

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ
УКРАЇНИ**

ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

МЕТОДИЧНІ ВКАЗІВКИ

**для самостійної роботи студентів та виконання практичних занять
з дисципліни**

***” Багатовимірний аналіз метеорологічних
полів та атмосферних процесів ”***

**для студентів III, IV курсів гідрометеорологічного інституту
та аспірантів**

Напрямок підготовки: гідрометеорологія

ЗАТВЕРДЖЕНО

**на засіданні методичної комісії
гідрометеорологічного інституту**

Протокол № ____ від _____ 2012 р.

ОДЕСА - 2012

Методичні вказівки для самостійної роботи студентів та виконання
практичних занять з дисципліни
''Багатовимірний аналіз метеорологічних полів та атмосферних процесів''
для студентів III, IV курсів та аспірантів
денної форми навчання
за напрямками ''Гідрометеорологія'', ''Екологія, охорона навколишнього
середовища та збалансоване природокористування''.

Укладачі: к.г.н., доцент Гончарова Л.Д.;
к.г.н., доцент Галич Є.А.;
д.т.н., професор Шкільний Є.П.

Одеса, ОДЕКУ, 2012 р. 39 с., укр. мова.

ЗМІСТ

Передмова.....	4
1 Загальні теоретичні положення	5
1.1 Задача створення прогностичної моделі.....	5
1.2 Підготовка вихідної інформації для побудови регресійної моделі	6
1.3 Поняття про частинний коефіцієнт кореляції.....	9
1.4 “Просіювання” предикторів та розрахунок параметрів моделі.....	12
1.5 Оцінка міри адекватності прогностичної моделі метеорологічного прогнозу.....	15
1.5.1 Множинний коефіцієнт кореляції.....	15
1.5.2 Перевірка статистичної гіпотези за допомогою критерію Фішера.....	19
2 Запитання для самоперевірки.....	20
3 Порядок виконання роботи	21
Література	21
Додаток А Приклад застосування регресійного аналізу при дослідженні метеорологічних полів.....	22
Додаток Б Вихідні дані для виконання роботи.....	30
Додаток В Відомості про комп’ютерну програму “STEP”.....	36
Додаток Г Значення критерію Стюдента.....	37
Додаток Д Значення критерію Фішера F для рівня значущості $\alpha = 0.05$	38

ПЕРЕДМОВА

Методичні вказівки складені для СРС при поглибленому вивченні та закріпленні лекційного матеріалу за темою “*Побудова регресійної моделі метеорологічного прогнозу*”, а також для здобуття практичних навичок з дисциплін “Багатовимірний аналіз метеорологічних полів та атмосферних процесів” (рівень бакалавр) та “Статистичні методи аналізу та обробки наукової інформації” (для аспірантів).

Регресійний аналіз є одним з методів багатовимірного статистичного аналізу, за допомогою якого реалізується задача створення прогностичної моделі метеорологічного прогнозу у вигляді рівняння множинної лінійної регресії.

Методичні вказівки містять у собі стислий огляд теоретичних положень, завдання на підготовку до практичної частини роботи, порядок її проведення та виконання, запитання до самоперевірки та розв’язання типової задачі.

Рекомендується розпочинати виконання практичної роботи після самостійної роботи по вивченню теоретичних положень з цієї теми (див. список літератури) та відповідей на контрольні запитання.

В якості вихідних даних для виконання практичної роботи використовуються результати наукових досліджень авторів щодо атмосферних процесів в Атлантико-Європейському регіоні. Варіанти вихідних даних та приклад застосування регресійного аналізу при дослідженні метеорологічних полів наводяться у додатках. Реалізація алгоритму «просіювання» предикторів методом покрокової регресії здійснюється за допомогою комп’ютерної програми “STEP” (автор доц. Верлан В.А.).

Методичні вказівки призначені для студентів денної форми навчання ОДЕКУ за напрямом гідрометеорологія: військова метеорологія, прикладна метеорологія та метеорологічні прогнози.

Також вони можуть бути використані для магістрів та аспірантів ОДЕКУ при дослідженнях статистичних взаємозв’язків в ланках кліматичної системи.

Мета роботи

Реалізувати алгоритм покрокової регресії, отримати параметри прогностичної регресійної моделі та з заданою ймовірністю побудувати рівняння множинної лінійної регресії для температури та опадів в окремих регіонах України на оптимальних предикторах. Реалізувати на ПК за допомогою комп’ютерної програми “STEP” алгоритм покрокової регресії на основі статистичних сукупностей підібраних потенційних предикторів та предиктанта і з заданою ймовірністю побудувати рівняння множинної регресії. Оцінити міру адекватності отриманої прогностичної моделі метеорологічного прогнозу та провести ретельний аналіз отриманих результатів.

1 ЗАГАЛЬНІ ТЕОРЕТИЧНІ ПОЛОЖЕННЯ

1.1 Задача створення прогностичної моделі

При складанні прогнозів метеорологічних величин і явищ прогностичними центрами використовуються чисельні моделі двох типів: гідродинамічні та статистичні. Досить часто статистичні моделі метеорологічних прогнозів розробляються у вигляді лінійного множинного рівняння регресії

$$u = b_0 + b_1 z_1 + \dots + b_i z_i + \dots + b_n z_n. \quad (1)$$

У рівнянні (1) величину u , що прогнозується, називають *предиктантом*, величини z_i ($i = \overline{1, n}$) – *предикторами*, а коефіцієнти b_i ($i = \overline{0, n}$) – *коефіцієнтами регресії*.

Якщо предиктори, що мають сенс фізичних факторів, впливаючих на формування предиктанта u , центровані та нормовані на середні квадратичні відхили величини, тобто

$$x_{ij} = \frac{z_{ij} - \bar{z}_i}{\sigma_{z_i}}, \quad (i = \overline{1, n}; j = \overline{1, m}), \quad (2)$$

то у рівнянні (1) вільний член дорівнює нулю і рівняння регресії приймає такий вигляд:

$$y = a_1 x_1 + a_2 x_2 + \dots + a_i x_i + \dots + a_n x_n. \quad (3)$$

Задача створення прогностичної моделі (3) полягає у тому, щоб на основі статистичних сукупностей предиктанта (y) і системи предикторів отримати значення коефіцієнтів регресії a_i ($i = \overline{1, l}$). Як правило, ця задача розв'язується за допомогою метода найменших квадратів (МНК).

Створення самої регресійної моделі метеорологічного прогнозу (3), тобто оцінка коефіцієнтів регресії $a_1, a_2, \dots, a_l$, є суто статистичною задачею, якщо предиктори визначені й для них сформовані статистичні сукупності. Але для створення складу предикторів треба розв'язати фізичну задачу, тобто на основі знань про фізику процесу формування метеорологічної величини або утворення метеорологічного явища, що прогножуються, необхідно визначити систему впливаючих факторів. Усі вони складають множину *потенційних (можливих)* предикторів.

Якщо множина потенційних предикторів сформована, то у подальшому потрібно перейти до розв'язування іншої статистичної задачі - вибору з системи потенційних предикторів таких, які з заданою ймовірністю чинять найбільший вплив на формування предиктанта y . Визначений таким чином склад l предикторів ($l \ll n$) називають *оптимальним складом*, а предиктори, що входять до цього складу - *статистично значущими*. Процедури, які дають можливість визначити зі складу потенційних статистично значущі предиктори, називають *процедурами "просіювання"* предикторів. Існує декілька методів, що дозволяють розв'язати цю задачу (за допомогою множинного коефіцієнта кореляції, частинного коефіцієнта кореляції, методом включення або методом покрокової регресії).

Отже, проблему створення статистичної моделі метеорологічного прогнозу (3) можна вирішувати у такій послідовності. Спочатку шляхом розв'язування зазначеної вище фізичної задачі створити склад потенційних предикторів. Після цього за допомогою будь-якого відомого методу "просіювання" вибрати з них множину статистично значущих предикторів, тобто створити систему оптимальних предикторів. Нарешті, на основі, наприклад, МНК розрахувати параметри моделі (3). Але можна цю задачу розв'язати іншим шляхом. Він полягає у тому, що використовуючи метод "просіювання" предикторів, який має назву *методу покрокової регресії*, можна перераховані етапи побудови регресійної моделі метеорологічного прогнозу виконати у комплексі, що значно спрощує розрахунки.

1.2 Підготовка вихідної інформації для побудови регресійної моделі

Вихідні дані:

вибірка предиктанта u об'ємом m

$$u: u_1, u_2, \dots, u_j, \dots, u_m \quad (j = \overline{1, m}) \quad (4)$$

матриця потенційних предикторів

$$Z = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1j} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2j} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{i1} & z_{i2} & \dots & z_{ij} & \dots & z_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nj} & \dots & z_{nm} \end{pmatrix} \quad (i = \overline{1, n}; j = \overline{1, m}). \quad (5)$$

Розрахункова частина:

- розрахунок $\bar{u}$ за формулою:

$$\bar{u} = \frac{\sum_{j=1}^m u_j}{m}; \quad (6)$$

- розрахунки дисперсії предиктанта

$$\sigma_u^2 = \frac{\sum_{j=1}^m (u_j - \bar{u})^2}{m} \quad (7)$$

і середнього квадратичного відхилу

$$\sigma_u = \sqrt{\sigma_u^2}; \quad (8)$$

- розрахунок вектора середніх значень предикторів

$$\bar{Z} = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \vdots \\ \bar{z}_i \\ \vdots \\ \bar{z}_n \end{pmatrix} \quad (i = \overline{1, n}) \quad (9)$$

за формулою

$$\bar{z}_i = \frac{\sum_{j=1}^m z_{ij}}{m} \quad (i = \overline{1, n}); \quad (10)$$

- розрахунок вектора дисперсій

$$\sigma_z^2 = \begin{pmatrix} \sigma_{z_1}^2 \\ \sigma_{z_2}^2 \\ \vdots \\ \sigma_{z_i}^2 \\ \vdots \\ \sigma_{z_n}^2 \end{pmatrix} \quad (i = \overline{1, n}) \quad (11)$$

та вектора середніх квадратичних відхилів предикторів

$$\sigma_z = \begin{pmatrix} \sigma_{z_1} \\ \sigma_{z_2} \\ \vdots \\ \sigma_{z_i} \\ \vdots \\ \sigma_{z_n} \end{pmatrix} \quad (i = \overline{1, n}) \quad (12)$$

за формулами

$$\sigma_{z_i}^2 = \frac{\sum_{j=1}^m (z_{ij} - \bar{z}_i)^2}{m} \quad (i = \overline{1, n}; j = \overline{1, m}), \quad (13)$$

$$\sigma_{z_i} = \sqrt{\sigma_{z_i}^2}; \quad (14)$$

- побудова сукупності центрованих та нормованих значень предиктанта

$$y: y_1, y_2, \dots, y_j, \dots, y_m, \quad (15)$$

де

$$y_j = \frac{u_j - \bar{u}}{\sigma_u} \quad (j = \overline{1, m}); \quad (16)$$

- створення матриці центрованих та нормованих значень предикторів

$$X = \{x_{ij}\}_{n \times m} \quad (i = \overline{1, n}; j = \overline{1, m}), \quad (17)$$

де

$$x_{ij} = \frac{z_{ij} - \bar{z}_i}{\sigma_{z_i}}. \quad (18)$$

Отже, статистична модель метеорологічного прогнозу будується на основі вибірки центрованих та нормованих значень предиктанта (15) та матриці центрованих та нормованих значень предикторів

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{pmatrix}. \quad (19)$$

1.3 Поняття про частинний коефіцієнт кореляції

Раніше (в дисципліні «Методи обробки та аналізу гідрометеорологічної інформації») ми розглядали парні коефіцієнти кореляції, які, як відомо, визначають лінійний кореляційний зв'язок між двома випадковими величинами (наприклад X та Y). При цьому ми вважали, що на випадкову величину Y діють й інші випадкові величини. Але нас не цікавило питання саме які це величини. В дійсності, предиктори, як правило, статистично зв'язані між собою і при побудові статистичних моделей такі зв'язки треба враховувати. Це можна здійснити за допомогою частинних коефіцієнтів кореляції. Дамо визначення частинного коефіцієнта кореляції й розглянемо алгоритм його оцінки.

Припустимо, що на випадкову величину y діють дві випадкові величини x_1 і x_2 . **Частинним коефіцієнтом кореляції між випадковими величинами y і x_1 ($r_{yx_1 \cdot x_2}$) називають коефіцієнт кореляції між ними за**

умови, що вплив іншої випадкової величини x_2 на y вже є врахованим. Таким же чином визначається частинний коефіцієнт кореляції $r_{yx_2 \cdot x_1}$.

Будемо вважати, що нам відома матриця кореляцій

$$R_x = \begin{pmatrix} 1 & r_{x_1 x_2} \\ r_{x_2 x_1} & 1 \end{pmatrix} \quad (20)$$

і вектор парних коефіцієнтів кореляцій між y і x_1 та x_2

$$R_{yx} = \begin{pmatrix} r_{yx_1} \\ r_{yx_2} \end{pmatrix}. \quad (21)$$

На їх основі сформуємо *розширену матрицю кореляцій*

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1 x_2} \\ r_{yx_2} & r_{x_1 x_2} & 1 \end{pmatrix} \quad (22)$$

Вона утворюється з матриці R_x шляхом додавання до неї рядка та стовпця, що складаються з координат вектора R_{yx} . На основі матриці (22) розрахуємо *мінори* $|R_x|$, $\ddot{A}_{yx_i}$, $\ddot{A}_{yx_i}^-$ ($i=1,2$). Мінори $\ddot{A}_{yx_i}$ складаються таким чином: стовпець, на першому місці котрого розташовується парна кореляція r_{yx_i} , переставляється на перше місце, а на його місце ставиться перший стовпець і, після цього, викреслюють перші рядок і стовпець. Очевидно

$$\ddot{A}_{yx_1} = r_{yx_1} - r_{yx_2} r_{x_1 x_2}, \quad (23)$$

$$\ddot{D}_{yx_2} = r_{yx_2} - r_{yx_1} r_{x_1 x_2}. \quad (24)$$

Означення мінора $\ddot{A}_{yx_1}^-$ має такий сенс: це мінор визначника $|\tilde{R}|$, який не утримує парної кореляції r_{yx_1} . Очевидно ми його будемо мати, якщо

викреслимо із мінора $|\tilde{R}|$ рядок і стовпець, що утримують цю парну кореляцію. Отже

$$\ddot{A}_{yx_1} = \begin{vmatrix} 1 & r_{yx_2} \\ r_{yx_2} & 1 \end{vmatrix} = 1 - r_{yx_2}^2, \quad (25)$$

$$\ddot{A}_{yx_2} = \begin{vmatrix} 1 & r_{yx_1} \\ r_{yx_1} & 1 \end{vmatrix} = 1 - r_{yx_1}^2. \quad (26)$$

Частинні коефіцієнти кореляції визначаються таким чином:

$$r_{yx_1 \bullet x_2} = \frac{\ddot{A}_{yx_1}}{\sqrt{|R_x| \ddot{A}_{yx_1}}} = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{\sqrt{(1 - r_{x_1 x_2}^2)(1 - r_{yx_2}^2)}}, \quad (27)$$

$$r_{yx_2 \bullet x_1} = \frac{\ddot{A}_{yx_2}}{\sqrt{|R_x| \ddot{A}_{yx_2}}} = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{\sqrt{(1 - r_{x_1 x_2}^2)(1 - r_{yx_1}^2)}}. \quad (28)$$

Виникає питання, яка суттєва інформація утримується в частинних коефіцієнтах кореляції? Щоб відповісти на нього, розглянемо такий приклад. Нехай парні коефіцієнти кореляції між випадковими величинами мають такі значення: $r_{yx_1} = 0.70$; $r_{yx_2} = 0.90$; $r_{x_1 x_2} = 0.80$. Що можна сказати про ці випадкові величини? Звісно те, що випадкова величина y характеризується дуже тісними кореляційними зв'язками і з величиною x_1 , і з величиною x_2 . Але треба звернути увагу на те, що дві останні випадкові величини теж зв'язані дуже тісним кореляційним зв'язком між собою. Отже, щоб визначити, яка з величин x_i дійсно чинить вплив на величину y , треба розрахувати частинні коефіцієнти кореляції за допомогою формул (27) і (28). Розрахунки дають такі значення: $r_{yx_1 \bullet x_2} = -0.08$; $r_{yx_2 \bullet x_1} = 0.79$. Таким чином, в дійсності на випадкову величину y чинить вплив випадкова величина x_2 . Кореляційний зв'язок y з величиною x_1 , якщо урахувати її зв'язок з величиною x_2 , є не тільки незначним, але навіть має обернений характер.

Щоб розповсюдити отриманий алгоритм розрахунків частинних коефіцієнтів кореляції на n змінних, треба побудувати розширену матрицю

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} & \dots & r_{yx_k} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_k} & \dots & r_{x_1x_n} \\ r_{yx_2} & r_{x_2x_1} & 1 & r_{x_2x_3} & \dots & r_{x_2x_k} & \dots & r_{x_2x_n} \\ r_{yx_3} & r_{x_3x_1} & r_{x_3x_2} & 1 & \dots & r_{x_3x_k} & \dots & r_{x_3x_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_k} & r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} & \dots & 1 & \dots & r_{x_kx_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_nx_1} & r_{x_nx_2} & r_{x_nx_3} & \dots & r_{x_nx_k} & \dots & 1 \end{pmatrix} \quad (29)$$

і на її основі визначити мінори $|R_x|$, $\ddot{A}_{yx_k}$, $\ddot{A}_{yx_k}^-$ ($k=1,2,\dots,n$) за тими ж правилами, як це було зроблено у попередньому випадку для матриці третього порядку.

Отримаємо

$$r_{yx_k \bullet x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n} = \frac{\ddot{A}_{yx_k}}{\sqrt{|R_x| \ddot{A}_{yx_k}^-}}, \quad (30)$$

мінори $|R_x|$, $\ddot{A}_{yx_k}$, $\ddot{A}_{yx_k}^-$ мають порядок n ; матриця R - порядок $n+1$.

Частинні коефіцієнти кореляції використовуються в ряді методів об'єктивного добору предикторів.

1.4 “Просіювання” предикторів та розрахунок параметрів моделі

1. Розрахунок вектора R_{yx} парних коефіцієнтів кореляцій між предиктантом та предикторами

$$R_{yx} = \begin{pmatrix} r_{yx_1} \\ r_{yx_2} \\ \vdots \\ r_{yx_i} \\ \vdots \\ r_{yx_k} \\ \vdots \\ r_{yx_n} \end{pmatrix} \quad (31)$$

за формулою

$$r_{yx_i} = \frac{\sum_{j=1}^m y_j x_{ij}}{m} \quad i = \overline{1, n} \quad (32)$$

2. Аналіз координат вектора R_{yx} та визначення найбільшого за модулем коефіцієнта кореляції між предиктантом та предиктором. Нехай $|r_{yx_k}| = \max$. Тоді $x_k \rightarrow x_1$, тобто x_k визначається як перший предиктор із системи оптимальних предикторів. Він ставиться на перше місце, а ті предиктори, що залишилися, перенумеровуються.
3. Знаходиться оцінка рівняння регресії першого кроку предиктанта y на перший предиктор

$$y^{(1)} = a_1^{(1)} x_1 \quad (33)$$

(номер в дужках позначає номер кроку).

Коефіцієнт рівняння регресії першого кроку визначається за формулою

$$a_1^{(1)} = \frac{\sum_{j=1}^m y_j x_{1j}}{\sum_{j=1}^m x_{1j}^2}. \quad (34)$$

Далі розраховуються всі нев'язки першого кроку між значеннями предиктанта y_j та його значеннями за допомогою моделі (33)

$$\varepsilon_{1j} = y_j - y_j^{(1)}, \quad \forall \quad j = \overline{1, m}. \quad (35)$$

Знаходяться коефіцієнти кореляції $r_{\varepsilon_1 x_s}$ ($s = \overline{2, n}$) між нев'язками першого кроку та всіма предикторами, починаючи з предиктора за номером 2, за формулою:

$$r_{\varepsilon_1 x_s} = \frac{\sum_{j=1}^m \tilde{\varepsilon}_{1j} x_{sj}}{\sqrt{\sum_{j=1}^m \tilde{\varepsilon}_{1j}^2 \sum_{j=1}^m x_{sj}^2}}, \quad (36)$$

де $\tilde{\varepsilon}_{1j}$ – центровані значення нев'язок.

Результат розрахунків – система коефіцієнтів кореляції

$$r_{\varepsilon_1 x_2}, r_{\varepsilon_1 x_3}, \dots, r_{\varepsilon_1 x_p}, \dots, r_{\varepsilon_1 x_n}, \quad (37)$$

які мають сенс частинних коефіцієнтів кореляції. Далі проводиться їх аналіз з метою вибору між ними найбільшого за модулем.

Нехай $\left| r_{\varepsilon_1 x_p} \right| = \max$, тоді $x_p \rightarrow x_2$, тобто x_2 вибирається як другий оптимальний предиктор. Інші $n-2$ предиктори перенумеровуються і відбувається перехід до другого кроку.

4. За допомогою МНК будується рівняння регресії другого кроку:

$$y^{(2)} = a_1^{(2)} x_1 + a_2^{(2)} x_2 \quad (38)$$

і розраховуються нев'язки другого кроку

$$\varepsilon_{2j} = y_j - y_j^{(2)}. \quad (39)$$

Це означає, що зі значень предиктанта вилучаються впливи першого x_1 і другого x_2 предикторів. Після цього розраховуються коефіцієнти кореляції $r_{\varepsilon_2 x_j}$ ($j = \overline{3, n}$), аналіз яких дає підставу для визначення третього оптимального предиктора x_3 і т.д. Процедура триває до тих пір, доки отримані після деякого l -того кроку всі частинні коефіцієнти кореляції $r_{\varepsilon_{l+1} x_j}$ ($j = \overline{l+1, n}$) втрачають статистичну значущість. Гіпотеза H_0 про це перевіряється за допомогою критерію Стюдента

$$t = \frac{\left| r_{\varepsilon_{l+1} x_j} \right|}{\sigma_{r_{\varepsilon_{l+1} x_j}}}, \quad (40)$$

де

$$\sigma_{r_{\varepsilon_{l+1} x_j}} = \frac{1 - r_{\varepsilon_{l+1} x_j}^2}{\sqrt{m-1}}. \quad (41)$$

Система нормальних рівнянь МНК для кожного k -того кроку, починаючи з другого, для визначення коефіцієнтів рівняння регресії

$$y^{(k)} = a_1^{(k)} x_1 + a_2^{(k)} x_2 + \dots + a_{k-1}^{(k)} x_{k-1} + a_k^{(k)} x_k \quad (42)$$

має вигляд

$$\left\{ \begin{array}{l} a_1^{(k)} \sum_{j=1}^m x_{1j}^2 + a_2^{(k)} \sum_{j=1}^m x_{1j} x_{2j} + \dots a_{k-1}^{(k)} \sum_{j=1}^m x_{1j} x_{(k-1)j} + a_k^{(k)} \sum_{j=1}^m x_{1j} x_{kj} = \sum_{j=1}^m y_j x_{1j} \\ a_1^{(k)} \sum_{j=1}^m x_{1j} x_{2j} + a_2^{(k)} \sum_{j=1}^m x_{2j}^2 + \dots a_{k-1}^{(k)} \sum_{j=1}^m x_{2j} x_{(k-1)j} + a_k^{(k)} \sum_{j=1}^m x_{2j} x_{kj} = \sum_{j=1}^m y_j x_{2j} \\ \dots \dots \dots \dots \dots \dots \dots \\ a_1^{(k)} \sum_{j=1}^m x_{1j} x_{kj} + a_2^{(k)} \sum_{j=1}^m x_{2j} x_{kj} + \dots a_{k-1}^{(k)} \sum_{j=1}^m x_{kj} x_{(k-1)j} + a_k^{(k)} \sum_{j=1}^m x_{kj}^2 = \sum_{j=1}^m y_j x_{kj} \end{array} \right. \quad (43)$$

(індекс k пробігає всі значення з послідовності $k : 2, 3, \dots, s, \dots, l$. Якщо $k = s$, то всі значення $x_{(s+1)l}, x_{(s+2)l}, \dots$ вважаються рівними нулю.

Таким чином, за методом покрокової регресії послідовно вибираються з n потенційних предикторів l статистично значущих, на основі котрих і будується лінійне рівняння множинної регресії. Воно має вигляд

$$y = a_1 x_1 + a_2 x_2 + \dots + a_l x_l \quad (44)$$

за умови, що всі змінні спочатку центруються й нормуються за рівностями (16) та (18). Разом з цим, при виконанні покрокової процедури, як допоміжний контролюючий параметр, розраховується відповідний множинний коефіцієнт кореляції, який може використовуватися й для відбору оптимальних предикторів і для оцінки вірогідності прогностичної моделі.

1.5 Оцінка міри адекватності прогностичної моделі метеорологічного прогнозу

1.5.1 Множинний коефіцієнт кореляції

На відміну від парного коефіцієнта кореляції *множинний коефіцієнт* характеризує *тісноту лінійного кореляційного зв'язку не з одним, а з цілою системою предикторів $x_1, x_2, \dots, x_n$, тобто є мірою адекватності регресійної прогностичної моделі*. Для того, щоб обґрунтувати це, а також отримати алгоритм розрахування множинного коефіцієнта кореляції, розглянемо рівняння регресії вигляду

$$y = b_1 x_1 + b_2 x_2. \quad (45)$$

Запишемо метрику найменших квадратів

$$\delta^2 = \sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = \sum_{j=1}^m [y_j - (b_1 x_{1j} + b_2 x_{2j})]^2 \quad (46)$$

і розкриємо праву його частину

$$\begin{aligned} \sum_{j=1}^m S_{y \bullet x_1 x_2}^2 &= \sum_{j=1}^m (y_j^2 - b_1 y_j x_{1j} - b_1 y_j x_{1j} - b_2 y_j x_{2j} - b_2 y_j x_{2j} + \\ &+ b_1^2 x_{1j}^2 + b_1 b_2 x_{1j} x_{2j} + b_1 b_2 x_{1j} x_{2j} + b_2^2 x_{2j}^2). \end{aligned} \quad (47)$$

Якщо провести групування членів правої частини таким чином

$$\begin{aligned} \sum_{j=1}^m S_{y \bullet x_1 x_2}^2 &= \sum_{j=1}^m y_j^2 - b_1 \sum_{j=1}^m y_j x_{1j} - b_2 \sum_{j=1}^m y_j x_{2j} + b_1 (b_1 \sum_{j=1}^m x_{1j}^2 + \\ &+ b_2 \sum_{j=1}^m x_{1j} x_{2j} - \sum_{j=1}^m y_j x_{1j}) + b_2 (b_1 \sum_{j=1}^m x_{1j} x_{2j} + b_2 \sum_{j=1}^m x_{2j}^2 - \sum_{j=1}^m y_j x_{2j}), \end{aligned}$$

то прийдемо до формули

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = \sum_{j=1}^m y_j^2 - b_1 \sum_{j=1}^m y_j x_{1j} - b_2 \sum_{j=1}^m y_j x_{2j}, \quad (48)$$

оскільки останні два члени дорівнюють нулю.

У рівнянні (48) будемо мати

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = m \sigma_y^2 - b_1 m \sigma_{x_1} \sigma_y r_{yx_1} - b_2 m \sigma_{x_2} \sigma_y r_{yx_2}. \quad (49)$$

Розділимо обидві частини рівності (49) на загальний об'єм статистичної сукупності m і введемо позначення

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 / m = \sigma_{y \bullet x_1 x_2}^2. \quad (50)$$

Очевидно *рівність (50) має сенс дисперсії нев'язки рівняння (45)*. Якщо її поділити на дисперсію предиктанта, то будемо мати *відносну дисперсію нев'язки* прогностичного рівняння

$$\tilde{\sigma}_{y \bullet x_1 x_2}^2 = \frac{\sigma_{y \bullet x_1 x_2}^2}{\sigma_y^2}. \quad (51)$$

Отже, з урахуванням формул (50) та (45), маємо

$$\tilde{\sigma}_{y \bullet x_1 x_2}^2 = 1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}. \quad (52)$$

Величина

$$R_{y \bullet x_1 x_2}^2 = \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \quad (53)$$

і характеризує тісноту лінійного кореляційного зв'язку між предиктантом y і предикторами x_1 і x_2 . Її називають *коефіцієнтом множинної детермінації*. Корінь квадратний з цієї величини

$$R_{y \bullet x_1 x_2} = \sqrt{R_{y \bullet x_1 x_2}^2} \quad (54)$$

є *коефіцієнтом множинної кореляції*.

З формул (52)-(54) випливає, що

$$R_{y \bullet x_1 x_2} = \sqrt{1 - \tilde{\sigma}_{y \bullet x_1 x_2}^2}. \quad (55)$$

Формула (55) є мірою адекватності прогностичної моделі.

Дійсно, $\tilde{\sigma}_{y \bullet x_1 x_2}^2$ характеризує ту частину предиктанта, котра не вичерпується моделлю. Чим вона менша, тим більш адекватною є модель. Якщо припустити, що $\tilde{\sigma}_{y \bullet x_1 x_2}^2 = 0$ (цього на практиці, звичайно, не буває), то модель тотожно відбиває процес, що моделюється. Але у цьому випадку $R_{y \bullet x_1 x_2} = 1$. Навпаки, за умови $\tilde{\sigma}_{y \bullet x_1 x_2}^2 = \sigma_y^2$, $\tilde{\sigma}_{y \bullet x_1 x_2}^2 = 1$ (такий прогноз називають випадковим), то $R_{y \bullet x_1 x_2} = 0$.

Отже множинний коефіцієнт кореляції змінюється в межах

$$0 < R_{y \bullet x_1 x_2} < 1. \quad (56)$$

Таким чином, чим більшим є значення множинного коефіцієнта кореляції, тим більш обумовленою є статистична модель.

Узагальнимо тепер отримані результати на рівняння, що утримує n предикторів. Для цього до кореляційної матриці другого порядку, що відповідає рівнянню (45), додамо рядок і стовпець коефіцієнтів кореляції між предиктантом та предикторами

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_1x_2} & 1 \end{pmatrix}. \quad (57)$$

Таку матрицю називають розширеною. Її визначник дорівнює

$$|\tilde{R}| = (1 - r_{x_1x_2}^2) \left(1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2} \right). \quad (58)$$

Перший співмножник правої частини рівності (58) є визначником матриці кореляції $|R_x|$. Якщо на нього поділити обидві частини співвідношення (58), то будемо мати

$$\frac{|\tilde{R}|}{|R_x|} = 1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}. \quad (59)$$

Другий член правої частини рівності (59) є множинним коефіцієнтом детермінації $R_{y \bullet x_1x_2}^2$. Тоді, як свідчить рівність (55), ліва частина (59) дорівнює відносній залишковій дисперсії $\tilde{\sigma}_{y \bullet x_1x_2}^2$. Отже з врахуванням цього маємо

$$R_{y \bullet x_1x_2} = \sqrt{1 - \frac{|\tilde{R}|}{|R_x|}}. \quad (60)$$

Поширюючи отриманий результат на n предикторів, запишемо формулу для визначення множинного коефіцієнта кореляції

$$R_{y \bullet x_1x_2 \dots x_n} = \sqrt{1 - \frac{|\tilde{R}|}{|R_x|}}, \quad (61)$$

де $\tilde{R}$ - розширена матриця для кореляційної матриці R_x n -ого порядку. Вона має вигляд:

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_n} \\ r_{yx_2} & r_{x_1x_2} & 1 & \dots & r_{x_2x_n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_nx_1} & r_{x_nx_2} & \dots & 1 \end{pmatrix}. \quad (62)$$

У цьому випадку

$$\frac{|\tilde{R}|}{|R_x|} = \tilde{\sigma}_{y \bullet x_1 x_2 \dots x_n}^2 \quad (63)$$

і рівність (61) можна записати так:

$$R_{y \bullet x_1 x_2 \dots x_n} = \sqrt{1 - \tilde{\sigma}_{y \bullet x_1 x_2 \dots x_n}^2}. \quad (64)$$

Множинний коефіцієнт кореляції $R_{y \bullet x_1 x_2 \dots x_n}$ характеризує тісноту кореляційного зв'язку між предиктантом та системою предикторів $x_1, x_2, \dots, x_n$, тобто є мірою адекватності моделі метеорологічного прогнозу $y = f(x_1, x_2, \dots, x_n)$, яка має вигляд лінійного рівняння множинної регресії. Чим ближчим до одиниці є множинний коефіцієнт кореляції, тим більш вірогідною буде регресійна модель метеорологічного прогнозу.

1.5.2 Перевірка статистичної гіпотези за допомогою критерію Фішера

Оцінити міру адекватності прогностичної моделі метеорологічного прогнозу можливо шляхом перевірки статистичної гіпотези H_0 про те, що відносна залишкова дисперсія $\tilde{\sigma}_{y \bullet x_1 x_2 \dots x_n}^2$ незначуще відрізняється від дисперсії предиктанта σ_y^2 .

Перевірка гіпотези H_0 відбувається за допомогою критерію Фішера, котрий формується таким чином

$$F = \frac{\sum_{j=1}^m (y_{\phi_j} - \bar{y}_{\phi})^2 / (m-1)}{\sum_{j=1}^m (y_{\phi_j} - y_{p_j})^2 / (m-k-1)}, \quad (65)$$

де m - об'єм вибірок;
 k - кількість предикторів, які включені в прогностичну модель;
 y_{ϕ} і y_p - відповідно фактичне та розрахункове значення предиктанта.

У рівнянні (65) у чисельнику розташована дисперсія предиктанта, а у знаменнику – залишкова дисперсія.

Гіпотеза H_0 не відхиляється, якщо $F < F_{\text{кр}}(\alpha, v_1, v_2)$ (для правосторонньої критичної області); $v_1 = m - 1$; $v_2 = m - k - 1$.

Вірогідність прогнозу на основі регресійної моделі можна також перевірити і за допомогою парного коефіцієнта кореляції між фактичним та розрахунковим значеннями предиктанта.

Зрозуміло, що модель тим краще відповідає величині, що прогнозується, чим ближчим за модулем до одиниці є цей коефіцієнт кореляції.

2 ЗАПИТАННЯ ДЛЯ САМОПЕРЕВІРКИ

1. Як називається регресійна модель метеорологічного прогнозу, якщо вона будується на множині предикторів?
2. Який метод використовується при визначенні коефіцієнтів лінійного множинного рівняння регресії?
3. Як побудувати регресійну модель метеорологічного прогнозу на статистично незалежних предикторах?
4. Який сенс має частинний коефіцієнт кореляції і яку суттєву інформацію він утримує?
5. Дати визначення множинного коефіцієнта кореляції і на яких етапах побудови регресійної моделі він використовується?
6. Які предиктори називаються потенційними, а які – оптимальними?
7. Які методи використовуються в основі добору оптимального складу предикторів при побудові регресійної моделі метеорологічного прогнозу?
8. У чому полягає процедура “просіювання” предикторів?
9. Який сенс методу покрокової регресії?
10. Як можна оцінити міру адекватності побудованої прогностичної регресійної моделі метеорологічного прогнозу?

3 ПОРЯДОК ВИКОНАННЯ РОБОТИ

1. Засвоїти основні теоретичні положення, що викладені у підручнику [1], с. 422-455 та на с. 4-20 цих методичних вказівок.
2. Проаналізувати наведені у підручнику [1], с. 495-503 та у цих методичних вказівках (Додаток А) приклади застосування регресійного аналізу для досліджень метеорологічних об'єктів.
3. Отримати у викладача варіант задачі (Додаток Б) для виконання практичної роботи та комп'ютерну програму "СТЕР" (призначення програми та особливості користування дивись у Додатку В).
4. Реалізувати на ПК алгоритм розв'язання задачі покрокової регресії та отримати:
 - середні значення та середні квадратичні відхили предикторів та предиктанта;
 - матрицю центрованих та нормованих значень предикторів та предиктанта;
 - матрицю кореляцій парних коефіцієнтів кореляцій між предиктантом і кожним з потенційних предикторів (r_{yx_i}) та між предикторами ($r_{x_i x_s}$);
 - фактичні значення критерію Стюдента для перевірки гіпотези про статистичну значущість отриманих парних коефіцієнтів кореляцій;
 - результати ітераційної покрокової процедури;
 - множинний коефіцієнт кореляції на кожному кроці.
5. Провести аналіз отриманих результатів та записати рівняння множинної лінійної регресії.
6. На рівні значущості $\alpha = 0,05$ оцінити міру адекватності побудованої прогностичної моделі метеорологічного прогнозу одним із відомих методів.
7. Оформити виконану роботу згідно з викладеним порядком та захистити у викладача.

ЛІТЕРАТУРА

1. Школьний Є.П., Лоева І.Д., Гончарова Л.Д. Обробка та аналіз гідрометеорологічної інформації. – Одеса, 1999. – 600 с.
2. Гончарова Л.Д., Школьний Є.П. Методи обробки та аналізу гідрометеорологічної інформації (збірник задач і вправ): Навчальний посібник. – Одеса, 2007. – 461 с.
3. Вентцель Е.С. Теория вероятностей. – М.: Наука, 1969. – 576 с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Выс. школа, 1977. – 480 с.
5. Митропольский А.К. Техника статистических вычислений. – М.: Гос. из-во физ. – мат. лит-ры, 1961. – 480 с.

ДОДАТОК А

Приклад застосування регресійного аналізу при дослідженні метеорологічних полів

Побудова лінійного рівняння множинної регресії

Побудувати лінійне рівняння множинної регресії, як моделі прогнозу опадів на ст.Одеса у лютому на центрованих та нормованих значеннях предикторів та предиктанта. На рівні значущості $\alpha = 0.05$ оцінити міру адекватності отриманої прогностичної моделі кліматичного прогнозу за допомогою парного коефіцієнта кореляції між фактичним та розрахунковим значеннями предиктанта.

В якості предикторів виступають часові ряди середнього місячного атмосферного тиску на двох станціях, що розташовані в Північній Атлантиці, зміна тиску на яких складає так зване Північно-Атлантичне коливання. Це ст. Рейк'явік (64° півн.ш., 22° зах.д., район Ісландської депресії) та ст. Понта-Дельгада (37°44' півн.ш., 25°40' зах.д., район Азорського антициклону).

Враховуючи те, що взимку західний перенос у помірних широтах добре виражений, а Північно-Атлантичне коливання є тим фізичним механізмом, від якого залежить формування кліматичних показників на території України, було вирішено використати в якості предикторів саме атмосферний тиск на вказаних станціях.

Із 12-ти характеристик баричного поля Північної Атлантики було визначено 6 найбільш впливових факторів, від яких залежить, на наш погляд, формування атмосферних опадів на ст. Одеса у лютому.

В табл. А.1 представлені вихідні дані для розв'язання поставленої задачі.

В якості 1-го, 2-го та 3-го предикторів виступають ряди атмосферного тиску на ст. Рейк'явік у листопаді (№1), грудні (№2), січні (№3), 4-го, 5-го та 6-го - ряди атмосферного тиску на ст. Понта-Дельгада у листопаді (№4), грудні (№5), січні (№6).

Оскільки прогноз опадів складається на лютий, то всі предиктори добиралися за попередній термін, враховуючи запізнення відгуків на 1-3 місяці на території України.

Реалізація алгоритму покрокової регресії проводилася на ПК за допомогою комп'ютерної програми "STEP". Ці результати наведені в табл. А.2.

Таблиця А.1 – Статистичні сукупності потенційних предикторів та предиктанта

Рік	№ п/п	Атмосферний тиск, гПа						Одеса, опади, мм
		ст. Рейк'явік			ст. Понта-Дельгада			
Місяці		XI	XII	I	XI	XII	I	II
1965	1	1017,7	997,5		1021	1024		
1966	2	1006,2	994	1009,2	1024	1026	1008	24
1967	3	1003	1008,7	1011	1023	1022	1019	88
1968	4	1004,5	1007,8	1001,5	1014	1020	1025	37
1969	5	1005,5	1001,5	1009,3	1019	1028	1010	135
1970	6	1003,5	1011,8	998,8	1021	1021	1006	48
1971	7	1008,2	997,9	999,8	1027,1	1023,1	1015,8	56
1972	8	1000,2	987,6	995	1023,7	1022,1	1020,8	5
1973	9	1010,7	1004,9	996,7	1019,8	1024,6	1021,7	58
1974	10	999,5	992,5	977,7	1021,3	1025,3	1021	11
1975	11	1000,5	1005,8	988,9	1023,2	1021,2	1023,4	5
1976	12	1000	1012,1	999,1	1023	1016,9	1024,8	11
1977	13	1002,6	999,6	1010,6	1017,3	1014,7	1018,6	52
1978	14	992,3	1005,9	996,9	1023,9	1011,7	1028,4	59
1979	15	995,5	991,2	1012	1021	1022	1014	23
1980	16	1014	997,9	1009,7	1015	1027,4	1017,8	4
1981	17	1005,4	1007,1	1005	1017,8	1018,3	1029,8	38
1982	18	993,6	991,9	1006,7	1023,3	1028,1	1017,7	33
1983	19	1010,2	997,3	986,1	1009,2	1020,8	1027,6	26
1984	20	998	993	989,1	1018,1	1023,8	1033,6	69
1985	21	1012	1007,2	1014,9	1015	1020,2	1014,7	66
1986	22	990,1	990,7	995,6	1025,7	1024,3	1032,2	55
1987	23	1005,2	1002,9	1013,6	1021,3	1012,6	1009,8	13
1988	24	1007,7	998,1	995,9	1013,5	1022,2	1024,8	15
1989	25	1008,4	1003,5	987,5	1017	1008,1	1027,4	10
1990	26	1009,5	994,7	981	1019	1023	1026,8	20
1991	27	995,6	1001,1	994,6	1022	1021,5	1022	24
1992	28	987,6	994,7	1004	1019,5	1023,7	1018	23
1993	29	991,7	993	984,2	1023,3	1029,9	1023,9	33
1994	30	997,9	990,5	995,2	1017,5	1022,8	1029,6	2
1995	31			992			1024,5	20

Таблиця А.2 – Результати реалізації алгоритму покрокової регресії

Середні предикторів та предиктанта						
1002.56	999.41	998.72	1019.98	1021.64	1021.22	35.43
Середньоквадратичні відхилення предикторів та предиктанта						
7.44	6.89	10.15	4.01	4.93	7.13	29.31
Матриця нормованих та центрованих значень предикторів та предиктанта						
1	2	3	4	5	6	y
2.04	-0.28	1.03	0.25	0.48	-1.85	-0.39
0.49	-0.79	1.21	1.00	0.88	-0.31	1.79
0.06	1.35	0.27	0.75	0.07	0.53	0.05
0.26	1.22	1.04	-1.49	-0.33	-1.57	3.40
0.40	0.30	0.01	-0.25	1.29	-2.13	0.43
0.13	1.80	0.11	0.25	-0.13	-0.76	0.70
0.76	-0.22	-0.37	1.78	0.30	-0.06	-1.04
-0.32	-1.71	-0.20	0.93	0.09	0.07	0.77
1.09	0.80	-2.07	-0.05	0.60	-0.03	-0.83
-0.41	-1.00	-0.97	0.33	0.74	0.31	-1.04
-0.28	0.93	0.04	0.80	-0.09	0.50	-0.83
-0.34	1.84	1.17	0.75	-0.96	-0.37	0.57
0.01	0.03	-0.18	-0.67	-1.41	1.01	0.80
-1.38	0.94	1.31	0.98	-2.02	-1.01	-0.42
-0.95	-1.19	1.08	0.25	0.07	-0.48	-1.07
1.54	-0.22	0.62	-1.24	1.17	1.20	0.09
0.38	1.12	0.79	-0.54	-0.68	-0.49	-0.08
-1.20	-1.09	-1.24	0.83	1.31	0.89	-0.32
1.03	-0.31	-0.95	-2.69	-0.17	1.74	1.15
-0.61	-0.93	1.59	-0.47	0.44	-0.91	1.04
1.27	1.13	-0.31	-1.24	-0.29	1.54	0.67
-1.68	-1.27	1.47	1.43	0.54	-1.60	-0.77
0.36	0.51	-0.28	0.33	-1.84	0.50	-0.70
0.69	-0.19	-1.11	-1.62	0.11	0.87	-0.87
0.79	0.59	-1.75	-0.74	-2.75	0.78	-0.53
0.93	-0.68	-0.41	-0.25	0.28	0.11	-0.39
-0.94	0.25	0.52	0.50	-0.03	-0.45	-0.42
-2.01	-0.68	-1.43	-0.12	0.42	0.38	-0.08
-1.46	-0.93	-0.35	0.83	1.68	1.17	-1.14
-0.63	-1.29	-0.66	-0.62	0.24	0.46	-0.53
Сумісна матриця R_x коеф. кореляції та коефіцієнтів значущості t_{ij}						
1.00	0.30	-0.13	-0.45	-0.08	0.11	0.19
1.80	1.00	0.10	-0.13	-0.49	-0.09	0.24
0.69	0.52	1.00	0.25	-0.01	-0.59	0.33
3.03	0.69	1.46	1.00	0.14	-0.33	-0.38
0.41	3.42	0.05	0.75	1.00	-0.07	-0.06
0.58	0.51	4.93	2.02	0.36	1.00	-0.19
1.05	1.35	2.01	2.43	0.33	1.06	1.00

КРОК 1-й Вільний член $A[0] = -1.0153739162E-10$ Коефіцієнти регресії: $A[1] = -0.384$ Коэф. значущості $t_a = 2.201$ Множинний коэф. кореляції $Rmk = 0.384$	Предиктор 4
КРОК 2-й Вільний член $A[0] = -1.0169853055E-10$ Коефіцієнти регресії: $A[1] = -0.501$ Коэф. значущості $t_a = 3.275$ $A[2] = 0.460$ Коэф. значущості $t_a = 3.006$ Множинний коэф. кореляції $Rmk = 0.587$	Предиктор 4 Предиктор 3
КРОК 3-й Вільний член $A[0] = -1.2021576358E-10$ Коефіцієнти регресії: $A[1] = -0.479$ Коэф. значущості $t_a = 3.177$ $A[2] = 0.441$ Коэф. значущості $t_a = 2.926$ $A[3] = 0.134$ Коэф. значущості $t_a = 0.889$ Множинний коэф. кореляції $Rmk = 0.602$	Предиктор 4 Предиктор 3 Предиктор 2
КРОК 4-й Вільний член $A[0] = -1.4530330937E-10$ Коефіцієнти регресії: $A[1] = -0.487$ Коэф. значущості $t_a = 3.244$ $A[2] = 0.440$ Коэф. значущості $t_a = 2.931$ $A[3] = 0.181$ Коэф. значущості $t_a = 1.209$ $A[4] = 0.099$ Коэф. значущості $t_a = 0.660$ Множинний коэф. кореляції $Rmk = 0.608$	Предиктор 4 Предиктор 3 Предиктор 2 Предиктор 5
КРОК 5-й Вільний член $A[0] = -1.6885573405E-10$ Коефіцієнти регресії: $A[1] = -0.508$ Коэф. значущості $t_a = 3.410$ $A[2] = 0.381$ Коэф. значущості $t_a = 2.554$ $A[3] = 0.168$ Коэф. значущості $t_a = 1.129$ $A[4] = 0.088$ Коэф. значущості $t_a = 0.588$ $A[5] = -0.110$ Коэф. значущості $t_a = 0.741$ Множинний коэф. кореляції $Rmk = 0.614$	Предиктор 4 Предиктор 3 Предиктор 2 Предиктор 5 Предиктор 6
КРОК 6-й Вільний член $A[0] = -1.8221144925E-10$ Коефіцієнти регресії: $A[1] = -0.523$ Коэф. значущості $t_a = 3.510$ $A[2] = 0.378$ Коэф. значущості $t_a = 2.540$ $A[3] = 0.179$ Коэф. значущості $t_a = 1.203$ $A[4] = 0.092$ Коэф. значущості $t_a = 0.619$ $A[5] = -0.112$ Коэф. значущості $t_a = 0.749$ $A[6] = -0.034$ Коэф. значущості $t_a = 0.231$ Множинний коэф. кореляції $Rmk = 0.615$	Предиктор 4 Предиктор 3 Предиктор 2 Предиктор 5 Предиктор 6 Предиктор 1

Відповідно до порядку виконання роботи, були реалізовані всі етапи: від підготовки вихідної інформації до розрахунків необхідних параметрів задачі.

Аналіз фактичних значень критерію Стюдента вказує на те, що між деякими предикторами зафіксовано статистично значущий зв'язок і побудувати модель на статистично незалежних предикторах неможливо. За визначеною методикою критичне значення критерію Стюдента за умов рівня значущості $\alpha = 0.05$ і числа степенів вільності $\nu = m - 1 = 30 - 1 = 29$, дорівнює: $t_{кр}(\alpha, \nu) = t_{кр}(0.05, 29) = 2.05$.

Отже виходить, що для трьох парних коефіцієнтів кореляції, що виражають лінійний кореляційний зв'язок між окремими предикторами, $t > t_{кр}(\alpha, \nu)$, а це означає, що гіпотеза H_0 відхиляється (с. 211-212 [2]).

Розрахункова частина, що представлена у першому розділі цих методичних вказівок та підручника [1] (с. 422-445, с. 449-455), розпочинається з аналізу вектора парних коефіцієнтів кореляції між предиктантом та переліченими потенційними предикторами:

$$R_{yx} = \begin{pmatrix} 0,188 \\ 0,236 \\ 0,332 \\ -\mathbf{0,384} \\ -0,060 \\ -0,189 \end{pmatrix}$$

Аналіз координат цього вектора вказує на те, що найбільшим за модулем є парний коефіцієнт кореляції між предиктантом та 4-тим предиктором. Він ставиться на перше місце ($x_4 \rightarrow x_1$), а ті, що залишилися, перенумеровуються. Множинний коефіцієнт кореляції першого кроку дорівнює $R_{y \bullet x_4}^{(1)} = 0,384$.

- Рівняння регресії першого кроку має вигляд:

$$y^{(1)} = -0,384x_4.$$

Враховуючи методику “просіювання” предикторів, проаналізуємо наступні кроки ітераційного процесу.

На другому кроці множинний коефіцієнт кореляції зростає і вже приймає значення $R_{y \bullet x_4 x_3}^{(2)} = 0,587$, що вказує на суттєвий вплив і 3-го предиктора на предиктант.

- Рівняння другого кроку має вигляд:

$$y^{(2)} = -0,501x_4 + 0,460x_3.$$

Аналіз результатів ітераційної процедури покрокової регресії (табл. А.2) вказує на те, що на 5-6-ому кроках ітераційний процес усталюється.

Дійсно, невеликі різниці між коефіцієнтами регресії 5-го та 6-го кроків виявляються у третьому знаку, коефіцієнти Стюдента, на основі яких проводилася перевірка гіпотези про статистичну значущість коефіцієнтів моделі на кожному кроці, змінюються мало і вони є високими, порівняно з $t_{кр}(0.05, 29) = 2.05$.

Задовільну адекватність показує прогностична модель з предикторами, для яких коефіцієнти Стюдента $t_a > 2,05$.

Множинний коефіцієнт кореляції майже не змінюється, тобто на 5-6-ому кроках настає його насичення (рис. А.1).

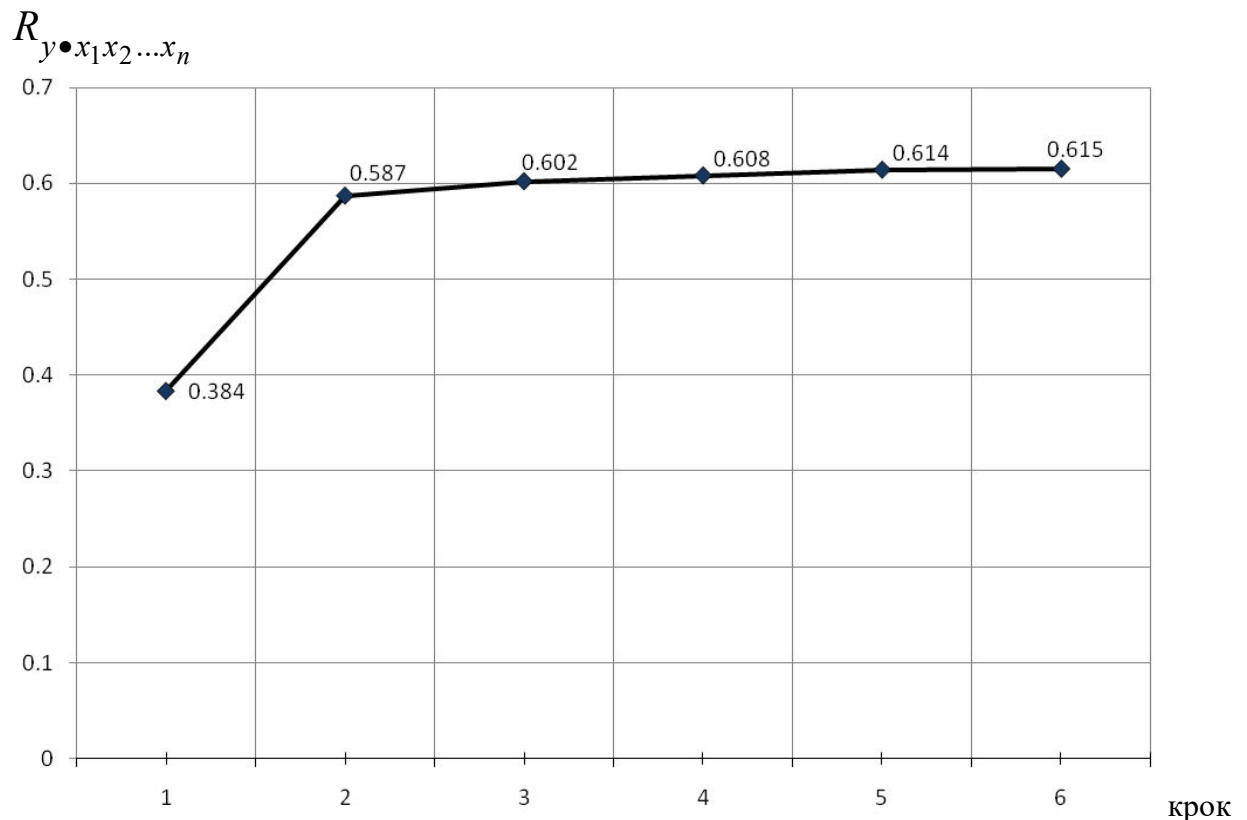


Рис. А.1 – Зміна множинного коефіцієнта кореляції

Отже, враховуючи всі вище перелічені умови, як модель прогнозу опадів на ст. Одеса у лютому буде побудована на двох оптимальних предикторах, а параметри моделі відповідають 6-ому кроку ітераційного процесу.

Таким чином, ми отримали таку прогностичну модель для опадів на ст. Одеса у лютому:

$$\hat{y}^{(6)} = 0,378x_3 - 0,523x_4.$$

В рівнянні регресії предиктори та предиктант є центрованими та нормованими на середній квадратичний відхил.

Значення множинного коефіцієнта кореляції $R_{y \bullet x_1 x_2 \dots x_6} = 0,615$ свідчить про те, що адекватність моделі є задовільною. Удосконалюємося в цьому ще й шляхом перевірки статистичної гіпотези про значущість парного коефіцієнта кореляції між фактичним та розрахунковим значеннями предиктанта (табл. А.3).

Таблиця А.3 – Статистичні сукупності фактичного (y_j) та модельного ($\hat{y}_j$) значень предиктанта

Предиктант	1	2	3	4	5	6	7	8	9	10
y_j	-0.39	1.79	0.05	3.40	0.43	0.70	-1.04	0.77	-0.83	-1.04
$\hat{y}_j^{(6)}$	0.26	-0.06	-0.29	1.17	0.13	-0.09	-1.07	-0.56	-0.76	-0.54

Предиктант	11	12	13	14	15	16	17	18	19	20
y_j	-0.83	0.57	0.80	-0.42	-1.07	0.09	-0.08	-0.32	1.15	1.04
$\hat{y}_j^{(6)}$	-0.41	0.05	0.28	-0.02	0.28	0.88	0.58	-0.90	1.05	0.85

Предиктант	21	22	23	24	25	26	27	28	29	30
y_j	0.67	-0.77	-0.70	-0.87	-0.53	-0.39	-0.42	-0.08	-1.14	-0.53
$\hat{y}_j^{(6)}$	0.53	-0.20	-0.28	0.43	-0.27	-0.02	-0.06	-0.48	-0.56	0.07

Розрахунки $r_{y\hat{y}}$ показали, що він дорівнює:

$$r_{y\hat{y}} = 0,58.$$

Оскільки об'єми вибірок є невеликими ($m=30$), то перевірку гіпотези будемо проводити наступним чином.

Гіпотеза H_0 : $r_{yy}^{\wedge} = 0,58$ на рівні значущості $\alpha = 0.05$ є статистично незначущим

Гіпотеза H_1 : Альтернативна гіпотеза.

Гіпотеза H_0 перевіряється за допомогою z -перетворення Фішера:

$$\hat{z} = \frac{1}{2} \ln \frac{1+r_{yy}^{\wedge}}{1-r_{yy}^{\wedge}} = \frac{1}{2} \ln \frac{1+0.58}{1-0.58} = 0.66,$$

$$\sigma_z = \frac{1}{\sqrt{m-3}} = \frac{1}{\sqrt{30-3}} = \frac{1}{\sqrt{27}} = 0.19.$$

Фактичне значення критерію Стюдента треба порівняти з $t_{кр}(\alpha, \nu)$

$$t = \frac{\left| \hat{z} \right|}{\sigma_z} = \frac{0.66}{0.19} = 3.49, \quad \text{а} \quad t_{кр}(0.05, 29) = 2.05.$$

$t > t_{кр}(\alpha, \nu)$ - таке співвідношення вказує на те, що $\hat{z}$ -перетворення з ймовірністю 0,95 є статистично значущим. Якщо $\hat{z}$ -перетворення є статистично значущим, то й коефіцієнт кореляції $r_{yy}^{\wedge} = 0,58$ з ймовірністю 95% є статистично значущим.

Висновок: лінійне рівняння множинної регресії, як моделі прогнозу опадів у лютому на ст. Одеса, побудоване на двох оптимальних предикторах (із 6 потенційних) і воно має вигляд:

$$\hat{y}^{(6)} = 0,378x_3 - 0,523x_4.$$

Опади на ст. Одеса у лютому залежать від атмосферного тиску на ст. Рейк'явік у січні та від атмосферного тиску на ст. Понта-Дельгада у листопаді. Вірогідність побудованої моделі складає 95%. Уточнення моделі проводилося на кожному кроці на основі перевірки статистичної гіпотези про значущість отриманих коефіцієнтів регресії.

ДОДАТОК Б
Вихідні дані для виконання роботи

Таблиця Б.1 – Місячна сума опадів (травень)

Рік	Одеса	Київ	Харків	Львів	Вінниця	Донецьк
1965	14	36	37	53	58	46
1966	30	66	30	80	64	72
1967	29	43	23	82	123	39
1968	7	52	30	69	62	32
1969	59	59	25	80	36	37
1970	134	84	64	141	69	122
1971	43	30	32	75	49	46
1972	30	47	37	107	67	26
1973	52	85	70	73	134	46
1974	63	45	59	111	69	87
1975	16	15	26	52	58	98
1976	51	60	26	68	39	31
1977	19	14	92	23	79	54
1978	62	46	103	172	55	116
1979	46	28	1	44	47	8
1980	40	43	46	54	55	66
1981	46	26	39	91	40	25
1982	9	25	20	38	14	11
1983	24	108	78	40	133	70
1984	5	63	19	100	65	31
1985	34	87	64	123	53	104
1986	5	17	55	52	13	36
1987	32	91	35	81	65	31
1988	46	75	82	57	67	9
1989	13	29	32	149	59	32
1990	48	46	78	89	44	33
1991	80	117	104	169	78	63
1992	34	73	85	94	35	132
1993	35	25	27	44	44	71
1994	18	61	103	57	54	55
1995	29	106	20	79	50	29
1996	23	28	39	92	39	59
1997	31	48	75	132	37	32
1998	34	31	41	65	43	58
1999	37	48	43	27	37	38
2000	28	73	46	66	47	30
2001	56	33	28	55	44	75
2002	2	65	48	91	74	25
2003	4	49	12	104	30	2
2004	104	53	102	85	31	72
2005	39	60	26	89	72	5

Таблиця Б.2 – Місячна сума опадів (червень)

Рік	Одеса	Київ	Харків	Львів	Вінниця	Донецьк
1965	16	44	46	107	89	10
1966	43	51	68	127	28	45
1967	20	74	68	161	87	97
1968	31	18	42	91	26	33
1969	66	81	23	180	86	163
1970	46	112	14	81	85	60
1971	48	78	54	91	152	64
1972	35	46	74	109	63	22
1973	20	17	26	110	42	59
1974	29	77	77	167	113	48
1975	33	64	45	106	140	15
1976	45	40	61	66	37	112
1977	38	92	88	55	67	68
1978	46	98	67	97	115	65
1979	37	10	7	87	14	8
1980	78	131	88	133	193	38
1981	11	49	15	85	75	41
1982	54	75	51	135	103	33
1983	37	62	39	64	82	44
1984	128	140	25	98	76	15
1985	63	84	119	161	175	52
1986	60	97	65	94	64	57
1987	16	67	75	58	94	145
1988	32	104	110	96	162	239
1989	65	162	111	157	119	47
1990	53	77	80	44	120	74
1991	61	97	43	65	109	16
1992	58	29	46	80	45	49
1993	54	87	82	83	83	31
1994	65	88	38	48	66	47
1995	42	55	78	57	129	140
1996	15	73	74	79	70	29
1997	99	83	95	55	134	97
1998	38	63	13	168	41	22
1999	57	47	25	71	16	41
2000	58	65	13	71	92	71
2001	56	152	116	157	155	190
2002	83	190	50	114	144	24
2003	44	26	89	54	28	109
2004	37	7	29	33	28	120
2005	51	111	104	84	58	56

Таблиця Б.3 – Середня місячна температура повітря (травень)

Рік	Одеса	Київ	Харків	Львів	Вінниця	Донецьк
1965	13.2	12.7	13.4	10.5	12.2	13.7
1966	15.5	16.4	16.9	14.2	15.0	15.4
1967	15.8	17.5	18.5	13.7	15.3	18.4
1968	17.4	15.1	16.7	13.3	14.9	17.9
1969	14.4	14.8	14.5	15.0	14.6	14.6
1970	14.4	15.0	16.0	11.9	13.7	15.5
1971	15.9	16.1	15.8	15.0	15.2	15.5
1972	16.1	16.0	16.3	13.9	14.8	16.4
1973	14.5	14.4	15.3	12.9	13.7	15.1
1974	13.3	12.6	12.8	11.1	12.0	13.5
1975	17.4	19.0	18.9	15.0	16.8	18.5
1976	13.7	12.6	12.9	12.2	12.2	13.4
1977	14.6	15.2	15.9	13.3	14.0	15.2
1978	13.5	12.6	12.8	11.3	11.7	12.4
1979	18.2	17.6	19.6	14.6	15.8	20.0
1980	12.7	11.3	12.5	9.7	11.0	13.4
1981	14.6	16.1	15.7	13.3	14.3	14.9
1982	15.5	15.1	14.2	13.8	14.8	14.7
1983	17.2	17.9	17.2	15.5	16.1	17.4
1984	15.1	16.7	18.7	13.4	14.5	17.9
1985	15.8	17.0	17.9	14.9	15.7	17.8
1986	16.5	16.7	15.2	14.7	14.9	14.9
1987	12.5	14.2	15.4	11.7	13.2	14.6
1988	15.0	15.2	14.8	13.7	14.3	14.6
1989	15.3	15.2	14.5	13.6	13.8	14.1
1990	14.9	14.6	13.6	12.9	13.7	13.9
1991	13.8	12.8	13.7	10.5	11.7	13.7
1992	14.7	13.6	13.6	11.9	12.4	13.6
1993	15.2	16.5	15.8	16.0	15.6	15.1
1994	15.6	13.1	13.1	12.6	13.4	13.6
1995	14.1	14.0	15.8	12.1	12.6	15.4
1996	17.5	18.6	18.9	16.1	17.6	18.6
1997	16.6	16.1	16.2	14.0	15.2	16.3
1998	15.2	15.2	15.9	13.1	13.4	15.4
1999	14.0	12.8	12.8	12.1	12.1	12.1
2000	16.7	15.4	14.0	14.7	14.9	13.7
2001	14.8	14.2	13.8	14.1	13.3	13.6
2002	16.9	16.5	15.5	16.5	15.9	15.6
2003	18.4	19.4	18.8	16.9	18.6	19.1
2004	14.2	13.2	13.8	11.9	12.7	14.3
2005	16.9	16.4	17.9	13.0	14.7	18.2

Таблиця Б.4 – Середня місячна температура повітря (червень)

Рік	Одеса	Київ	Харків	Львів	Вінниця	Донецьк
1965	19.9	17.6	18.4	16.3	16.9	19.5
1966	17.4	16.9	17.4	15.9	15.8	17.4
1967	19.1	18.2	18.5	16.4	17.3	18.0
1968	19.8	20.1	20.0	18.3	19.0	19.9
1969	19.2	17.3	18.6	15.6	16.4	19.3
1970	18.4	17.2	17.6	16.0	16.5	17.5
1971	18.9	18.1	19.1	15.7	16.8	18.7
1972	21.3	19.7	22.5	16.8	18.3	23.0
1973	18.6	17.9	18.5	15.2	16.3	17.7
1974	18.6	17.0	17.9	14.1	16.1	18.1
1975	21.9	20.7	23.3	16.5	18.7	23.1
1976	18.4	16.2	16.6	14.9	15.6	16.7
1977	18.2	17.2	17.3	16.4	16.9	17.6
1978	18.4	16.6	16.4	14.9	16.0	16.6
1979	21.2	21.5	21.3	19.0	19.9	21.0
1980	18.1	17.1	17.9	14.8	16.0	18.2
1981	21.4	21.1	22.8	17.4	19.6	22.3
1982	18.7	16.9	16.5	15.8	16.2	16.8
1983	19.6	18.2	18.1	16.0	16.7	18.4
1984	17.8	15.7	18.0	13.9	14.5	18.6
1985	18.0	16.7	17.4	14.0	15.7	17.6
1986	20.7	19.3	20.5	16.3	17.7	18.6
1987	19.3	18.2	18.7	16.5	17.4	19.9
1988	19.3	18.3	19.6	15.5	16.8	19.2
1989	19.0	19.4	20.6	15.3	17.0	19.7
1990	19.0	16.9	16.3	15.8	16.3	17.0
1991	20.0	18.5	20.7	16.0	17.4	20.8
1992	18.9	18.7	18.8	17.6	17.4	18.5
1993	18.7	16.7	16.5	15.4	15.8	16.9
1994	18.8	16.3	16.4	15.6	15.6	16.5
1995	21.1	20.1	21.7	16.9	18.2	21.6
1996	20.5	18.8	18.9	16.8	17.9	19.0
1997	19.7	18.3	19.5	16.7	17.2	19.4
1998	21.1	19.8	22.0	17.6	18.6	21.6
1999	22.8	22.6	22.8	18.7	20.8	22.2
2000	19.8	17.9	17.8	16.9	17.2	17.9
2001	18.2	16.7	16.9	14.7	15.7	17.1
2002	20.4	18.4	18.9	17.0	17.0	19.2
2003	20.5	18.0	17.3	17.4	17.7	17.5
2004	19.0	17.7	17.3	16.2	16.5	17.0
2005	19.3	17.3	17.6	16.3	16.6	18.5

Таблиця Б.5 – Середньомісячні значення індексу
Північно-Атлантичного коливання (ПАК)

№ п/п	Рік	Місяць				
		I	II	III	IV	V
1	1965	-0.12	-1.55	-1.51	0.72	-0.62
2	1966	-1.74	-1.39	0.56	-0.75	0.22
3	1967	-0.89	0.19	1.51	0.18	-0.99
4	1968	0.13	-1.29	0.4	-1.08	-1.76
5	1969	-0.83	-1.55	-1.56	1.53	0.55
6	1970	-1.5	0.64	-0.96	-1.3	1.14
7	1971	-1.13	0.24	-0.84	-0.24	0.5
8	1972	0.27	0.32	0.72	-0.22	0.95
9	1973	0.04	0.85	0.3	-0.54	-0.44
10	1974	1.34	-0.14	-0.03	0.51	-0.24
11	1975	0.58	-0.62	-0.61	-1.6	-0.52
12	1976	-0.25	0.93	0.75	0.26	0.96
13	1977	-1.04	-0.49	-0.81	0.65	-0.86
14	1978	0.66	-2.2	0.7	-1.17	1.08
15	1979	-1.38	-0.67	0.78	-1.71	-1.03
16	1980	-0.75	0.05	-0.31	1.29	-1.5
17	1981	0.37	0.92	-1.19	0.36	0.2
18	1982	-0.89	1.15	1.15	0.1	-0.53
19	1983	1.59	-0.53	0.95	-0.85	-0.07
20	1984	1.66	0.72	-0.37	-0.28	0.54
21	1985	-1.61	-0.49	0.2	0.32	-0.49
22	1986	1.11	-1	1.71	-0.59	0.85
23	1987	-1.15	-0.73	0.14	2	0.98
24	1988	1.02	0.76	-0.17	-1.17	0.63
25	1989	1.17	2	1.85	0.28	1.38
26	1990	1.04	1.41	1.46	2	-1.53
27	1991	0.86	1.04	-0.2	0.29	0.08
28	1992	-0.13	1.07	0.87	1.86	2.63
29	1993	1.6	0.5	0.67	0.97	-0.78
30	1994	1.04	0.46	1.26	1.14	-0.57
31	1995	0.93	1.14	1.25	-0.85	-1.49
32	1996	-0.12	-0.07	-0.24	-0.17	-1.06
33	1997	-0.49	1.7	1.46	-1.02	-0.28
34	1998	0.39	-0.11	0.87	-0.68	-1.32
35	1999	0.77	0.29	0.23	-0.95	0.92
36	2000	0.6	1.7	0.77	-0.03	1.58
37	2001	0.25	0.45	-1.26	0	-0.02
38	2002	0.44	1.1	0.69	1.18	-0.22
39	2003	0.16	0.62	0.32	-0.18	0.01
40	2004	-0.29	-0.14	1.02	1.15	0.19
41	2005	1.52	-0.06	-1.83	-0.3	-1.25

Таблиця Б.6 – Середньомісячні значення індексу
Північно-Каспійського коливання (ПКК)

№ п/п	Рік	Місяць				
		I	II	III	IV	V
1	1965	-0.82	0.86	-0.33	0.62	-0.75
2	1966	-0.78	-1.08	-0.16	-0.2	0.23
3	1967	-0.08	0.55	-0.09	0.39	-1.3
4	1968	0.2	-0.63	-0.09	0.05	-1.63
5	1969	0.57	-0.76	0.02	-0.22	-0.25
6	1970	-0.48	-1.65	-1.72	-2.09	0.66
7	1971	-0.99	1.08	-0.83	0.78	0.37
8	1972	1.04	-0.18	0.68	-0.74	-0.67
9	1973	1.69	-0.47	1.37	-0.59	0.04
10	1974	0.84	-0.54	-0.32	2.03	-0.55
11	1975	-0.11	2.0	-0.35	-1.49	-0.33
12	1976	-0.04	1.44	0.16	0.56	-0.19
13	1977	-0.02	-1.49	-0.5	-1.8	-0.23
14	1978	-0.43	-1.04	-1.31	0.65	1.21
15	1979	-1.39	-0.23	-1.8	-0.38	-1.46
16	1980	0.29	0.61	-0.55	0.71	-0.23
17	1981	-0.26	0.34	-0.6	1.56	0.3
18	1982	0.49	1.33	0.14	-0.03	-0.76
19	1983	0.69	0.04	0.05	-1.24	-1.5
20	1984	-1.7	0.33	-0.3	1.22	-0.46
21	1985	-0.87	1.2	-0.28	-0.68	-0.16
22	1986	-1.36	-0.39	-0.2	-2.06	0.4
23	1987	0.07	-0.36	-0.24	1.77	-1.06
24	1988	-0.51	-1.21	-1.17	-0.06	1.03
25	1989	2.3	0.18	-0.73	-0.8	1.7
26	1990	0.55	-0.72	0.94	0.48	1.11
27	1991	0.9	-0.21	0.31	-0.34	1.26
28	1992	2.39	1.27	0.07	0.05	2.21
29	1993	0.65	1.8	1.15	0.71	1.07
30	1994	-0.96	0.65	-0.78	-0.52	-0.01
31	1995	-0.9	-1.28	-1.68	-0.1	-0.59
32	1996	0.43	-0.72	0.44	0.93	-1.2
33	1997	1.71	-0.02	1.47	0.61	-0.19
34	1998	0.71	1.48	0.74	-1.61	0.68
35	1999	-0.86	-0.54	-0.4	0.21	1.15
36	2000	0.65	-0.22	0.48	-0.7	1.15
37	2001	-0.49	0.04	-1.85	-1.17	1.13
38	2002	0.67	-0.87	-0.06	1.36	1.25
39	2003	-0.48	1.1	1.69	0.94	-0.05
40	2004	-0.94	0.29	-0.18	0.66	-0.1
41	2005	0.17	0.08	0.29	-0.33	-1.1

ДОДАТОК В

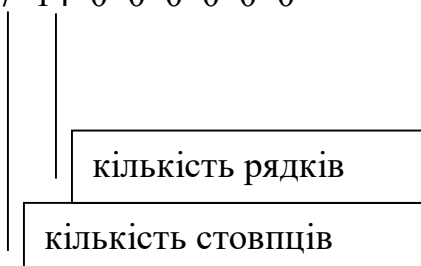
Особливості формування файлу вихідних даних для розрахунків на ПК

Зразок файлу вихідних даних до програми “STEP”:

7 14 0 0 0 0 0 0

1017.7	997.5	1009.2	1021	1024	1008.0	3.4
1006.2	994.0	1011.0	1024.0	1026.0	1019.0	-2.6
1003.0	1008.7	1001.5	1023.0	1022.0	1025.0	-0.5
1004.5	1007.8	1009.3	1014.0	1020.0	1010.0	-3.5
1005.5	1001.5	998.8	1019.0	1028.0	1006.0	0.2
1003.5	1011.8	999.8	1021.0	1021.0	1015.8	-0.5
1008.2	997.9	995.0	1027.1	1023.1	1020.8	-2.8
1000.2	987.6	996.7	1023.7	1022.1	1021.7	1.0
1010.7	1004.9	977.7	1019.8	1024.6	1021.0	0.8
999.5	992.5	988.9	1021.3	1025.3	1023.4	-0.4
1000.5	1005.8	999.1	1023.2	1021.2	1024.8	-5.5
1000.0	1012.1	1010.6	1023.0	1016.9	1018.6	3.5
1002.6	999.6	996.9	1017.3	1014.7	1028.4	-0.6
992.3	1005.9	1012.0	1023.9	1011.7	1014.0	-1.0

Обов’язкова інформація: 7 14 0 0 0 0 0 0



Приклад для 6 предикторів та одного предиктанта з об’ємами вибірок $m=14$.
Останній стовпчик (7-й) – предиктант.

ДОДАТОК Г
Значення критерію Стьюдента
 для рівня значущості α і числа степенів вільності ν

ν	Двостороння критична область, α							
	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
6	1.44	1.94	2.45	3.14	3.71	4.32	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.03	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	3.83	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	3.69	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	3.58	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	3.50	4.02	4.44
12	1.36	1.78	2.20	2.68	3.05	3.43	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.37	3.85	4.22
14	1.34	1.76	2.14	2.62	2.98	3.33	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.29	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.25	3.69	4.02
17	1.33	1.74	2.11	2.57	2.90	3.22	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.20	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.17	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.15	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.14	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.12	3.51	3.79
23	1.32	1.71	2.07	2.50	2.81	3.10	3.48	3.77
24	1.32	1.71	2.06	2.49	2.90	3.09	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.08	3.45	3.73
26	1.32	1.71	2.06	2.48	2.78	3.07	3.44	3.71
27	1.31	1.70	2.05	2.47	2.77	3.06	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.05	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.04	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.03	3.39	3.65
40	1.30	1.68	2.02	2.42	2.70	2.97	3.31	3.55
60	1.30	1.67	2.00	2.39	2.66	2.91	3.23	3.46
120	1.29	1.66	1.98	2.36	2.62	2.85	3.16	3.37
∞	1.28	1.64	1.96	2.33	2.58	2.81	3.09	3.29

ДОДАТОК Д
Значення критерію Фішера для рівня значущості 0,05

ν_1 – число степенів вільності для більшої дисперсії

ν_2 – число степенів вільності для меншої дисперсії

ν_2	ν_1							
	15	20	24	30	40	60	120	∞
10	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.538
11	2.179	2.646	2.609	2.571	2.531	2.490	2.448	2.405
12	2.617	2.544	2.506	2.466	2.426	2.384	2.341	2.296
13	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.206
14	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.131
15	2.404	2.328	2.288	2.247	2.204	2.160	2.114	2.066
16	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.010
17	2.308	2.230	2.190	2.148	2.104	2.058	2.010	1.960
18	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.917
19	2.234	2.156	2.114	2.071	2.026	1.980	1.930	1.878
20	2.203	2.124	2.083	2.039	1.994	1.946	1.896	1.843
21	2.176	2.096	2.054	2.010	1.965	1.917	1.866	1.812
22	2.151	2.071	2.028	1.984	1.938	1.890	1.838	1.783
23	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.757
24	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.733
25	2.089	2.008	1.964	1.919	1.872	1.822	1.768	1.711
26	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.691
27	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.672
28	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.654
29	2.028	1.945	1.901	1.854	1.806	1.754	1.698	1.638
30	2.015	1.932	1.987	1.841	1.792	1.740	1.684	1.622
40	1.925	1.839	1.793	1.744	1.693	1.637	1.577	1.509
60	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.389
120	1.751	1.659	1.608	1.554	1.495	1.429	1.352	1.254
∞	1.666	1.571	1.517	1.459	1.394	1.318	1.221	1.000

МЕТОДИЧНІ ВКАЗІВКИ

для СРС та виконання практичних занять

з дисципліни

*”Багатовимірний аналіз метеорологічних
полів та атмосферних процесів”*

Укладачі: к.г.н., доцент Гончарова Л.Д.;
к.г.н., доцент Галич Є.А.;
д.т.н., професор Шкільний Є.П.

Підп. до друку
Умовн. друк. арк..

Формат
Тираж

Папір
Зам. №

Надруковано з готового оригінал-макета

Одеський державний екологічний університет
65016, Одеса, вул. Львівська, 15
