u)r поведіння, динаміка (економічного показника)
belief переконання, думка, довіра
prior believes апіорні припущення
best кращий, найкращий, оптимальний
bias зсув, зміщення; систематична помилка; **asymptotic** асимптотичне зміщення; **downward** заниження, зсув униз;
upward завищення, зміщення вгору
biased зміщений;
severely сильно зміщений
binary бінарний, двійковий
binomial біноміальний
block блок
bootstrap бутстреп
bootstrapping проведення бутстрепного аналізу
bound границя, грань;
a priori апіорна границя
confidence довірча границя;
Cramer-Rao границя Крамера-Рао;
lower нижня границя
upper верхня границя
bounded обмежений

Box-Jenkins стосовний до методу Бокса-Дженкінса

break-point точка злому, точка зміни тенденції

С

calculation обчислення, розрахунок

causal причинний

causality причинність;

instantaneous миттєве (одночасне) причинний вплив

Granger причинність по Грейнджеру

causation причинність, заподіяння

centile процентиль

center (амер. ≡ centre)

centre центр || центрувати

centred центрований

chain ланцюг, ланцюжок;

causal причинний ланцюжок;

Markov ланцюг Маркова; марковський ланцюг

change зміна;

structural структурні зміни

chart діаграма, карта, схема;

band стрічкова діаграма;

bar стовпчаста діаграма

checking перевірка;

diagnostic діагностична перевірка .

chi-square хі-квадрат;

choice вибір, вибірка;

binary бінарний вибір;

discrete дискретна вибірка ;

model вибірмоделі;

multiple множинний вибір

circular циклічний, круговий

classification класифікація, угруповання;

closed замкнутий

cluster кластер, скупчення, група

clustering розбивка сукупності на кластери

coefficient коефіцієнт;

of determination коефіцієнт детермінації;

of determination adjusted for degrees of freedom коефіцієнт детермінації

скоректований на кількість ступенів вільності;

of variation коефіцієнт варіації;

adjusted of determination скоректований коефіцієнт детермінації;

correlation коефіцієнт кореляції;

multiple correlation коефіцієнт множинної кореляції;

regression коефіцієнт регресії;

sample correlation вибірковий коефіцієнт кореляції;

skewness коефіцієнт асиметрії;

slope коефіцієнт нахилу;

Spearman's rank correlation коефіцієнт рангової кореляції Спірмена

coherence когерентність, погодженість

cointegrated коінтегрований

cointegration коінтеграція

collectiondata збір даних

collinearity колінійність

combination комбінація, сполучення

cointegrating коінтегруюча комбінація

complementorthogonal ортогональне доповнення

component компонента, складова частина, складова, член

cyclic(al) циклічна складова

harmonic гармонійна складова

orthogonal ортогональні компоненти

principal головні компоненти

smooth гладка складова

compound складний || суміш

concentration концентрація, купчастість

condition умова || обумовлювати, обумовити, розглядати умовний розподіл

equilibrium умова рівноваги

normalizing умова нормалізації, умова нормування

orthogonality умова ортогональності

stability умова стійкості

stationarity умова стаціонарності

conditional умовний

confidence довіра; довірчий

confirmation підтвердження, підкріплення

confounding змішування

congruency відповідність

congruent підходящий, відповідний, що гармонує

conjecture здогад, припущення

conjugate сполучений

consistency слушність, несуперечність

consistent слушний

autocorrelation слушний при автокореляції

heteroscedasticity слушний при гетероскедастичності

constancy незмінність, сталість

parameter сталість параметрів

constant константа, постійна; постійний

constrained обмежений, умовний (про оптимізацію)

constraint обмеження

contingency успряженість ознак, випадковість, випадок

contribution внесок

converge сходиться

convergence збіжність

in distribution збіжність по розподілу;

in law збіжність за законом;

in mean square середньоквадратична збіжність;

in probability збіжність по ймовірності;

almost sure збіжність майже напевно, з імовірністю 1;

convergent збіжний

correction виправлення, виправлення, коректування (Див. тж. **adjustment**)

autocorrelation виправлення на автокореляцію

bias виправлення на зсув

continuity виправлення на безперервність

degrees of freedom виправлення на кількість ступенів вільності

error виправлення помилки

grouping виправлення на угруповання

correlated корельований

serially ~ серіально корельований

correlation кореляція

canonical канонічна кореляція

multiple множинна кореляція.
множинний коефіцієнт кореляції
rank рангова кореляція
simple проста кореляція
serial серіальна кореляція
spurious помилкова кореляція
correlogram коррелограма
partial частинна коррелограма, графік частинної автокорреляційної функції
residual коррелограма залишків
correspondence відповідність, аналогія, співвідношення
covariance коваріація, змішаний другий центральний момент
criterion критерій
Akaike information інформаційний критерій Акаїке
choice критерій вибору
information інформаційний критерій
Schwartz information інформаційний критерій Шварца
critique критика
cross- перехресний, взаємний
cross-classification перехресна класифікація
cross-correlation крос-кореляція, взаємна кореляція
cross-covariance крос-коваріація, перехресна коваріація
cross-elasticity перехресна еластичність
cross-plot $\equiv$ scatter

cross-product взаємодія (парне); попарні добутки
cross-spectrum крос-спектр
cumulative кумулятивний, інтегральний, накопичений
curve крива
logistic логістична крива

regression регресійна крива
smooth гладка крива
cycle цикл
cyclic(al) циклічний, періодичний

D

data дані
count ~ дискретні дані
categorized ~ категоризовані дані
differenced ~ різниці вихідних даних
grouped ~ згруповані дані
input ~ вхідні дані, вихідні дані
integrated ~ інтегровані дані
missing ~ пропущені, відсутні дані
noisy ~ зашумлені дані (виміряні з більшими помилками)
ordered упорядковані дані
original первинні дані
raw неопрацьовані дані, “сирі” дані
ranked упорядковані дані
sample вибіркві дані
dating датування
decile дециль
decomposition декомпозиція, розкладання
Cholesky ~ розклад Холецкого;
orthogonal ортогональне розкладання
spectral спектральне розкладання
square root розкладання квадратного кореня
sum of squares розкладання суми квадратів
definition визначення
deflate виражати в постійних цінах
deflator дефлятор (індекс цін)
degenerate вироджений
degree ступінь, порядок, градус

of accuracy ступінь точності

of belief ступінь упевненості, ступінь довіри

of dependence ступінь залежності

of freedom ступеня вільності, кількість ступенів вільності

deletion видалення

variable видалення змінних

density щільність, щільність розподілу

conditional умовна щільність розподілу, щільність умовного розподілу
marginal маргінальна щільність
normal щільність нормального розподілу
prior апіорна щільність розподілу
posterior апостеріорна щільність розподілу
probability щільність імовірності
spectral спектральна щільність
dependence залежність
linear лінійна залежність
dependent залежний
derivative похідна
numerical чисельна похідна
deseasonalize видаляти сезонність, очищати від сезонної складової
design план (експерименту), конструкція
of experiments планування експериментів
balanced урівноважений план
factorial факторний план
incomplete неповний план
orthogonal ортогональний план
random випадковий план, рандомізований план
detect виявляти
determinant визначник, детермінант
deterministic детермінований
detrending видалення тренда
deviance сума квадратів відхилень вибірових значень від їх середнього
deviation відхилення, ухил;

absolute абсолютне відхилення;

least absolute [method] метод найменших модулів;
standard стандартне відхилення, середньоквадратичне відхилення
diagnostic діагностичний
diagnostics діагностика, діагностична статистика
diagonal діагональний || діагональ;

principal головна діагональ

diagonalization діагоналізація, приведення матриці до діагонального виду

diagram діаграма, графік

scatter діаграма розсіювання

difference різниця, розходження

differentiation диференціювання;

matrix ~ матричне диференціювання;

vector ~ векторне диференціювання

dimension розмірність

disaggregation дезагрегація, розбивка

discard відкидати

discrepancy розбіжність, розходження, нев'язка

discrete дискретний

disequilibrium нерівновага, порушення рівноваги || нерівноважний

dispersion дисперсія, розсіювання

distance відстань;

least [method] метод найменших відстаней

distribution розподіл;

asymptotic асимптотичний розподіл;

bell-shaped дзвоноподібний розподіл

Bernoulli розподіл Бернуллі;

beta бета-розподіл;

bimodal бімодальний розподіл;

bivariate normal двовимірний нормальний розподіл;

chi-square(d) розподіл хі-квадрат;

Cauchy розподіл Коші;

conditional умовний розподіл;

continuous неперервний розподіл;

degenerate вироджений розподіл; **discrete** дискретний розподіл;

empiric(al) емпіричний розподіл;

Erlang розподіл Ерланга;

exponential експонентний розподіл;

F F-розподіл,

Fisher розподіл Фішера;

gamma гамма-розподіл

gaussian гаусівський розподіл;

geometric геометричний розподіл;

hypergeometric гіпергеометричний розподіл;

inverse обернена функція розподілу;
joint спільний розподіл;
Laplace розподіл Лапласа;
large-sample розподіл великої вибірки;
limiting граничний розподіл;
logistic логістичний розподіл;
lognormal логнормальний розподіл;
multimodal багатомодальний розподіл,
багатомодальний розподіл;
multinomial поліноміальний розподіл;
multivariate багатовимірний розподіл;
multivariate normal багатовимірний
нормальний розподіл;
normal нормальний розподіл;
Poisson розподіл Пуассона;
posterior апостеріорний розподіл;
prior апіорний розподіл;
sample вибіркового розподілу, емпіричний
розподіл;
small-sample розподіл малої вибірки;
Student розподіл Стюдента;
symmetric симетричний розподіл;
t t-розподіл;
unconditional безумовний розподіл;
uniform рівномірний розподіл;
unimodal одномодальний розподіл
disturbance збурення;
non-spherical несферичні збурювання
random випадкове збурення;
spherical сферичні збурення
divergence розбіжність

domain область;
of function область визначення функції;
frequency частотна область
downward спадний }
draw тягти, кидати (жереб), генерувати
(випадкові числа);
to sample робити вибірку, генерувати
вибірку
drawing робити вибірку, вибір;
with replacement вибір з поверненням;
upward тенденція росту
dummy[variable] фіктивна змінна;
seasonal сезонна фіктивна змінна
dynamic динамічний

dynamics динаміка;
system системна динаміка, динаміка
системи

Е
econometric економетричний
econometrics економетрика, економетрія
effect ефект, результат, наслідок, вплив
end граничний ефект, кінцевий ефект
fixed фіксовані ефекти (у дисперсійному
аналізі)
lagged запізнілий, відкладений
ефект/вплив
long-run довгостроковий ефект
random випадкові ефекти (у
дисперсійному аналізі)
seasonal сезонний ефект
short-run короткостроковий ефект
efficiency ефективність
asymptotic асимптотична ефективність
asymptotic relative відносна
асимптотична ефективність
efficient ефективний
one-step однокроковий ефективний
eigenvalue власне число, власне значення
eigenvector власний вектор
elasticity еластичність
cross перехресна еластичність
unit одинична еластичність
elimination виключення
ellipsoid еліпсоїд
confidence довірчий еліпсоїд
embed вкладати, занурювати
embedding вкладення
empiric(al) емпіричний, отриманий з
досвіду, **endogeneity** ендегенність
endogenous ендегенний, внутрішньо
системний
equation рівняння
likelihood рівняння правдоподібності
normal нормальні рівняння
recurrence рекурентні рівняння
regression рівняння регресії
simultaneous система одночасних рівнянь
Yule-Walker рівняння Юла-Уолкера
equilibrium рівновага

equivalence еквівалентність, рівносильність
asymptotic асимптотична еквівалентність
observational еквівалентність с точки зору спостережень, нерозрізненість
equivalent еквівалентний, рівносильний || еквівалент
asymptotic(al) асимптотичний еквівалент
asymptotically асимптотично еквівалентний
locally локально еквівалентний
ergodic ергодичний
ergodicity ергодичність
erroneous помилковий, помилковий

error помилка **due to ...** помилка, обумовлена ...
in variables помилки в змінних
autoregressive авторегресійні помилки
asymptotic standard асимптотична стандартна помилка
coefficient standard стандартна помилка коефіцієнта
forecast помилка прогнозу
mean-root-square середньоквадратична похибка
mean-square середньоквадратична помилка, **measurement** помилка виміру
observation(al) помилка спостереження
predictive (prediction) помилка прогнозування
random випадкова помилка
rounding помилка округлення
sample вибіркова помилка
specification помилка специфікації
standard стандартна помилка(, середньоквадратична помилка
systematic систематична помилка
type I помилка першого роду
type II помилка другого роду
white noise білий шум
establish установлювати
established установлений
estimable що може бути оцінений, що піддається оцінці

estimand оцінювана величина
estimate оцінка, оцінювати, **to estimate**
recursively оцінювати рекурентно
a priori апіорна оцінка
approximate наближена оцінка
biased зміщена оцінка
Bayes байесівська оцінка
exact maximum likelihood оцінка безпосередньо методом максимальної правдоподібності
maximum likelihood оцінка максимальної правдоподібності
non-parametric непараметрична оцінка
point точкова оцінка
regression регресійна оцінка
sample вибіркова оцінка
semiparametric напівпараметрична оцінка
subjective суб'єктивна оцінка
unbiased незміщена оцінка
estimated оцінений
expansion розкладання
asymptotic асимптотичне розкладання
Edgeworth розкладання Еджворта
serial розкладання в ряд
stochastic стохастичне розкладання
Taylor розкладання в ряд Тейлора, розвинення в ряд Тейлора
expectation очікування, математичне очікування, перший момент
formation формування очікувань
adaptive адаптивні очікування
conditional умовне [математичне] очікування
data-based очікування, засновані на даних
backward-looking ретроспективні (побудовані на минулому досвіді) очікування
rational раціональні очікування
static статичні очікування
unconditional безумовне [математичне] очікування
unbiased незміщені очікування
expected очікуваний
experiment експеримент, досвід

Monte-Carlo експеримент Монте-Карло
factorial факторний експеримент
exponential експонентний
extent розмір, протяг, ступінь
extrapolation екстраполяція,
екстраполювання
extremum екстремум

F

F-distribution F-розподіл (Фішера)
factor фактор, множник, коефіцієнт
autonomous автономний фактор
common загальний множник
exogenous екзогенний фактор
scale масштабний множник
factorial факторний || факторіал
factoring розкладання на множники,
подання у вигляді добутку
factorization розкладання, розкладання
на множники
failure невдача
predictive неадекватність прогнозу
falsification фальсифікація, спростування
falsify фальсифікувати, спростовувати
family сімейство
of distributions сімейство розподілів
exponential експонентне сімейство
розподілів
feasible припустимий, можливий,
здійснений
feedback зворотний зв'язок
feedforward попередження,
попередження
finite обмежений, кінцевий
filter фільтр || фільтрувати
Kalman фільтр Калмана
linear лінійний фільтр
seasonal сезонний фільтр
filtering фільтрація
finite-lag з кінцевим лагом
first-order першого порядку
fit підбір, відповідність, підбрана
функція || підбирати, підганяти
good точний підбір
poor поганий підбір
fitting підбір (як процес), припасування

fixed постійний, незмінний, фіксований
in repeated samples незмінний у
повторних вибірках
forecast прогноз, пророкування
back зворотний прогноз
conditional умовний прогноз
dynamic динамічний прогноз
multi-step багатокроковий прогноз
one-step однокроковий прогноз
static статичний прогноз
forecasting прогнозування
form вид, форма
functional функціональна форма
quadratic квадратична форма
reduced наведена форма
restricted обмежена форма
structural структурна форма
forward уперед || спрямований уперед
fraction дріб, частка
fractional дробовий
frequency частота (подій), частота
(коливань), періодичність || частотний
data періодичність даних
harmonic гармонійна частота
frontier границя
stochastic стохастична границя
function функція
autocorrelation автокореляційна
функція
beta бета-функція
characteristic характеристична функція
concentrated likelihood концентрована
функція правдоподібності
cumulative distribution [кумулятивна,
інтегральна] функція розподілу
demand функція попиту
density функція щільності розподілу
discriminant дискримінантна функція
distribution функція розподілу
inverse distribution зворотна функція
розподілу
likelihood функція правдоподібності
logistic логістична функція
loss функція втрат
partialautocorrelation частинна
автокореляційна функція

penalty штрафна функція
power функція потужності
production виробнича функція
spectral density функція спектральної щільності
spectral спектральна функція
spline сплайн-функція
functional

G

gain виграш, збільшення
efficiency виграш в ефективності
gaussian гаусівський
general-to-specific від загального до частинного
generation генерація
generator генератор
global глобальний
goodness точність, узгодженість

of fit точність підбору, відповідність прогнозу моделі даним, узгодженість
grid сітка (на графіку), ґрати
group група
grouping угруповання, групування
groupwise група
guess здогад, припущення

H

Hessian гессіан, матриця Гессе
harmonic гармонійний || гармоніка
hazard ризик, шанс
heteroscedasticity гетероскедастичність
autoregressive conditional авторегресійна умовна гетероскедастичність
conditional умовна гетероскедастичність
multiplicative мультиплікативна гетероскедастичність
high-frequency висока частота || високочастотний
high-order вищого порядку, старшого порядку
histogram гістограма
homogeneous однорідний
homogeneity однорідність
homoroscedasticity гомоскедастичність

horizon обрій
forecasting обрій прогнозування
household домашнє господарство
hull оболонка
convex опукла оболонка
hypothesis гіпотеза
to accept a прийняти гіпотезу,
to reject a відкинути, відхилити гіпотезу,
to test a перевірити гіпотезу
alternative альтернативна гіпотеза;
composite складна гіпотеза;
false помилкова гіпотеза;
implicit неявна гіпотеза, яка мається на увазі;
linear лінійна гіпотеза
non-parametric непараметрична гіпотеза
null нульова гіпотеза
parametric параметрична гіпотеза
simple проста гіпотеза
true вірна гіпотеза

I

identifiability ідентифіковність

asymptotic
асимптотична ідентифіковність
identifiable ідентифіковний
identification ідентифікація
model ідентифікація моделі
identified ідентифікований
exactly точно ідентифікований
poorly погано ідентифікований
identity тотожність
ill-conditioned погано обумовлений
impact імпульс, поштовх, вплив, вплив
implicit неявний, який мається на увазі
imply мати на увазі, значити, містити в собі
inaccuracy неточність, помилка
inadequacy неадекватність, невідповідність
model неадекватність моделі
induced індукований, породжений, викликаний
increase збільшення, приріст || зростати

increment приріст

independence незалежність

of irrelevant alternatives

незалежність від сторонніх альтернатив

asimptotic асимптотична незалежність

conditional умовна незалежність

linear лінійна незалежність

stochastic статистична незалежність

independent незалежний

of ... не залежний від..., незалежний від...

asimptotically асимптотично незалежний

linearly лінійно-незалежні

mutually взаємно незалежні

serially серіально незалежний

index індекс

indicator індикатор, показник

inefficiency неефективність

inefficient неефективний

inelastic нееластичний, негнучкий

inexact неточний

inequality нерівність

Bonferroni нерівність Бонферроні

Chebyshev's нерівність Чебишова

Cramer-Rao нерівність Крамера-Рао

Jensen нерівність Йенсена

inference висновок

Bayesian байєсівський висновок

statistical статистичні висновки,

статистичний висновок

infinite нескінченний, нескінченно великий

influence вплив, вплив

influential впливовий

information інформація

a priori апіорна інформація

extraneous зовнішня інформація

prior апіорна, попередня інформація

sample вибіркова інформація

initial початковий, первісний

innovation інновація

inquiry дослідження, обстеження

insignificant незначущий

insignificant at ... % level незначущий на рівні ...%

instability нестабільність, нестійкість

instantaneous миттєвий, негайний

instrument інструмент

instrumental інструментальний

integral інтеграл, ціле число ||

інтегральний, цілий

probability інтеграл імовірностей

integrated інтегровний

integrated of order n інтегровний n -го порядку

integratedness інтегровність

integration інтегрування

interaction взаємодія

intercept відрізок, що відтинається на координатній осі; константа, вільний член (у рівнянні регресії)

interdependence взаємна залежність

internal внутрішній

interpolation інтерполяція, інтерполяція

interval інтервал, відрізок, проміжок

confidence довірчий інтервал

non-central confidence нецентральний довірчий інтервал

prediction інтервал прогнозу

probability інтервал імовірності

intrinsic внутрішній, властивий

intrinsically внутрішньо

invariance інваріантність

under linear transformations

інваріантність стосовно лінійних перетворень

parameter інваріантність параметрів

invariant інваріантний || інваріант

inverse зворотна величина

generalized узагальнена зворотна матриця

irregularity неправильність, нерегулярність

J

J-test J-критерій

Jacobian якобіан, матриця Якобі

jumpfunction ступінчата функція

joint спільний

K

kernel ядро

kurtic ексцесивний / куртозисний

kurtosis четвертий центральний момент, куртозис

excess ексцес

L

lack недолік, відсутність

of fit поганий підбір (проорокування моделі не відповідають даним, неадекватність)

lag лаг, запізнювання, затримка

adjustment лаг підлаштування

Almon лаг Алмон

autoregressive distributed авторегресійний розподілений лаг

distributed розподілений лаг

exponential експоненціальний лаг

expectation-formation лаг у формуванні очікувань

expectations лаг очікувань

finite кінцевий лаг

geometric геометричний лаг

linear лінійний лаг

mean середній лаг

polynomial поліноміальний лаг

polynomial distributed

поліноміальний розподілений лаг

rational раціональний лаг

lagged запізнілий, з лагом (про змінну)

lagging запізнювання, відставання || запізнілий

lagrangian лагранжиан, функція Лагранжа

large-sample стосовний до великої вибірки

latent прихований, латентний

law закон

of large numbers закон великих чисел

empirical емпіричний закон

least-square(s) найменших квадратів, що має відношення до методу найменших квадратів

leastsquares[method] метод найменших квадратів

autoregressive авторегресійний метод найменших квадратів

generalized узагальнений метод найменших квадратів

indirect непрямий метод найменших квадратів

non-linear нелінійний метод найменших квадратів

ordinary звичайний метод найменших квадратів

recursive рекуррентний метод найменших квадратів

length довжина

lag довжина лага

level рівень

accuracy рівень точності

confidence рівень довіри

conventional significance

загальноприйнятий, традиційний рівень значущості (звичайно 5%)

risk рівень ризику

significance рівень значущості

leverage впливовість, показник впливовості (спостережень)

likelihood правдоподібність

maximum максимальна правдоподібність, метод максимальної правдоподібності

limit межа, границя || обмежувати

confidence довірча границя, довірча межа

lower нижня границя

probability імовірнісна межа

stochastic випадкова межа

tolerance припустимі межі

upper верхня границя

limited обмежений

line лінія, пряма

regression лінія регресії

trend лінія тренда, тренд

linear лінійний

intrinsically внутрішньолінійний
linearity лінійність
linearization лінеаризація
loadingfactor факторне навантаження
local локальний
locally локально
logarithmic логарифмічний
logit логіт
ordered упорядкований логіт
logistic логістичний
loglikelihood[function] логарифмічна функція правдоподібності
loglinear логлінійний
lognormal логнормальний
log-transform логарифм змінної, логарифмічне перетворення
long-run довгостроковий
loss втрата, втрати

of efficiency втрата ефективності, втрата ефективності

of information втрата інформації
minimum expected мінімально очікувані втрати
low-frequency низькочастотний
low-order низького порядку, молодшого порядку

M

macroeconometrics макроеконометрія
magnitude величина, розміри
marginal маргінальний (у статистиці) граничний (в економічній теорії)
martingale мартингал
matrix матриця
asymptotic covariance асимптотична коваріаційна матриця
correlation кореляційна матриця
covariance матриця коваріацій, коваріаційна матриця
block-diagonal блочно-діагональна матриця
diagonal діагональна матриця
dispersion дисперсійна матриця
distance матриця відстаней

Hessian матриця Гессе
idempotent ідемпотентна матриця
identity одинична матриця
impact матриця впливу
information інформаційна матриця
orthogonal ортогональна матриця
orthonormal ортонормована матриця
negative-definite від'ємно визначена матриця
positive-definite додатньо визначена матриця
projection матриця проєкції
pseudoinverse псевдообернена матриця
singular особлива, вироджена матриця
symmetric(al) симетрична матриця, симетрична матриця
transition probability матриця перехідних імовірностей
transpose of транспонована матриця
triangular трикутна матриця
mean середнє, середнє значення, [математичне] очікування
population генеральне середнє
sample вибіркове середнє
value середнє значення
measurability вимірність
measure міра
of ..._{fit} показник точності підбора
 Точність підбора звичайно вимірюють коефіцієнтом детермінації
leverage показник впливовості (спостережень)
rank correlation міра рангової кореляції
measurement вимір
mechanism механізм
causal причинний механізм
error correction коригувальний механізм
median медіана
mesokurtic маючий нормальний/нульовий ексцес, мезокуртичний
method метод, спосіб, прийом

of least squares метод найменших квадратів

of maximum likelihood метод
максимальної правдоподібності

of moments метод моментів
generalized of moments узагальнений
метод моментів

generalized instrumental variables
узагальнений метод інструментальних
змінних

instrumental variables метод
інструментальних змінних

least distance метод найменших відстаней

least squares метод найменших квадратів

Marquardt метод Марквардта

maximum likelihood метод максимальної
правдоподібності

Newton метод Ньютона

Newton-Rafson метод Ньютона-Рафсона

principal components метод головних
компонентів

microeconometrics мікроеконометрія

minimand таке, що мінімізується

misspecification неправильна
специфікація, помилка специфікації

functional form неправильна
специфікація функціональної форми

mixture суміш

of distributions суміш розподілів

mode мода

model модель

autoregression (autoregressive)
авторегресійна модель

**autoregressive (autoregression) moving
average** змішана модель авторегресії й
ковзного середнього

autoregressive distributed lag
авторегресійна модель із розподіленням
лагом

binary choice модель бінарного вибору

Box-Jenkins модель Бокса-Дженкінса

Brookings Брукинские моделі

choice модель вибору

cobweb паутинообразна модель

congruent підходяща модель

distribution-free вільна від розподілу
модель

discrete choice модель дискретного
вибору

dynamic динамічна модель

error correction модель виправлення
помилки

errors-in-variables модель із помилками
в змінних

fixed-effects модель із фіксованими
ефектами

general linear загальна лінійна модель

linear лінійна модель

loglinear логлінійна (лінійна в
логарифмах) модель

multiplicative мультиплікативна модель

moving average модель ковзного
середнього

nested вкладена модель

non-nested не вкладена модель

non-parametric непараметрична модель

parametric параметрична модель

parsimonious ощадлива (з т. зр. кількості
параметрів), проста модель

partial adjustment модель часткового
пристосування

quadratic квадратична модель

qualitative dependent variable модель із
якісної залежної змінної

random coefficients модель із
випадковими коефіцієнтами

random-effects модель із випадковими
ефектами

rival конкуруючі моделі

St Louis сент-луисская модель

stochastic стохастическая модель

structural структурна модель

translog транслоговая модель

variable parameter модель із
параметрами, що змінюються

model(l)ing моделювання, побудова
моделі

general-to-specific моделювання від
загального до частки

moment момент

n-th n-й момент

central центральний момент
conditional умовний момент
mixed змішаний момент
non-central нецентральний момент
population момент генеральної сукупності
product змішаний момент
Monte-Carlo стосовний до методу Монте-Карло
Motion рух
Brownian броуновское рух
multicollinearity мультиколінеарність
multinomial мультиноміальний, множинний, поліноміальний, багаточленний || багаточлен або поліном
multiple множинний
choice множинний вибір
coefficient множинний коефіцієнт
correlation множинна кореляція
regression множинна регресія
multiplication множення
matrix матричне множення
pointwise по координатне множення
multiplier множник, мультиплікатор (у кейнсіанстві)
dynamic динамічний мультиплікатор
Lagrange множник Лагранжа
long-run довгостроковий мультиплікатор
short-run короткостроковий мультиплікатор
multivariate багатомірний

N

negligible незначний
neighborhood околі
nest вкладати
nested вкладений (про гіпотезу або модель)
nesting вкладення (однієї гіпотези/моделі в іншу)
artificial штучне вкладення
noise шум, перешкоди
white білий шум
non-causality відсутність причинного зв'язку
non-central нецентральний (про розподіл) **non-centrality** нецентральність

non-degenerate не вироджений
non-linear нелінійний

in variables нелінійний у змінних
intrinsically внутрішньо нелінійний
non-linearity нелінійність
non-negative не негативний
non-negligible значний, помітний, істотний
non-nested не вкладений (про гіпотезу)
non-normality ненормальність, невідповідність нормальному розподілу
non-parametric непараметричний
non-random не випадковий
non-stationary не стаціонарний
non-stochastic не стохастичний, не випадковий
norm норма || нормувати
normal нормальний, стосовний до нормального розподілу

asymptotically normal
 асимптотически нормальний
normality нормальність, відповідність нормальному розподілу
asymptotic асимптотична нормальність
normalization нормування, нормалізація
null нульовий, стосовний до нульової гіпотези
null[hypothesis] нульова гіпотеза
under the в умовах нульової гіпотези
number число

condition число обумовленості
random випадкове число, випадкова величина
numerical чисельний
numerically чисельно

O

observation спостереження
artificial штучне спостереження
influential впливове спостереження
outlying різко, що відхиляється спостереження, викид

observational спостережний,
 призначений для спостережень; який є
 результатом спостережень
observed спостережуваний
omitted пропущений
one-sided однобічний
one-step однокроковий
operator оператор
difference оператор різниці, різницевий
 оператор
expectation оператор математичного
 очікування
lag лаговий оператор
variance оператор дисперсії
optimal оптимальний, найкращий
optimization оптимізація
constrained умовна оптимізація
order порядок, кратність ||
 упорядковувати || порядковий

cointegration порядок коінтеграції
ordered упорядкований
ordering упорядкування, порядок
orthogonal ортогональний
orthogonality ортогональність
orthogonalization ортогоналізація
oscillation коливання
outcome результат
outlier викид, що різко виділяється
 значенням
over-estimate переоцінювати, давати
 завищену оцінку || завищена оцінка
over-identified
 свержидентифицированный
overlap перекривати
overlapping щоперекривається
over-perform перевершувати по робочих
 характеристиках
over-predict давати завищене
 пророкування
over-reject відкидати нульову гіпотезу
 занадто часто (про критерії)
over-smoothed перезгладжений

P
P-value≡probability value

pairwise попарно || попарний
paradigm парадигма
Bayesian байєсовська парадигма
parameter параметр
non-centrality параметр нецентральності
nuisance параметр, що заважає
orthogonal ортогональний параметр
redundant надлишковий параметр
restricted обмежений параметр, параметр
 із обмеженнями
sample [вибіркова] оцінка параметра
scale параметр масштабу
time-dependent залежний від часу
 параметр
time-varying мінливий у часі параметр
true вірний параметр
unknown невідомий параметр
variable змінний параметр
parameter-free вільний від параметра
parametric параметричний
hypothesis гіпотеза
model модель
parametrization параметризація
parsimonious ощадливий, простий
parsimony економія
partial частковий, неповний
autocorrelation автокореляція
partition розбивка; розбивати, розділяти
matrix блокова розбивка матриці
sum-of-squares розкладання суми
 квадратів
pattern схема, структура, зразок
seasonal структура сезонності, характер
 сезонності
penalty штраф
percentage відсоток, процентне
 відношення, частка
percentile процентиль
performance поведіння,
 характеристика, якість роботи
model якість моделі (наскільки добре
 “працює” модель)
period період, часовий інтервал, строк
base базисний період
estimation період оцінювання
periodic періодичний, циклічний

periodogram періодограма, графік спектральної функції
permutation перестановка
perturbation збурювання, пертурбація
pilot пробний, попередній, експериментальний
plan план
plausible правдоподібний, імовірний
plot графік
normal нормальний графік
point точка
turning точка екстремума
point-optimal точково-оптимальний
polynomial багаточлен, поліном, поліноміальний
characteristic характеристичний багаточлен
lag лаговий багаточлен
matrix матричний багаточлен
orthogonal ортогональний багаточлен
population генеральна сукупність
posterior апостеріорний
post-multiplication множення праворуч
power потужність (критерію), сила, ступінь
asymptotic асимптотическая потужність
forecasting точність прогнозу
test потужність критерію
powerful потужний
most найбільш потужний
uniformly most рівномірно найбільш потужний
precision точність
pre-determined заданий, визначений наперед, заздалегідь встановлений
prediction прогноз, прогнозування
biased зміщений прогноз
predictive прогнозований
pre-multiplication множення ліворуч
pre-sample довибірковий
pre-test попереднє тестування
pre-whitening попереднє відбілювання
principle принцип
encompassing принцип охоплення
parsimony принцип економії (в описі дійсності), принцип простоти

prior попередній, апіорний
probability імовірність
conditional умовна ймовірність
fiducial фідучіальна ймовірність
posterior апостеріорна ймовірність
prior апіорна ймовірність
subjective суб'єктивна ймовірність
transition ймовірність переходу
problem проблема, задача, питання
identification проблема ідентифікації
procedure процедура, метод дії
Hildreth-Lu метод Хилдрета-Лу
process процес
adjustment процес пристосування
autoregressive авторегресійний процес
cointegrated коінтегрований процес
correlated корельований процес
data generating дані процес, що породжує
difference-stationary стаціонарний у перших разницях процес
ergodic ергодичний процес
explosive “вибуховий” процес
gaussian гаусівський процес
integrated інтегрований процес
invertible оборотний процес
moving-average процес ковзного середнього
Markoff(Markov) марковський процес
Poisson пуассонівський процес
random випадковий процес
random walk процес випадкового блукання
stationary стаціонарний процес
stochastic стохастичний процес, випадковий процес
trend-stationary стаціонарний щодо тренда процес
processing обробка
data обробка даних
product добуток
inner внутрішній добуток (матриць), скалярний добуток
Kronecker добуток Кронекера
outer зовнішній добуток (матриць)

outer of gradient зовнішній добуток градієнта
scalar скалярний добуток
projection проекція
orthogonal ортогональна проекція
property властивість
asymptotic асимптотичні властивості
finite-sample властивості в кінцевій вибірці
proportion співвідношення, пропорція, частка
proxy[variable] замітник, ерзац-змінна (не дуже якісний замітник змінної, котра повинна стояти в рівнянні)
pseudo-inverce псевдообернений
pseudo-random псевдовипадковий

Q

Q-statistic Q-статистика
qualitative якісний
quantile квантиль
quantitative кількісний
quarterly кварталний
quartile квантиль
quasi- квазі-, псевдо-
quasi-random псевдовипадковий
quotient частка, коефіцієнт, відношення

R

randomization рандомізація
range діапазон, розмах, інтервал
sample (sampling) розмах вибірки
rank ранг

of a matrix ранг матриці
full повний ранг
ranking упорядкування, ранжування
rate норма, ставка, коефіцієнт, частка

of convergence швидкість збіжності
growth темп росту, темп приросту
survival коефіцієнт дожиття
ratio відношення, пропорція, коефіцієнт
likelihood відношення правдоподібності
variance дисперсійне відношення

reciprocal зворотна величина
recognition розпізнавання, розрізнення
pattern розпізнавання образів
recursive рекурентний, рекурсивний
reduce скорочувати, зменшувати, перетворювати, зводити
reduced наведений, скорочений
reduction приведення, відомість, скорочення, зменшення, перетворення
data скорочення даних
variance зниження дисперсії
redundant надлишковий, зайвий
regime режим
region область
confidence довірча область
critical критична область
stationarity область стаціонарності
regressand регресанд, залежна змінна в регресії
regression регресія
artificial штучна регресія
auxiliary допоміжна регресія
censored цензурована регресія
double-length подвоєна регресія
dynamic динамічна регресія
Gauss-Newton регресія Гаусса-Ньютона
harmonic гармонійна регресія
logistic логістична регресія
multiple множинна регресія
non-linear нелінійна регресія
non-parametric непараметрична регресія
nonsense безглузда регресія
orthogonal ортогональна регресія
polynomial поліноміальна регресія
ridge гребнева регресія, рідж-регресія
semiparametric напівпараметрична регресія
simple проста регресія
spurious помилкова регресія
static статична регресія
switching регресія, що перемикається
truncated усічена регресія
regressive регресійний
regressor регресор, незалежна змінна в регресії
reject відкидати,

relation відношення, зв'язок, залежність
negative негативний зв'язок
positive позитивний зв'язок
strong сильний зв'язок
relationship відношення, співвідношення, залежність, зв'язок
cointegrating коінтегруюче відношення
functional функціональний зв'язок
spurious помилкова залежність
relative відносний, порівняльний
reliability вірогідність, надійність
replication повторення
residual залишковий
autocorrelation залишкова автокореляція
generalized узагальнені залишки
squared квадрат остачі
standardized нормовані залишки
residual-based на основі залишків
resolution orthogonal ортогональне розкладання
response відгук, реакція, відповідь
responsiveness чутливість
restricted обмежений
restriction обмеження, функція обмеження
to impose a накласти обмеження
on parameters обмеження на параметри
a priori апіорне обмеження
equality обмеження у вигляді рівності
identifying ідентифікуючі обмеження
inequality обмеження у вигляді нерівності
linear лінійне обмеження
non-linear нелінійне обмеження
zero обмеження, що полягає в тому, що параметр дорівнює нулю
revision перегляд, виправлення
data перегляд даних
risk ризик
Bayes байєсівський ризик
robust робастний, стійкий до відхилень від прийнятих припущень
robustness робастність, стійкість до відхилень від прийнятих припущень
root корінь
explosive “вибуховий” корінь

latent eigenvalue
square квадратний корінь
stable стабільний корінь
unit одиничний корінь
root-mean-square середньоквадратичний
rotation обертання
rule правило
of thumb практичне, емпіричне правило/прийом
sample вибірка || вибірковий
with replacement вибірка з поверненням
artificial штучна вибірка
balanced урівноважена вибірка
biased зміщена вибірка, необ'єктивна вибірка
finite кінцева вибірка
infinite нескінченна вибірка
moderate вибірка середнього розміру
ordered упорядкована вибірка
random випадкова вибірка
representative представницька вибірка, репрезентативна вибірка
small мала вибірка
truncated усічена вибірка

S

sampling вибір, вибірка
with replacement вибір з поверненням
correlated корельований вибір
selective виборчий вибір
scalar скаляр
scale масштаб, шкала, розмір
scaling масштабування, зміна масштабу
scatter[plot,diagram] діаграма розсіювання
scheme схема, процес

autoregressive авторегресійна схема, авторегресійний процес

Markoff(Markov) марковський процес
screening відсівання
seasonal сезонний
seasonal[variable] сезонна змінна

seasonality сезонність

second-order другого порядку

selection відбір, вибір

bandwidth **вибір** ширини вікна

model вибір моделі

sample відбір спостережень

selectivity вибірковість

semi-logarithmic напівлогарифмічний

semi-parametric напівпараметричний

sensitivity чутливість

sequence послідовність

sequential послідовний

serial серіальний, послідовний

serially серіально, послідовно

series ряд

integrated time інтегрований часовий ряд

stationary стаціонарний ряд

time часовий ряд

set множина, набір, система, серія

of equations система рівнянь

of restrictions набір обмежень

data набір даних

information інформаційна множина

shift зрушення, зміна, перемикання

regime зміна режиму

short-run короткостроковий

sign знак

correct вірний знак, правильний знак

wrong невірний знак

significance значущість, рівень значущості

regression значущість регресії

statistical статистична значущість

test значущість критерію, рівень значущості критерію

significant значущий

significant at ... % level значущий на рівні ...%

simulation моделювання, імітація, імітаційне моделювання, імітаційний експеримент

Monte-Carlo метод Монте-Карло

Simultaneous одночасний, спільний

Simultaneity одночасність, спільність

singular вироджений, сингулярний

size розмір, величина

sample розмір вибірки

test розмір критерію

skew асиметричний, скошений

skewness асиметрія

slope кутовий коефіцієнт, нахил (про лінійну функцію)

small-sample стосовний до малої вибірки

smooth гладкий || згладжувати

smoothing згладжування

exponential експонентне згладжування

kernel ядерне згладжування

linear лінійне згладжування

smoothness гладкість

source джерело, початок, першопричина

of variation джерело варіації

space простір

error простір помилок

estimation простір оцінок

parameter простір параметрів

probability імовірнісний простір

sample вибірковий простір

span[лінійна] оболонка || охоплювати, натягати (на вектори)

spatial просторовий

specification специфікація

dynamic динамічна специфікація

model специфікація моделі, опис моделі

system специфікація системи

spectra *множ. ч. від spectrum*

spectral спектральний

analysis спектральний аналіз

function спектральна функція

spectrum спектр, спектральна функція, спектральна щільність || спектральний

band смуговий спектр

spline сплайн

cubic кубічний сплайн

spread розкид, варіація

square квадрат, квадратний; площа

root квадратний корінь

least[method] метод найменших квадратів
weighted least [method] зважений метод найменших квадратів
mean середній квадрат
stability стабільність, стійкість
parameter стабільність параметрів
structural стабільність структури, структурна стабільність
stabilization стабілізація
standardize стандартизувати, нормувати, нормалізувати
standardized стандартизований, нормований, нормалізований
starting початковий
state стан
static статичний || статика
statical статичний, нерухомий, стаціонарний
stationarity стаціонарність
strict стаціонарність у вузькому розумінні
weak слабка стаціонарність
stationary стаціонарний, сталий
covariance стаціонарний
difference стаціонарний у перших різницях, стаціонарний за приростами
trend стаціонарний щодо тренда
statistic статистика, статистичний показник
descriptive описова статистика
Durbin-Watson статистика Дарбіна-Уотсона
likelihood ratio статистика відношення правдоподібності
non-parametric непараметрична статистика
order порядкова статистика
sufficient достатня статистика
summary зведення статистик
trace статистика сліду
statistical статистичний
statistically статистично
statistics статистика (як наука) , статистичні методи, статистичні дані
steady-state сталий, стаціонарний
stochastic стохастичний, випадковий
stratification розшарування, стратифікація

structure структура
lag структура лага
structural структурний
sufficiency достатність
sufficient достатній
sum сума

of squares сума квадратів
cumulative of squares кумулятивна сума квадратів
explained of squares пояснена сума квадратів
residual of squares сума квадратів залишків, залишкова сума квадратів
total of squares загальна сума квадратів
surface поверхня
surprise несподіванка, несподіване відхилення || несподіваний
survey обстеження, вибіркове обстеження, опитування
survival дожиття, виживання
switch перемикання, зміна
regime зміна режиму
switching що перемикається
system система, система рівнянь
recursive рекурсивна система
simultaneous система одночасних рівнянь
static статична система

T
t-distribution t-розподіл, розподіл Стюдента
t-statistic t-статистика, статистика Стюдента
table таблиця
frequency частотна таблиця
tail хвіст (розподілу)
technique метод, спосіб, техніка
temporal тимчасовий, тимчасовий
tentative пробний
term член
constant константа, вільний член
leading-order член вищого порядку
quadratic квадратичний член
test критерій, тест || перевірити
to ... a hypothesis перевірити гіпотезу

chi-square критерій хі-квадрат
Chow критерій Чоу
conditional moment критерій умовного моменту
Durbin-Watson критерій Дарбіна-Уотсона
efficient ефективний критерій
exact точний критерій
Glejser критерій Глейзера
Godfrey критерій Годфрея
Goldfeld-Quandt критерій Голдфелда-Квандта
goodness-of-fit критерій згоди
homogeneity критерій однорідності
information matrix критерій інформаційної матриці
invariant інваріантний критерій
Jarque-Bera критерій Жака-Бера
Lagrange multiplier критерій множника Лагранжа
Ljung-Box критерій Льюнга-Бокса
likelihood ratio критерій відношення правдоподібності
misspecification критерій помилки специфікації
most powerful найбільш потужний критерій
non-linearity критерій нелінійності
non-parametric непараметричний критерій
normality критерій нормальності
omitted variables критерій пропущених змінних
one-sided однобічний критерій

one-tail $\equiv$ **one-sided**
overidentifying restrictions критерій надідентифікуючих обмежень
parameter constancy критерій сталості параметрів
parameter-free вільний від параметра критерій
predictive failure критерій [не]адекватності пророкувань
preliminary попередній критерій
randomization критерій рандомізації
rank ранговий критерій

specification критерій правильності специфікації
two-sided двосторонній критерій
unbiased незміщений критерій

uniformly most powerful рівномірно найбільш потужний критерій
unit root критерій одиничних корінь
variable addition критерій додавання змінних
variable deletion критерій видалення змінних
Wald критерій Вальда
White's критерій Уайта
testing перевірка, перевірка гіпотез, тестування
hypothesis перевірка гіпотез
sequential послідовна перевірка гіпотез
theorem теорема

of total probability формула повної ймовірності
central limit центральна гранична теорема
Gauss-Markov теорема Гаусса-Маркова
theory теорія
asymptotic асимптотична теорія
threshold поріг, критичний рівень
time
time-dependent залежний від часу
time-ordered упорядкований за часом
time-series часовий ряд
time-varying мінливий у часі
tool інструмент
trace слід (матриці)
trade-off зворотна залежність, альтернатива, вибір
bias-variance зворотне відношення між зсувом і дисперсією (оцінки)
power-size зворотне відношення між потужністю й розміром (критерію)
transform перетворювати, трансформувати || перетворення, результат перетворення
Fourier перетворення Фур'є
transformation перетворення

Box-Cox перетворення Боксу-Коксу
Koysck перетворення Ліжко
orthogonal ортогональне перетворення
translog транслоговий
transitional перехідний, проміжний
trend тренд, тенденція
common загальний тренд
cyclic(al) циклічний тренд
deterministic детермінований тренд
linear лінійний тренд
multiplicative мультиплікативний тренд
polynomial поліноміальний тренд
quadratic квадратичний тренд
stochastic стохастичний тренд
upward
 зростаючий тренд, тенденція до росту
trending тренд, що має (про змінну)
trial випробування, проба
Bernoulli випробування Бернуллі
independent незалежні випробування
truncate усікати, відкидати
truncated усічений
truncation усікання
lag усікання лага, відкидання лагів
two-factor двофакторний
two-sample двовибірковий
two-sided двосторонній (про критерії)

U

unbiased незміщений
uncentred нецентрований
unconditional безумовний
uncorrelated некорельований
undated недатований, без дат
under-estimate недооцінювати, давати занижену оцінку || занижена оцінка, недооцінка
uniform рівномірний, однорідний
uniformly рівномірно, однорідно
unit одиничний || одиниця, елемент

of measure одиниця виміру

Economic економічна одиниця, господарська одиниця
observation одиниця спостереження

sample елемент вибірки
unique єдиний
univariate одномірний
unobservable неспостережуваний
unobservable неспостережувана змінна
unrestricted необмежений, без обмежень, безумовний
unrestrictedly без обмежень, необмежено
unstable нестійкий
update обновляти (напр. дані)
upward спрямований нагору, що рухається нагору

V

valid дійсний, що зберігає силу, обґрунтований
statistically статистично обґрунтований
validity дійсність, законність, обґрунтованість
value значення
actual фактичне значення
anomalous аномалія, аномальні значення
approximate наближене значення
average середнє [значення]
critical критичне значення
estimated оцінка, оцінене значення, розрахункове значення
expected очікуване значення, математичне очікування
fitted розрахункове значення, підібране значення
initial вихідне (початкове) значення
mean середнє [значення], математичне очікування
median медіана

P≡ probability

sample вибіркове значення
standardized нормована величина
starting початкове значення
stationary стаціонарне значення
table табличне значення
variability мінливість
variable змінна, величина || що змінюється

to drop a опустити змінну, не включити змінну
artificial штучна змінна
auxilliary допоміжна змінна
binary бінарна змінна, двійкова змінна (приймаюча тільки два значення)
concomitant супутня змінна
controllable керована змінна
constructed сконструйована змінна
dependent залежна змінна
deterministic детермінована змінна
dichotomous дихотомічна змінна
discrete дискретна змінна
dummy фіктивна змінна
endogenous ендегенна змінна
excluded виключена змінна
exogenous екзогенна змінна
explanatory пояснююча змінна
independent незалежна змінна
indicator індикаторна змінна
instrumental інструментальна змінна
instrumental[method] метод інструментальних змінних
lagged змінна з лагом
latent прихована змінна
leading випереджальна змінна
limited dependent обмежена залежна змінна

omitted пропущена змінна
predetermined визначена змінна
proxy перемінна-мінлива-ерзац-змінна
qualitative dependent якісна залежна змінна
random випадкова змінна, випадкова величина
redundant зайва змінна
seasonal сезонна змінна
stationary стаціонарна змінна
stochastic випадкова змінна, випадкова величина
unobservable неспостережувана змінна
vector random векторна випадкова величина, випадковий вектор
variance дисперсія, другий центральний момент
about mean дисперсія щодо середнього, **about regression** дисперсія щодо регресії
bounded обмежена дисперсія
conditional умовна дисперсія
error дисперсія помилки
generalized узагальнена дисперсія
limiting гранична дисперсія
relative відносна дисперсія
residual залишкова дисперсія

Д.7 Статистичні таблиці

Статистичні таблиці

Д.7.1 Значення функції Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,00000	0,00399	0,00798	0,01197	0,01596	0,01994	0,02392	0,02791	0,03188	0,03586
0,10	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06750	0,07143	0,07535
0,20	0,07926	0,08317	0,08707	0,09096	0,09484	0,09871	0,10257	0,10642	0,11026	0,11409
0,30	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,40	0,15542	0,15910	0,16276	0,16640	0,17003	0,17365	0,17724	0,18082	0,18439	0,18794
0,50	0,19146	0,19498	0,19847	0,20195	0,20540	0,20884	0,21226	0,21566	0,21905	0,22241
0,60	0,22575	0,22907	0,23237	0,23566	0,23892	0,24216	0,24538	0,24857	0,25175	0,25491
0,70	0,25804	0,26115	0,26424	0,26731	0,27035	0,27338	0,27638	0,27935	0,28231	0,28524
0,80	0,28815	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,90	0,31594	0,31859	0,32122	0,32382	0,32639	0,32895	0,33147	0,33398	0,33646	0,33892
1,00	0,34135	0,34375	0,34614	0,34850	0,35083	0,35314	0,35543	0,35769	0,35993	0,36215
1,10	0,36434	0,36650	0,36865	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,20	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40148
1,30	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774
1,40	0,41925	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43057	0,43189
1,50	0,43320	0,43348	0,43575	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44409
1,60	0,44520	0,44630	0,44739	0,44845	0,44950	0,45053	0,45155	0,45254	0,45352	0,45449
1,70	0,45544	0,45637	0,45729	0,45819	0,45907	0,45994	0,46080	0,46164	0,46246	0,46328
1,80	0,46407	0,46485	0,46562	0,46638	0,46712	0,46785	0,46856	0,46926	0,46995	0,47062
1,90	0,47129	0,47194	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47671
2,00	0,47725	0,47779	0,47831	0,47882	0,47933	0,47982	0,48030	0,48078	0,48124	0,48169
2,10	0,48214	0,48257	0,48300	0,48342	0,48383	0,48422	0,48462	0,48500	0,48537	0,48574
2,20	0,48610	0,48645	0,48679	0,48713	0,48746	0,48778	0,48809	0,48840	0,48870	0,48899
2,30	0,48928	0,48956	0,48983	0,49010	0,49036	0,49062	0,49086	0,49111	0,49135	0,49158
2,40	0,49180	0,49203	0,49224	0,49245	0,49266	0,49286	0,49306	0,49325	0,49343	0,49362
2,50	0,49379	0,49397	0,49413	0,49430	0,49446	0,49462	0,49477	0,49492	0,49506	0,49520
2,60	0,49534	0,49548	0,49561	0,49573	0,49586	0,49598	0,49610	0,49621	0,49632	0,49643
2,70	0,49654	0,49664	0,49674	0,49684	0,49693	0,49703	0,49711	0,49720	0,49728	0,49737
2,80	0,49745	0,49753	0,49760	0,49768	0,49775	0,49782	0,49788	0,49795	0,49801	0,49808
2,90	0,49814	0,49820	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,00	0,49865	0,49870	0,49874	0,49878	0,49882	0,49886	0,49890	0,49893	0,49897	0,49900
3,10	0,49903	0,49907	0,49910	0,49913	0,49916	0,49919	0,49921	0,49924	0,49927	0,49929
3,20	0,49932	0,49934	0,49936	0,49938	0,49940	0,49943	0,49945	0,49946	0,49948	0,49950
3,30	0,49952	0,49954	0,49955	0,49957	0,49958	0,49960	0,49961	0,49963	0,49964	0,49965
3,40	0,49967	0,49968	0,49969	0,49970	0,49971	0,49522	0,49973	0,49974	0,49975	0,49976
3,50	0,49977	0,49978	0,49979	0,49979	0,49980	0,49981	0,49982	0,49982	0,49983	0,49984
3,60	0,49984	0,49985	0,49986	0,49986	0,49987	0,49988	0,49988	0,49988	0,49989	0,49989
3,70	0,49989	0,49990	0,49990	0,49991	0,49991	0,49991	0,49992	0,49992	0,49992	0,49993
3,80	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995	0,49995
3,90	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997	0,49997
4,00	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998	0,49998	0,49998
4,10	0,49998	0,49998	0,49998	0,49998	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999
4,20	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999

Д.7.2 Таблиця значень F -критерія Фішера на рівні значущості

$$\alpha = 0,05$$

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
1	2	3	4	5	6	7	8	9	10	11
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48

Д.7.2 Таблиця значень F -критерія Фішера на рівні значущості $\alpha = 0,05$ (продовження)

1	2	3	4	5	6	7	8	9	10	11
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

Д.7.3 Критичні значення t -критерія Стюдента на рівні значущості 0,10, 0,05, 0,01 (двосторонній)

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	00,10	0,05	0,01		00,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Д.7.4. Квантили розподілу Хі-квадрат

$\alpha \backslash v$	0,005	0,010	0,025	0,050	0,100	0,900	0,950	0,975	0,990	0,995
1	0,000039	0,00016	0,00098	0,0039	0,0158	2,71	3,84	5,02	6,63	7,88
2	0,010	0,0201	0,0506	0,1026	0,2107	4,61	5,99	7,38	9,21	10,6
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	2,2	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	0,6	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	8,6	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	63,67
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	106,6	112,3	116,3
100	67,33	70,06	74,22	77,93	82,36	118,5	124,34	129,6	135,8	140,2
120	83,85	86,92	91,58	95,70	100,62	140,2	146,57	152,2	159,0	163,6

Д.7.5 Статистики Дарбіна-Уотсона d_L, d_U на 5%-ому рівні значущості

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	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	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	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

Д.7.5. (продовження) Статистики Дарбіна-Уотсона d_L, d_U на 5%-ому рівні значущості для $k = \overline{6...10}$

	$k = 6$		$k = 7$		$k = 8$		$k = 9$		$k = 10$	
N	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-
11	0,203	3,005	-	-	-	-	-	-	-	-
12	0,268	2,832	0,171	3,149	-	-	-	-	-	-
13	0,328	2,692	0,230	2,985	0,147	3,266	-	-	-	-
14	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,360	-	-
15	0,447	2,472	0,343	2,727	0,251	2,979	0,175	3,216	0,111	3,438
16	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,090	0,155	3,304
17	0,554	2,318	0,451	2,537	0,356	2,757	0,272	2,975	0,198	3,184
18	0,603	2,257	0,502	2,461	0,407	2,667	0,321	2,873	0,244	3,073
19	0,649	2,206	0,549	2,396	0,456	2,589	0,369	2,783	0,290	2,974
20	0,692	2,162	0,595	2,339	0,502	2,521	0,416	2,704	0,336	2,885
21	0,732	2,124	0,637	2,290	0,547	2,4(iO	0,461	2,633	0,380	2,806
22	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571	0,424	2,734
23	0,804	2,061	0,715	2,208	0,628	2,360	0,545	2,514	0,465	2,670
24	0,837	2,035	0,751	2,174	0,666	2,318	0,584	2,464	0,506	2,613
25	0,868	2,012	0,784	2,144	0,702	2,280	0,621	2,419	0,544	2,560
30	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251	0,712	2,363
40	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,945	2,149
50	1,219	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110	2,044
60	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939	1,222	1,984
70	1,433	1,802	1,401	1,837	1,369	1,873	1,337	1,910	1,305	1,948
80	1,480	1,801	1,428	1,831	1,425	1,861	1,397	1,893	1,369	1,925
90	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881	1,420	1,909
100	1,550	1,803	1,528	1,826	1,506	1,850	1,484	1,874	1,462	1,898
150	1,651	1,817	1,637	1,832	1,622	1,847	1,608	1,862	1,594	1,877
200	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863	1,665	1,874

Д.7.6. Статистики Дарбіна-Уотсона d_L, d_U на 1%-ому рівні значущості

n	$k'=1$		$k'=2$		$k'=3$	
	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46
16	0.84	1.09	0.74	1.25	0.63	1.44
17	0.87	1.10	0.77	1.25	0.67	1.43
18	0.90	1.12	0.80	1.26	0.71	1.42
19	0.93	1.13	0.83	1.26	0.74	1.41
20	0.95	1.15	0.86	1.27	0.77	1.41
21	0.97	1.16	0.89	1.27	0.80	1.41
22	1.00	1.17	0.91	1.28	0.83	1.40
23	1.02	1.19	0.94	1.29	0.86	1.40
24	1.04	1.20	0.96	1.30	0.88	1.41
25	1.05	1.21	0.98	1.30	0.90	1.41
26	1.07	1.22	1.00	1.31	0.93	1.41
27	1.09	1.23	1.02	1.32	0.95	1.41
28	1.10	1.24	1.04	1.32	0.97	1.41
29	1.12	1.25	1.05	1.33	0.99	1.42
30	1.13	1.26	1.07	1.34	1.01	1.42
31	1.15	1.27	1.08	1.34	1.02	1.42
32	1.16	1.28	1.10	1.35	1.04	1.43
33	1.17	1.29	1.11	1.36	1.05	1.43
34	1.18	1.30	1.13	1.36	1.07	1.43
35	1.19	1.31	1.14	1.37	1.08	1.44
36	1.21	1.32	1.15	1.38	1.10	1.44
37	1.22	1.32	1.16	1.38	1.11	1.45
38	1.23	1.33	1.18	1.39	1.12	1.45
39	1.24	1.34	1.19	1.39	1.14	1.45

Література

1. Бідюк П.І., Терентьєв О.М., Просянкін-Жарова Т.І. Прикладна статистика. – Київ: НТУУ «КПІ», 2012. – 510 с.
2. Бідюк П.І., Меньяйленко О.С., Половцев О.В. Методи прогнозування. – Луганськ: Альма-матер, 2008. – 305 с.
3. Бідюк П.І., Романенко В.Д., Тимошук О.Л. Аналіз часових рядів. – Київ: Видавництво «Політехніка», 2013. – 607 с.
4. Бідюк П.І., Харрінгтон Т., Ткач Б.П. Математична статистика. – Київ: Видавничий Дім «Персонал», 2017. – 348 с.
5. Коваленко І.П. Математична статистика. – Київ: Видавничий Дім «Слово», 2012. – 496 с.
6. Бобик О.І., Берегова Г.І., Копитко Б.І. Теорія ймовірностей і математична статистика. – Київ: Професіонал, 2007. – 560 с.
7. Жиров О.Л. Економетрія [Електронний ресурс] : методичні вказівки до виконання комп'ютерних практикумів для студентів напряму підготовки 6.030601 «Менеджмент» – Київ : НТУУ «КПІ», 2013. – 155 с. –
8. Лук'яненко І. Г., Краснікова Л. І. Економетрика: Підручник. — К.: Товариство "Знання", КОО, 1998. — 494 с. (з табл., граф.).
9. Бідюк П.І., Половцев О.В. Аналіз та математичне моделювання економічних процесів перехідного періоду, Київ, НТУУ «КПІ», 1999, 230с.
10. Gelman A., Carlin J.B., Stern H.S., Rubin D.B. Bayesian Data Analysis. – New York: CRC Press, 2004. – 670p. (надається в електронному вигляді).