

## МАТРИЧНА МНОЖИННА РЕГРЕСІЯ ТА СУЧАСНІ МЕТОДИ БІОМЕТРІЇ ДЛЯ ПРОГНОЗУВАННЯ БІОЛОГІЧНИХ ПОКАЗНИКІВ: ПРИКЛАДИ

І.М. НАЗАРАГА, Я.Р. НАЗАРАГА

**Анотація.** Розглянуто приклади прогнозування біологічних показників із застосуванням сучасних методів біометрії та методів на основі матричної множинної регресії. З метою розв'язання задачі оцінювання методом найменших квадратів для множинної матричної регресії використано розроблений математичний апарат сингулярного розкладу і техніку псевдообернення за Муром–Пенроузом у межах розвитку концепції кортежних операторів. Емпіричними даними для розрахунків стали дані експерименту, проведеного в навчально-науковому центрі «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка. Розрахунки виконано у Microsoft Office Excel та в середовищі Wolfram Mathematica, проаналізовано результати за різними варіантами. Точність прогнозування за критерієм APE (абсолютний відсоток похибки) згідно з алгоритмом на основі матричної множинної регресії (похибка від 0% до 10%) вища, ніж для сучасних методів біометрії (деякі похибки перевищують 30%). Показано, що матрична множинна регресія може бути ефективним інструментом прогнозування в біології з прийнятною для цього точністю.

**Ключові слова:** матрична множинна регресія, методи біометрії, біологічні показники, прогнозування, сингулярний розклад.

### ВСТУП

Сучасна біологія широко використовує математичний апарат для моделювання та прогнозування процесів у живих системах і формалізації механізмів, покладених в основу біологічних процесів. Для передбачення поведінки таких систем дедалі частіше вдаються до віртуальних експериментів із використанням математичних засобів, що дає змогу контролювати, вимірювати чи прогнозувати ключові змінні, а також окремо розглядати пов'язані процеси, які в реальних експериментах розділити неможливо. У зв'язку з цим полегшується інтерпретація отриманих результатів і з'являється можливість проводити віртуальні експерименти без участі живих істот, які в житті неприпустимі з етичних міркувань.

Таким чином, реалізація та апробація математичних алгоритмів прогнозування біологічних показників є актуальними завданнями для сучасної біологічної науки.

### АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ ТА ПОСТАНОВКА ПРОБЛЕМИ

Сучасними методами, які найчастіше використовувалися в біології, є методи біометрії (біологічної статистики) [1–3], але завдяки біохімії і біофізиці дедалі популярнішими стають складні математичні моделі та алгоритми.

Натепер є багато причин збільшення кількості різноманітних математичних методів у біологічних науках [4], одними з яких є збільшення кількості нової експериментальної інформації, яку неможливо систематизувати без нового математичного апарату, а також пошук нових аспектів та знань, що впливають після застосувань такого апарату до відомих положень та законів біології, сформульованих без застосування математики. Застосування таких математичних методів та моделей дає змогу не тільки обробляти дані реальних дослідів, а і прогнозувати перебіг біологічних процесів під час віртуальних експериментів [5]. При цьому виникає потреба у створенні обчислювальних алгоритмів і розробленні програмного забезпечення, що надійно і ефективно використовує розроблені алгоритми [5].

У працях [6–11] запропоновано нові методи аналізування об'єктів як складних багаторозмірних табличних структур (матриць) і досліджується сучасний математичний апарат, який дозволяє оперувати із цими структурами. Зокрема, В.С. Донченком та його учнями у працях [8–10] запропоновано концепцію кортежних операторів, що дозволяє перенести техніки сингулярного розкладу і псевдообернення на евклідові простори матриць фіксованої розмірності.

Таким чином, для розв'язання задач прогнозування запропоновано до розгляду клас матричних функцій набору матричних аргументів (матричну множинну регресію), а також алгоритм оцінювання методом найменших квадратів для вектора невідомих параметрів згаданого класу матричних функцій [9]. Застосування такого підходу для прогнозування макроекономічних показників економіки України продемонстровано у праці [11], де на розглянутих прикладах показано його ефективність. Оскільки для біологічних задач природними представниками досліджуваних об'єктів є табличні структури, то запропоновані методи можна успішно застосовувати і для прогнозування біологічних показників.

**Мета дослідження** — з'ясувати можливість використання запропонованого у праці [9] алгоритму для розв'язання задач прогнозування біологічних показників з прийнятною точністю.

Для досягнення мети необхідно:

- розрахувати прогнозні значення досліджуваних показників із використанням сучасних методів біометрії;
- розрахувати прогнозні значення показників із застосуванням запропонованого алгоритму на основі матричної множинної регресії;
- проаналізувати результати прогнозування за різними варіантами.

Емпіричними даними для розрахунків стали дані експерименту, проведеного у науковій лабораторії кафедри біохімії навчально-наукового центру «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка.

## ТЕОРЕТИЧНІ ОСНОВИ МЕТОДУ ПРОГНОЗУВАННЯ НА БАЗІ МАТРИЧНОЇ МНОЖИННОЇ РЕГРЕСІЇ

Теоретичні засади методу аналізу та прогнозування об'єктів, що виражені матрицями, детально подано у працях [6–10].

Нехай маємо  $M$  матриць спостережень:  $X_1, X_2, \dots, X_M$ . означимо  $O_i$ ,  $i = \overline{1, N}$  — матричні кортежі-рядки:

$$O_1 = (X_1 \dot{\vdots} X_K) \quad O_2 = (X_2 \dot{\vdots} X_{K+1}), \dots, \quad O_N = (X_N \dot{\vdots} X_{K+N-1}),$$

$M > N + K$ ,  $K$  — довжина основи.

Називатимемо  $O_i, i = \overline{1, N}$  основами (опорами) прогнозу, а  $Y_i = X_{K+i}, i = \overline{1, N}$  — відгуками прогнозу. Вхідні та вихідні дані моделі задають матричні пари  $(O_1, Y_1), (O_2, Y_2), \dots, (O_N, Y_N)$ .

Визначимо множинну матричну регресію з матричними значеннями [8] як матричну функцію  $Y$  матричного кортежу-рядка  $\alpha_{1,K} = (X_1 : \dots : X_K)$   $X_k \in R^{m \times n}$   $k = \overline{1, K}$  такого вигляду  $Y(\alpha_{1,K}) = \beta_1 X_1 + \dots + \beta_K X_K$   $\beta_k \in R^1$ ,  $k = \overline{1, K}$ , що визначаються на основі спостережень  $(\alpha_{1,K}^{(i)}, Y_i)$   $i = \overline{1, N}$ , де  $\alpha_{1,K}^{(i)}$ ,  $i = \overline{1, N}$  —  $i$ -а компонента матричного кортежу-рядка  $\alpha_{1,K} = (X_1 : \dots : X_K)$ . Визначимо аналогічно, як і у працях [9, 10], оператор  $\wp_{\alpha_{N,K}} : R^K \rightarrow R^{N, m \times n}$  — матричний кортежний оператор з евклідового простору  $R^K$  в евклідов простір стовпчикових кортежів довжини  $N$  з матриць розмірності  $m \times n$  покомпонентним скалярним добутком для стовпчикових

$$\text{кортежів } \chi_1 = \begin{pmatrix} A_1 \\ \dots \\ A_N \end{pmatrix} \in R^{N, m \times n}, \quad \chi_2 = \begin{pmatrix} B_1 \\ \dots \\ B_N \end{pmatrix} \in R^{N, m \times n} \text{ у вигляді } (\chi_1, \chi_2)_{N, m \times n} = \\ = \sum_{i=1}^N (A_i, B_i)_{tr} \text{ і оператор, спряжений до } \wp_{\alpha_{N,K}} \text{ — } \wp_{\alpha_{N,K}}^* : R^{N, m \times n} \rightarrow R^K.$$

### АЛГОРИТМ ПРОГНОЗУВАННЯ ПОКАЗНИКІВ НА ОСНОВІ МАТРИЧНОЇ МНОЖИННОЇ РЕГРЕСІЇ

1. Побудова стовпчикових «компонентних» кортежів  $\chi_j, j = \overline{1, K}$   $\chi_Y$

$$\chi_j = \begin{pmatrix} O_1^{(j)} \\ O_2^{(j)} \\ \dots \\ O_N^{(j)} \end{pmatrix}, \text{ де } O_i^{(j)} \text{ — } j\text{-а компонента } O_i, i = \overline{1, N}; \quad \chi_Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_N \end{pmatrix} \text{ — матрич-}$$

ний стовпчиковий кортеж, отриманий зі значень спостережень  $Y_i, i = \overline{1, N}$ .

2. Побудова матриці Грама  $\mathbf{F}$  стовпчикових «компонентних» кортежів

$$\mathbf{F} = (\chi_i, \chi_j)_{i,j=\overline{1,K}}:$$

$$\mathbf{F} = \begin{pmatrix} (\chi_1, \chi_1) & (\chi_1, \chi_2) & \dots & (\chi_1, \chi_K) \\ (\chi_2, \chi_1) & (\chi_2, \chi_2) & \dots & (\chi_2, \chi_K) \\ \dots & \dots & \dots & \dots \\ (\chi_K, \chi_1) & (\chi_K, \chi_2) & \dots & (\chi_K, \chi_K) \end{pmatrix},$$

де  $(\chi_i, \chi_j)_{i,j=\overline{1,K}}$  — скалярний добуток матричних кортежів, який визначається покомпонентним скалярним добутком відповідних матриць.

Зауваження: для матриць  $A = (a_{ij})$  а  $i = \overline{1, m}, j = \overline{1, n}$ , скалярний добуток визначається таким чином:  $(A, B) = \sum_{i=\overline{1, m}, j=\overline{1, n}} a_{ij} b_{ij} = \text{tr} A^T B$ .

3. Знаходження розв'язку задачі на власні значення для матриці  $F$ . Обчислення ненульових сингулярностей  $(v_i, \lambda_i^2)$  матриці Грама  $F$  стовпчикових «компонентних» кортежів,  $i = \overline{1, r}, r = \text{rank } F$ .

Зауваження: ненульові сингулярності  $(v_i, \lambda_i^2), i = \overline{1, r}$  записуємо у відповідності до зростання власних значень.

4. Обчислення власних кортежів  $\psi_i : \psi_i = \frac{1}{\lambda_i} \sum_{k=1}^K \chi_k v_{ki}, i = \overline{1, r}$  оператора  $\mathcal{P}_{\alpha_{N,K}} \mathcal{P}_{\alpha_{N,K}}^*$   $\psi_i$  — стовпчиковий кортеж довжини  $N$ .

5. Знаходження скалярних добутків  $sk_i = (\psi_i, \chi_Y)_{N, m \times n}, i = \overline{1, r}$ .

6. Обчислення  $\hat{\beta}$  за формулою  $\hat{\beta} = \sum_{k=1}^r \lambda_k^{-1} v_k (\psi_k, \chi_Y)_{N, m \times n}$ .

7. Знаходження прогнозної функції як умовного математичного сподівання:

$$\begin{aligned} \hat{Y} &= \hat{Y}(\alpha_{N+1, N+K}) = E(Y(\alpha_{N+1, N+K}) | \alpha_{N+1, N+K}) = \\ &= E(Y(Y(\alpha_{N+1, N+K}) | X_{N+1}, \dots, X_{N+K})) = \sum_{k=1}^K \hat{\beta}_k X_{N+k}, \end{aligned}$$

## ЕМПІРИЧНІ ДАНІ ДЛЯ РОЗРАХУНКІВ

Для розглянутих у роботі прикладів використано дані досліджень, проведених у науковій лабораторії кафедри біохімії навчально-наукового центру «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка [12, 13] під керівництвом О.М. Савчука, завідувача кафедри біохімії, доктора біологічних наук, професора. Експерименти проводились на білих нестатевозрілих пацюках жіночої статі з початковою масою 90–110 г відповідно до стандартів Конвенції про біоетику Ради Європи 1997 р., Європейської конвенції про захист хребетних тварин, що застосовуються для експериментальних та інших наукових цілей, загальних етичних принципів експериментів на тваринах, затверджених першим Національним конгресом біоетики України (вересень 2001 р.) та інших міжнародних угод та національного законодавства у цій сфері. Під час експерименту досліджувався вплив водного екстракту стручків *Phaseolus vulgaris* на фізіологічні параметри та запалення пацюків при опіках стравоходу 2-го ступеня.

Тварин розділили на 4 групи: група 1 (G1) — контрольна група тварин (фізіологічне введення); група 2 (G2) — контрольна група тварин, яким вводили екстракт *Phaseolus vulgaris* у дозі 1 мг / кг; група 3 (G3) — пацюки з опіками стравоходу 2-го ступеня та фізіологічним введенням; група 4 (G4) — пацюки з опіками стравоходу 2-го ступеня, яким вводили екстракт *Phaseolus vulgaris* у дозі 1 мг/кг.

Під час експерименту вимірювалися такі показники: приріст маси тіла відносно початкового значення, % (P1); споживання рідини, г (P2); спожи-

вання корму, г (P3). У табл. 1–3 подано значення показників P1, P2, P3 для досліджуваних груп G1 – G4 на 6-й, 9-й, 12-й, ..., 30-й дні спостережень.

**Таблиця 1.** Значення показників для досліджуваних груп на 6-й, 9-й та 12-й дні

Група	6-й день			9-й день			12-й день		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
G1	30,77	30,00	19,00	51,68	30,00	20,60	74,75	36,00	21,60
G2	44,33	40,00	19,00	65,18	25,00	18,80	82,59	25,00	17,80
G3	24,94	27,08	18,41	38,24	33,75	22,00	49,10	31,67	18,20
G4	24,37	26,54	16,00	39,97	30,62	19,00	60,73	30,38	18,80

**Таблиця 2.** Значення показників для досліджуваних груп на 15-й, 18-й та 21-й дні

Група	15-й день			18-й день			21-й день		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
G1	114,60	36,00	22,00	137,87	46,00	21,60	151,68	40,00	22,60
G2	95,34	27,00	21,00	111,34	34,00	20,00	122,87	30,00	20,00
G3	59,84	42,00	20,30	68,70	43,00	20,60	69,65	34,70	19,40
G4	73,55	37,50	22,60	83,38	42,50	21,30	86,73	39,10	18,30

**Таблиця 3.** Значення показників для досліджуваних груп на 24-й, 27-й та 30-й дні

Група	24-й день			27-й день			30-й день		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
G1	162,52	42,00	23,20	172,58	42,00	25,00	182,45	44,00	23,00
G2	129,76	34,00	21,20	136,84	35,00	22,40	142,31	35,00	21,60
G3	73,62	37,50	18,40	78,97	43,10	16,00	85,16	35,00	16,40
G4	90,60	49,20	19,40	97,51	40,60	17,40	102,74	34,80	18,60

## РЕЗУЛЬТАТИ ПРОГНОЗУВАННЯ З ВИКОРИСТАННЯМ УБУДОВАНИХ ЗАСОБІВ MICROSOFT OFFICE EXCEL

Прогнозування досліджуваних показників для кожної з груп здійснювалося на 1 крок вперед (на 30-й день за наявними даними за 6-й – 27-й дні), при цьому використано сучасні методи біометрії (інструменти Excel: лінії тренду). Зокрема у варіанті 1 застосовано лінійну апроксимацію, у варіанті 2 — експоненційну апроксимацію, а у варіанті 3 — логарифмічну апроксимацію (табл. 4).

**Таблиця 4.** Прогнозні значення показників P1, P2, P3 для досліджуваних груп на 30-й день у трьох варіантах апроксимації

Група	Варіант 1			Варіант 2			Варіант 3		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
G1	208,53	46,63	25,03	284,53	47,46	25,11	176,80	43,96	23,99
G2	157,83	32,96	22,23	181,57	33,08	22,20	138,69	30,82	21,33
G3	91,38	44,14	17,51	106,47	44,73	17,42	81,13	42,20	18,45
G4	115,26	48,96	19,57	142,04	50,72	19,46	101,54	45,16	20,00

Ступінь наближення фактичних та прогнозних даних характеризується коефіцієнтами детермінації  $R^2$  та коефіцієнтами кореляції  $R$ , значення яких для кожного з варіантів подано у табл. 5 і 6.

**Таблиця 5.** Коефіцієнти детермінації прогнозу значень показників P1, P2, P3 для досліджуваних груп у трьох варіантах апроксимації

Група	Варіант 1			Варіант 2			Варіант 3		
	$R^2$ (P1)	$R^2$ (P2)	$R^2$ (P3)	$R^2$ (P1)	$R^2$ (P2)	$R^2$ (P3)	$R^2$ (P1)	$R^2$ (P2)	$R^2$ (P3)
G1	0,960	0,696	0,895	0,880	0,721	0,893	0,951	0,747	0,864
G2	0,971	0,030	0,659	0,902	0,054	0,649	0,978	0,004	0,504
G3	0,931	0,493	0,242	0,851	0,514	0,251	0,986	0,596	0,099
G4	0,917	0,754	0,015	0,817	0,782	0,022	0,985	0,769	0,123

**Таблиця 6.** Коефіцієнти кореляції прогнозу значень показників P1, P2, P3 для досліджуваних груп у трьох варіантах апроксимації

Група	Варіант 1			Варіант 2			Варіант 3		
	R (P1)	R (P2)	R (P3)	R (P1)	R (P2)	R (P3)	R (P1)	R (P2)	R (P3)
G1	0,980	0,834	0,946	0,938	0,849	0,945	0,975	0,864	0,930
G2	0,985	0,173	0,812	0,950	0,232	0,806	0,989	0,063	0,710
G3	0,965	0,702	0,492	0,922	0,717	0,501	0,993	0,772	0,315
G4	0,958	0,868	0,122	0,904	0,884	0,148	0,992	0,877	0,351

Зауважимо, що згідно зі шкалою Чеддока (табл. 7), показнику  $R$  дається якісна оцінка, див. напр. [14].

**Таблиця 7.** Якісна оцінка коефіцієнта кореляції

$R$	(0 – 0,3)	[0,3 – 0,5)	[0,5 – 0,7)	[0,7 – 0,9)	[0,9 – 1)
Оцінка	Слабка	Помірна	Помітна	Висока	Дуже висока

Як бачимо з табл. 6, коефіцієнти кореляції іноді не перевищують 0,3, що свідчить про слабкий ступінь наближення прогнозних значень показників до фактичних. Про високий ступінь наближення можна говорити для показників, клітинки табл. 6 для яких виділено. У розділі «Обговорення результатів обчислень» наведено похибки прогнозу для кожного з варіантів та проаналізовано отримані результати.

## РЕЗУЛЬТАТИ ОБЧИСЛЕНЬ ЗА АЛГОРИТМОМ

На основі даних табл. 1 очевидним чином формуємо матриці спостережень  $X_1$ ,  $X_2$  і  $X_3$ , на основі даних табл. 2 — матриці  $X_4$ ,  $X_5$  і  $X_6$ , а на основі даних табл. 3 — матриці  $X_7$ ,  $X_8$ ,  $X_9$ . Маємо 9 матриць спостережень:  $X_1, X_2, \dots, X_9$ . Поданий вище алгоритм реалізовано у середовищі Wolfram Mathematica, за яким виконано розрахунки. Наведемо детальну схему розрахунку для варіанта групування даних,  $K = 4$ ,  $N = 4$ .

1. Формування основи прогнозу:

$$O_1 = (X_1 X_2 X_3 X_4), Y_1 = X_5; O_2 = (X_2 X_3 X_4 X_5), Y_2 = X_6;$$

$$O_3 = (X_3 X_4 X_5 X_6), Y_3 = X_7; O_4 = (X_4 X_5 X_6 X_7), Y_4 = X_8.$$

2. Побудова стовпчикових «компонентних» кортежів:

$$\chi_1 = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}, \chi_2 = \begin{pmatrix} X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix}, \chi_3 = \begin{pmatrix} X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix}, \chi_4 = \begin{pmatrix} X_4 \\ X_5 \\ X_6 \\ X_7 \end{pmatrix}, \chi_Y = \begin{pmatrix} X_5 \\ X_6 \\ X_7 \\ X_8 \end{pmatrix}.$$

3. Обчислення матриці Грама:

$$\mathbf{F} = \begin{pmatrix} 86665,96 & 104641,28 & 118441,75 & 129760,48 \\ 104641,28 & 129031,62 & 147449,86 & 162622,31 \\ 118441,75 & 147449,86 & 171160,04 & 189822,56 \\ 129760,48 & 162622,31 & 189822,56 & 212665,51 \end{pmatrix}, r = \text{rank } \mathbf{F} = 0,4.$$

4. Обчислення ненульових сингулярностей матриці Грама  $\mathbf{F}$  стовпчикових «компонентних» кортежів:

$$v_1 = (0,312599; -0,700455; 0,609736; -0,199665), \lambda_1 = 18,3022;$$

$$v_2 = (-0,507724; 0,408739; 0,563729; -0,507306), \lambda_2 = 23,9231;$$

$$v_3 = (-0,709866; -0,355271; 0,148895; -0,589663), \lambda_3 = 80,0334;$$

$$v_4 = (-0,374951; -0,464841; -0,5369; -0,59588), \lambda_4 = 769,5520.$$

8. Обчислення власних кортежів  $\psi_i, i = \overline{1, r}$ : отримуємо чотири власні кортежі-матриці розмірності  $16 \times 3$ , які не наводимо через їх великий розмір.

5. Обчислення скалярних добутків:

$$sk_1 = -3,0496; sk_2 = -12,9760; sk_3 = 65,2573; sk_4 = -491,4360.$$

6. Знаходження  $\hat{\beta}$ :

$$\hat{\beta} = (-0,1161; -0,0978; 0,0569; 1,1698).$$

7. Прогнозування показників на наступний період (30-й день) (табл. 8).

**Таблиця 8.** Прогнозні значення досліджуваних показників

Група	30-й день, прогноз		
	P1	P2	P3
G1	189,74	45,11	23,43
G2	147,30	36,13	22,15
G3	88,83	35,70	16,04
G4	106,80	33,67	18,73

### ОБГОВОРЕННЯ РЕЗУЛЬТАТІВ ОБЧИСЛЕНЬ

Для визначення точності прогнозування показників використаємо формальний критерій точності (absolute percentage error):  $APE = \left| \frac{z - \hat{z}}{z} \right| \cdot 100$ , де  $z$  —

фактичне значення відповідного показника, а  $\hat{z}$  — його прогнозне значення. Якість запропонованого методу перевірено з використанням ретроспективного оцінювання прогнозу. Зауважимо, що загальноприйнятим є твердження, що значення  $APE$ , яке менше за 10%, відповідає високій точності прогнозу, а отже, і якості моделі чи підходу; від 10 до 20 % — добрій точності; від 20 до 30 % — задовільній; більше як 30 % — незадовільній. Роз-

раховані значення похибок прогнозу значень показників за вказаним критерієм для кожного з варіантів подано у табл. 9.

**Таблиця 9.** Точність прогнозних показників для різних варіантів прогнозування

Група	Варіант 1			Варіант 2			Варіант 3			Варіант 4		
	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
G1, %	14,30	5,99	8,80	55,95	7,86	9,19	3,09	0,10	4,32	4,00	2,52	1,89
G2, %	10,91	5,83	2,91	27,59	5,50	2,79	2,54	11,94	1,27	3,51	3,22	2,53
G3, %	7,30	26,12	6,76	25,02	27,80	6,24	4,73	20,56	12,49	4,30	2,00	2,18
G4, %	12,19	40,70	5,23	38,26	45,76	4,64	1,17	29,78	7,52	3,95	3,26	0,68

Як бачимо з табл. 9, для варіантів 1, 2 і 3 значення похибок для деяких значень (виділених) варіюють від 20% до 30%, що відповідає задовільній точності прогнозу, для варіантів 1 та 2 деякі значення (виділені) перевищують 30%, що свідчить про незадовільну точність прогнозу. Для варіанта 4 похибки не перевищують 10%, що свідчить про високу точність прогнозу і, відповідно, якість методу прогнозування. Таким чином, згідно з обчисленими значеннями критерію *APE* (від 0% до 10%), точність прогнозування вказаних біологічних показників за алгоритмом, який базується на матричній множинній регресії, є високою.

## ВИСНОВКИ

Розглянуто приклади прогнозування біологічних показників із застосуванням сучасних методів біометрії та алгоритму на основі матричної множинної регресії. Емпіричними даними для обчислень стали дані експерименту, проведеного в науковій лабораторії кафедри біохімії навчально-наукового центру «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка. Розрахунки проведені у Microsoft Office Excel та середовищі Wolfram Mathematica. Згідно з обчисленими значеннями похибок прогнозу значень показників за критерієм *APE*, точність прогнозу за алгоритмом, який базується на матричній множинній регресії і, відповідно, якість методу прогнозування вказаних показників є високою (похибки від 0% до 10%). Точність прогнозу за критерієм *APE* для одного з розглянутих методів біометрії є задовільною (похибки становлять від 0% до 30%), для двох інших — незадовільною (похибки для деяких показників понад 30%).

Таким чином, виконані розрахунки показують, що методи прогнозування на основі матричної множинної регресії є конкурентоспроможними і можуть бути використані для розв'язання задач прогнозування біологічних показників з прийнятною для цього точністю.

## ЛІТЕРАТУРА

1. О.Г. Близнюченко, *Біометрія: Монографія*. Полтава: Редакційно-видавничий відділ "Terra", 2003, 346 с.
2. *Biomedical statistics* (website devoted to statistical analysis of biomedical data). [Online]. Available: <http://www.biomedicalstatistics.info/en>.
3. М.П. Горощко, С.І. Миклуш, та П.Г. Хомюк, *Біометрія*. Львів: Камула, 2004, 285 с.
4. Н.В. Кепчик, *Математические методы в биологии в контексте университетского образования*. [Электронный ресурс]. Доступно: <https://bsu.by/Cache/pdf/96113.pdf>.
5. А. Кизилова и И. Сбальцарини, "Пространственно-временное моделирование в биологии", *Биомолекула*, 2012. [Электронный ресурс]. Доступно: <https://biomolecula.ru/articles/prostranstvenno-vremennoe-modelirovanie-v-biologii>.



6. V. Donchenko, I. Nazaraga, and O. Tarasova, "Vectors and matrixes least square method: foundation and application examples", *International Journal Information Theories and Applications*, vol. 20, no. 4, pp. 311–322, 2013.
7. V. Donchenko, I. Nazaraga, and O. Tarasova, "Matrixes least squares method: examples of its application in macroeconomics and TV-media business", *Eastern-European Journal of Enterprise Technologies*, vol. 4, no. 4(70), pp. 42–46, 2014. doi: <https://doi.org/10.15587/1729-4061.2014.26292>.
8. V. Donchenko, T. Zinko, and F. Skotarenko, "Feature Vectors" in Grouping Information Problem in Applied Mathematics: Vectors and Matrixes", in *Proc. Int. Conf. Problems of Computer Intellectualization, Kyiv, Ukraine–Sofia, Bulgaria, 2012*, pp. 111–124. [Online]. Available: [http://foibg.com/ibs\\_isc/ibs-28/ibs-28-p13.pdf](http://foibg.com/ibs_isc/ibs-28/ibs-28-p13.pdf).
9. В.С. Донченко та О.В. Тарасова, "Матрична множинна регресія", *Вісник КНУ імені Тараса Шевченка. Серія: фіз.-мат. науки*, № 2, с. 133–138, 2015.
10. В.С. Донченко, Т.П. Зінько, та Ф.М. Скотаренко, "Концепція кортежності для лінійних операторів та її реалізація для матричних кортежів", *Журнал обчислювальної та прикладної математики*, № 3 (120), с. 127–140, 2015.
11. I. Nazaraha, "Predictions of macroeconomic indicators based on matrix multiple regression: examples", *System Research and Information Technologies*, no. 1, pp. 119–131, 2018. doi: <https://doi.org/10.20535/SRIT.2308-8893.2018.1.10>.
12. A. Pasichnyk, V. Dmytryk, and Ya. Rayetska, "The effect of aqueous extract of Phaseolus Vulgaris pods on the some biochemical parameters in the conditions of esophagus burn of second degree in rats", in *Book of Abstr. of XIII Int. Scientific Conf. for Students and PhD Students Youth and Progress of Biology, Ukraine, Lviv, 25 – 27 April 2017*, pp. 67–68.
13. Yana Raetska et al., "The effect of Phaseolus vulgaris pods extract on cytokines profile in the condition of alkali burn esophagus 2 degree", *Biomedical Research and Therapy*, vol. 6, no. 9, pp. 3352–3358, 2019. [Online]. Available: <http://www.bmrat.org/index.php/BMRAT/article/view/563>. doi: 10.15419/bmrat.v6i9.563
14. В.П. Сторожук та ін., *Теорія статистики: курс лекцій. Частина 1*. Тернопіль: Економічна думка, 2006, 224 с.

Надійшла 27.09.2021

#### INFORMATION ON THE ARTICLE

**Inna M. Nazaraha**, ORCID: 0000-0001-8256-515X, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: [inna\\_na@ukr.net](mailto:inna_na@ukr.net)

**Yaroslav R. Nazaraha**, Igor Sikorsky Kyiv Polytechnic Institute, Ukraine, e-mail: [yarikn21@gmail.com](mailto:yarikn21@gmail.com)

#### МАТРИЧНАЯ МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И СОВРЕМЕННЫЕ МЕТОДЫ БИОМЕТРИИ ДЛЯ ПРОГНОЗИРОВАНИЯ БИОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ: ПРИМЕРЫ / И.М. Назарага, Я.Р. Назарага

**Аннотация.** Рассмотрены примеры прогнозирования биологических показателей с применением современных методов биометрии и методов на основе матричной множественной регрессии. С целью решения задач оценивания методом наименьших квадратов для множественной матричной регрессии использованы математический аппарат сингулярного разложения и техника псевдообращения по Муру–Пенроузу в пределах развития концепции кортежных операторов. Эмпирическими данными для расчетов стали данные эксперимента, проведенного в учебно-научном центре «Институт биологии и медицины» (Киевский национальный университет имени Тараса Шевченко). Расчеты выполнены в Microsoft Office Excel и Wolfram Mathematica, проанализированы результаты. Точность прогнозирования по критерию АРЕ (абсолютный процент ошибки) согласно алгоритму, основанном на матричной множественной регрессии (погрешность от 0% до 10%), выше, чем для современных методов биометрии (некоторые погрешности превышают 30%). Показано, что матричная множественная регрессия может быть эффективным инструментом прогнозирования в биологии с приемлемой для этого точностью.

**Ключевые слова:** матричная множественная регрессия, методы биометрии, биологические показатели, прогнозирование, сингулярное разложение.

**MATRIX MULTIPLE REGRESSION AND MODERN BIOMETRIC METHODS FOR PREDICTION OF BIOLOGICAL INDICATORS: EXAMPLES** / I.M. Nazaraha, Ya.R. Nazaraha

**Abstract.** In this article, examples of prediction of biological indicators are considered. In this case, the classical methods of biometrics and methods based on matrix multiple regression are used. In order to solve the problem of estimation by the method of least squares for multiple matrix regression, a mathematical apparatus for the singular value decomposition (SVD) and pseudo-inversion technique for Moore–Penrose was used within the development of the concept of tuple operators. The empirical data for calculations were data from an experiment conducted at the Educational and Scientific Center “Institute of Biology and Medicine” (Taras Shevchenko National University of Kyiv). The calculations were made in Microsoft Office Excel and Wolfram Mathematica. The algorithm based on matrix multiple regression has the prediction accuracy in terms of the APE (absolute percentage error) criterion (the error is from 0% to 10%) higher than the accuracy of modern methods of biometrics (some errors are greater than 30%). As shown in the examples, matrix multiple regression can be an effective prediction instrument in biology with an acceptable planning processes accuracy.

**Keywords:** matrix multiple regression, methods of biometrics, biological indicators, prediction, singular-value decomposition.

**REFERENCES**

1. O.H. Blyzniuchenko, *Biometrics: monograph*, (in Ukrainian). Poltava: “Terra”, 2003, 346 p.
2. *Biomedical statistics*. [Online]. Available: <http://www.biomedicalstatistics.info/en>.
3. M.P. Horoshko, S.I. Myklush, and P.H. Khomiuk, *Biometrics*, (in Ukrainian). Lviv: Kamula, 2004, 285 p.
4. N.V. Kepchik, *Mathematical methods in biology in the context of university education*, (in Russian). [Online]. Available: <https://bsu.by/Cache/pdf/96113.pdf>.
5. A. Kizilova and I. Sbal'czarini, “Space-time modeling in biology”, (in Russian), *Bio-molecula*, 2012. [Online]. Available: <https://biomolecula.ru/articles/prostranstvenno-vremennoe-modelirovanie-v-biologii>.
6. V. Donchenko, I. Nazaraga, and O. Tarasova, “Vectors and matrixes least square method: foundation and application examples”, *International Journal Information Theories and Applications*, vol. 20, no. 4, pp. 311–322, 2013.
7. V. Donchenko, I. Nazaraga, and O. Tarasova. “Matrixes least squares method: examples of its application in macroeconomics and TV-media business”. *Eastern-European Journal of Enterprise Technologies*. vol. 4, no. 4(70), pp. 42–46, 2014. doi: <https://doi.org/10.15587/1729-4061.2014.26292>.
8. V. Donchenko, T. Zinko, and F. Skotarenko, “Feature Vectors” in Grouping Information Problem in Applied Mathematics: Vectors and Matrixes”, in *Proc. Int. Conf. Problems of Computer Intellectualization, Kyiv, Ukraine–Sofia, Bulgaria, 2012*, pp. 111–124. [Online]. Available: [http://foibg.com/ibs\\_isc/ibs-28/ibs-28-p13.pdf](http://foibg.com/ibs_isc/ibs-28/ibs-28-p13.pdf).
9. V. Donchenko and O. Tarasova. “Matrix multiple regression”. (in Ukrainian). *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physics and Mathematics*, no. 2, pp. 133–138, 2015.
10. V. Donchenko, T. Zinko, and F. Skotarenko, “Cortege conception for linear operators and its implementation for matrix corteges”, (in Ukrainian), *Journal of Computational and Applied Mathematics*, no. 3 (120), pp. 127–140, 2015.
11. I. Nazaraha, “Predictions of macroeconomic indicators based on matrix multiple regression: examples”, *System Research and Information Technologies*, no. 1, pp. 119–131, 2018. doi: <https://doi.org/10.20535/SRIT.2308-8893.2018.1.10>.
12. A. Pasichnyk, V. Dmytryk, and Ya. Rayetska, “The effect of aqueous extract of Phaseolus Vulgaris pods on the some biochemical parameters in the conditions of esophagus burn of second degree in rats”, in *Book of Abstr. of XIII Int. Scientific Conf. for Students and PhD Students Youth and Progress of Biology, Ukraine, Lviv, 25 – 27 April 2017*, pp. 67–68.
13. Yana Raetska et al., “The effect of Phaseolus vulgaris pods extract on cytokes profile in the condition of alkali burn esophagus 2 degree”, *Biomedical Research and Therapy*, vol. 6, no. 9, pp. 3352–3358, 2019. [Online]. Available: <http://www.bmrat.org/index.php/BMRAT/article/view/563>. doi: 10.15419/bmrat.v6i9.563
14. V.P. Storozhuket al., *Theory of Statistics: Lecture Course, Part 1*, (in Ukrainian). Ternopil: Ekonomichna dumka, 2006, 224 p.