

УДК 004.412:519.237

Сергій ПРИХОДЬКО

sergiy.prykhodko@nuos.edu.ua

Наталія ПРИХОДЬКО

natalia.prykhodko@nuos.edu.ua

Олег КУДІН

oleg.kudin@nuos.edu.ua

Дмитро КОШОВИЙ

м. Миколаїв

ПОБУДОВА РІВНЯННЯ ТРАНСФОРМОВАНОГО ЕЛІПСУ ПЕРЕДБАЧЕННЯ ЗА ДАНИМИ ТРИВАЛОСТІ РОЗРОБКИ 3D-МОДЕЛЕЙ СЕКЦІЙ КОРПУСІВ СУДЕН

В статті вперше побудовано рівняння трансформованого еліпсу передбачення для двовимірних негаусівських даних тривалості розробки 3D-моделей секцій корпусів суден в залежності від маси секцій на основі перетворення Джонсона сім'ї S_B . Наводиться приклад застосування цього рівняння до визначення викидів у зазначених двовимірних негаусівських даних.

Ключові слова: рівняння трансформованого еліпсу передбачення; перетворення Джонсона сім'ї S_B ; двовимірні негаусівські дані; тривалість; 3D-модель секції корпусу судна.

Постановка проблеми

При статистичному аналізі двовимірних даних використовується, як правило, еліпс [1, 2]. Завдяки еліпсу передбачення можна виявити двовимірні викиди (bivariate outliers) в даних, тому він часто застосовується для цього, зокрема в інформаційних технологіях. Статистичні методи, що базуються на еліпсі передбачення, використовують припущення, що дані апроксимуються двовимірним розподілом Гаусу. Але таке припущення може бути застосованим тільки в окремих випадках. Це призводить до необхідності побудови рівняння трансформованого еліпсу передбачення для конкретного набору даних, розподіл якого суттєво відрізняється від нормального. Саме такими даними є тривалості розробки 3D-моделей секцій корпусів суден в залежності від маси секцій.

Аналіз останніх досліджень і публікацій

В роботі [3] був запропонований метод побудови трансформованого еліпсу передбачення для двовимірних негаусівських даних. Цей метод складається з наступних

етапів. Спочатку двовимірні негаусівські дані нормалізують за допомогою бієктивного двовимірного нормалізуючого перетворення. Після цього отримують рівняння еліпсу передбачення для нормалізованих даних. На останньому етапі будують рівняння трансформованого еліпсу передбачення для двовимірних початкових даних на основі рівняння еліпсу передбачення для нормалізованих даних та зворотного перетворення.

Формулювання цілей

Метою роботи є побудова рівняння трансформованого еліпсу передбачення для двовимірних негаусівських даних тривалості розробки 3D-моделей секцій корпусів суден в залежності від маси секцій із використанням методу, запропонованого у роботі [3].

Основний матеріал

Як вказано вище, метод побудови трансформованого еліпсу передбачення [3] складається з трьох етапів.

На першому етапі вектор двовимірних негаусівських даних $\mathbf{X} = \{X_1, X_2\}^T$ норма-

лізують у двовимірний гаусівський вектор $\mathbf{Z} = \{Z_1, Z_2\}^T$ на основі рівняння

$$\mathbf{Z} = \psi(\mathbf{X}). \quad (1)$$

Перетворення (1) має зворотне перетворення

$$\mathbf{X} = \psi^{-1}(\mathbf{Z}). \quad (2)$$

На другому етапі будують рівняння еліпсу передбачення для нормалізованих даних

$$(\mathbf{Z} - m_Z)^T \mathbf{S}^{-1} (\mathbf{Z} - m_Z) = \frac{2(N^2 - 1)}{N(N - 2)} F_{2, N-2, \alpha}, \quad (3)$$

де $F_{2, N-2, \alpha}$ – квантиль F -розподілу; N – об'єм вибірки; m_Z – вектор середніх значень, $m_Z = (m_{Z_1}, m_{Z_2})^T$; α – рівень значимості; \mathbf{S} – матриця коваріації

$$\mathbf{S} = \begin{pmatrix} S_{Z_1}^2 & S_{Z_1 Z_2} \\ S_{Z_1 Z_2} & S_{Z_2}^2 \end{pmatrix}.$$

Ліву частину виразу (3) можна представити у вигляді

$$\begin{aligned} & \frac{S_{Z_1}^2 S_{Z_2}^2}{S_{Z_1}^2 S_{Z_2}^2 - S_{Z_1 Z_2}^2} \left[\frac{(Z_1 - m_{Z_1})^2}{S_{Z_1}^2} + \frac{(Z_2 - m_{Z_2})^2}{S_{Z_2}^2} - \right. \\ & \left. - \frac{2S_{Z_1 Z_2} (Z_1 - m_{Z_1})(Z_2 - m_{Z_2})}{S_{Z_1}^2 S_{Z_2}^2} \right] = \\ & = \frac{2(N^2 - 1)}{N(N - 2)} F_{2, N-2, \alpha}. \quad (4) \end{aligned}$$

На третьому етапі рівняння трансформованого еліпсу передбачення для двовимірних негаусівських даних будується на основі рівняння еліпсу передбачення (3) для нормалізованих даних та перетворення (1). З урахуванням (4) рівняння трансформованого еліпсу передбачення має вигляд

$$\begin{aligned} & \frac{[\psi_1(X_1) - m_{Z_1}]^2}{S_{Z_1}^2} + \frac{[\psi_2(X_2) - m_{Z_2}]^2}{S_{Z_2}^2} - \\ & - \frac{2S_{Z_1 Z_2} [\psi_1(X_1) - m_{Z_1}][\psi_2(X_2) - m_{Z_2}]}{S_{Z_1}^2 S_{Z_2}^2} = \\ & = \frac{2(N^2 - 1)(S_{Z_1}^2 S_{Z_2}^2 - S_{Z_1 Z_2}^2)}{N(N - 2) S_{Z_1}^2 S_{Z_2}^2} F_{2, N-2, \alpha}. \quad (5) \end{aligned}$$

У якості нормалізуючого перетворення (1) в роботі обрано перетворення Джонсо-

на сім'ї S_B . Нормалізуюче перетворення Джонсона має вигляд [4]

$$\mathbf{Z} = \boldsymbol{\gamma} + \boldsymbol{\eta} \mathbf{h}[\boldsymbol{\lambda}^{-1}(\mathbf{X} - \boldsymbol{\varphi})] \sim N_m(\mathbf{0}_m, \mathbf{S}), \quad (6)$$

де $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$, $\boldsymbol{\varphi}$ та $\boldsymbol{\lambda}$ – параметри нормалізуючого перетворення Джонсона;

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^T; \quad \boldsymbol{\eta} = \text{diag}(\eta_1, \eta_2);$$

$$\boldsymbol{\varphi} = (\varphi_1, \varphi_2)^T; \quad \boldsymbol{\lambda} = \text{diag}(\lambda_1, \lambda_2);$$

$$\mathbf{h}[(y_1, y_2)] = \{h_1(y_1), h_2(y_2)\}^T,$$

де $h_i(\cdot)$ – одна з функцій перетворення

$$h = \begin{cases} \ln(y), & \text{for } S_L \text{ (log normal) family;} \\ \ln[y/(1-y)], & \text{for } S_B \text{ (bounded) family;} \\ \text{Arsh}(y), & \text{for } S_U \text{ (unbounded) family;} \\ y & \text{for } S_N \text{ (normal) family.} \end{cases}$$

Тут $y = (x - \varphi)/\lambda$; $\text{Arsh}(y) = \ln(y + \sqrt{y^2 + 1})$.

З урахуванням (6) рівняння трансформованого еліпсу передбачення (5) для двовимірних негаусівських даних тривалості розробки 3D-моделей секцій корпусів суден в залежності від маси секцій можна записати як

$$\begin{aligned} & \frac{\left[\left(\gamma_1 + \eta_1 \ln \frac{x_1 - \varphi_1}{\lambda_1 + \varphi_1 - x_1} \right) - m_{Z_1} \right]^2}{S_{Z_1}^2} + \\ & + \frac{\left[\left(\gamma_2 + \eta_2 \ln \frac{x_2 - \varphi_2}{\lambda_2 + \varphi_2 - x_2} \right) - m_{Z_2} \right]^2}{S_{Z_2}^2} - \\ & - \frac{2S_{Z_1 Z_2} \left[\left(\gamma_1 + \eta_1 \ln \frac{x_1 - \varphi_1}{\lambda_1 + \varphi_1 - x_1} \right) - m_{Z_1} \right] \times}{S_{Z_1}^2 S_{Z_2}^2} \times \\ & \times \left[\left(\gamma_2 + \eta_2 \ln \frac{x_2 - \varphi_2}{\lambda_2 + \varphi_2 - x_2} \right) - m_{Z_2} \right] = \\ & = \frac{2(N^2 - 1)(S_{Z_1}^2 S_{Z_2}^2 - S_{Z_1 Z_2}^2)}{N(N - 2) S_{Z_1}^2 S_{Z_2}^2} F_{2, N-2, \alpha}. \quad (7) \end{aligned}$$

Розглянемо приклад застосування рівняння (7) для визначення двовимірних викидів у негаусівських даних тривалості розробки 3D-моделей секцій корпусів суден в залежності від маси секцій.

Приклад. Вибірка із 144 двовимірних даних тривалості розробки 3D-моделей секцій корпусів суден в залежності від маси секцій наведена в табл. 1. Колонки табл. 1 мають наступні позначки; n – порядковий номер пари даних «маса секції, тривалість розробки 3D-моделі секції»; X_1 – маса сек-

ції в тонах; X_2 – тривалість розробки 3D-моделі секції в годинах. Для нормалізації початкових даних було обрано перетворення Джонсона (6) сім'ї S_B згідно з [5]. Оцінювання параметрів цього перетворення здійснювалося за методом максимальної правдоподібності.

Табл. 1. Тривалість розробки 3D-моделей секцій корпусів суден в залежності від маси секцій

n	X ₁	X ₂	n	X ₁	X ₂	n	X ₁	X ₂	n	X ₁	X ₂
1	40,62	111	37	41,57	68	73	1,69	41	109	64,58	49
2	40,62	45	38	38,53	82	74	2,25	39	110	95,31	44
3	49,51	233	39	43,6	101	75	3,56	43	111	86,64	98
4	49,51	17	40	35,7	67	76	78,19	328	112	61,13	54
5	33,25	132,5	41	11,89	72	77	78,26	51	113	67,24	65
6	33,25	90,5	42	5,66	28	78	78,39	337	114	52,39	74
7	25,54	101	43	17,14	110	79	78,41	44	115	70,5	193
8	24,63	114	44	2,1	39	80	50,93	221	116	55,4	158
9	41,81	188	45	2,1	32	81	50,93	83	117	50,3	173
10	25,67	106	46	27,06	124	82	59,54	388	118	72,78	331
11	21,1	70	47	24,91	47	83	60,36	402	119	94,9	242
12	14,02	46	48	14,94	25	84	77,23	340	120	84,7	144
13	14,02	24	49	18,82	49	85	78,24	319	121	67,68	138
14	5,92	56	50	39,42	93	86	85,62	154	122	59,02	125
15	9,64	28	51	46,29	148	87	82,5	148	123	64,16	208
16	37,07	17	52	12,13	145	88	73,74	154	124	66,04	292
17	37,14	17	53	10,5	159	89	75,4	179	125	56,19	250
18	37,22	18	54	17,46	143	90	67,3	194	126	56,95	74
19	33,97	16	55	10,67	151	91	83,57	154	127	75,81	167
20	44,36	34	56	11,35	197	92	55,31	202	128	66,77	187
21	37,65	12	57	13,16	179	93	61,49	144	129	65,79	193
22	37,25	13	58	12,61	56	94	75,58	215	130	57,27	127
23	37,15	12	59	14,34	113	95	78,18	212	131	54,95	112
24	44,15	26	60	21,04	84	96	60,44	127	132	57,89	134
25	39,81	31	61	25,63	70	97	91,15	275	133	50,95	136
26	24,83	31	62	24,55	109	98	51,36	183	134	50,3	96
27	6,15	20	63	15,88	124	99	55,22	226	135	56,12	46
28	13,08	34	64	15,83	180	100	56,68	156	136	55,86	116
29	35,07	69	65	20,15	224	101	74,16	152	137	57,91	126
30	34,15	24	66	20,38	172	102	83,9	134	138	98,17	289
31	8,01	73	67	17,15	264	103	96,54	48	139	89,51	263
32	41,48	71	68	17,23	118	104	53,91	48	140	54,56	302
33	41,34	101	69	22,64	74	105	90,43	50	141	54,15	252
34	29,69	108	70	2,64	47	106	91,25	64	142	99,56	222
35	48,66	90	71	0,69	23	107	98,35	55	143	72,13	311
36	36,06	100	72	11,53	66	108	90,32	21	144	73,07	183

В результаті нормалізації було отримано такі параметри перетворення Джонсона сім'ї S_B : $\gamma_1 = 0,153813$; $\eta_1 = 0,697331$; $\varphi_1 = -1,660918$; $\lambda_1 = 104,0445$; $\gamma_2 = 0,968767$; $\eta_2 = 0,769265$; $\varphi_2 = 6,432764$; $\lambda_2 = 427,4441$. Отримано середні значення перетворених даних Z : $m_{z_1} = -5,4598E-09$; $m_{z_2} = 1,0296E-09$. Для нормалізованих даних матриця коваріації S має значення

$$S_N = \begin{Bmatrix} 1,000000 & 0,379791 \\ 0,379791 & 0,999999 \end{Bmatrix}.$$

На рис. 1 представлено еліпс передбачення нормалізованих даних, що побудований за рівнянням (4) для $F_{2,142,0.05} = 3,060$, і двовимірні нормалізовані дані з розробки 3D-моделей секцій корпусів суден в залежності від маси секцій.

На рис. 2 представлено трансформований еліпс передбачення для початкових даних з табл. 1, що побудований за рівнянням (7), та двовимірні дані про тривалості розробки 3D-моделей секцій корпусів суден в залежності від маси секцій.

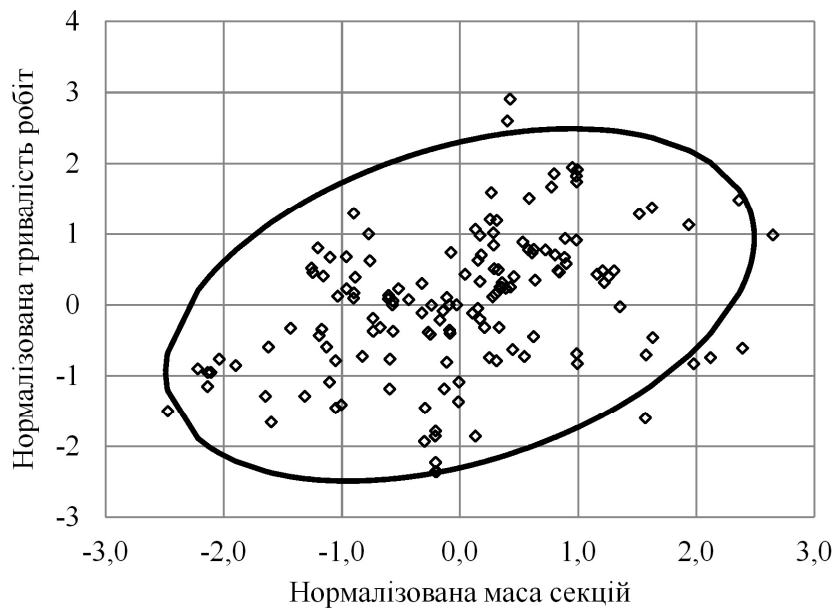


Рис. 1. Еліпс передбачення для нормалізованих даних і двовимірні нормалізовані дані з розробки 3D-моделей секцій корпусів суден

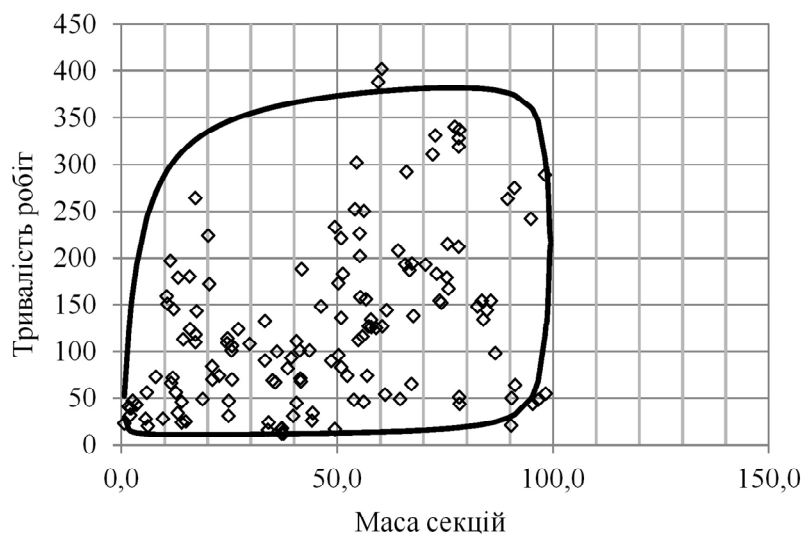


Рис. 2. Трансформований еліпс передбачення і вибірка із 144 двовимірних даних про тривалість робіт з розробки 3D-моделей секцій корпусів суден

Завдяки трансформованому еліпсу передбачення визначено вісім точок даних (пари даних: 71, 82, 83, 103, 107, 108, 110 та 142), які є двовимірними викидами (bivariate outliers). Реалізація отриманого рівняння (7) виконана в системі MS Excel 2010.

Висновки

Вперше отримано рівняння трансформованого еліпсу передбачення для дво-

мірних негасівських даних на основі двовимірного нормалізуючого перетворення Джонсона сім'ї S_B для тривалостей робіт з розробки 3D-моделей секцій корпусів суден в залежності від маси секцій, що дозволяє визначити наявність двовимірних викидів (bivariate outliers) у подібних наборах негаусівських даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Chew, V. Confidence, prediction and tolerance regions for the multivariate normal distribution [Text] / V. Chew // Journal of the American Statistical Association, Vol. 61. – 1966. – P. 605-617.
2. Friendly, Michael. Elliptical Insights: Understanding Statistical Methods Through Elliptical Geometry [Text] / Michael Friendly, Georges Monette, John Fox // Statistical Science, Vol. 28. – 2013. – No. 1. – P. 1-39.
3. Prykhodko, S.B. Constructing the transformed prediction ellipses on the basis of normalizing transformations for bivariate non-Gaussian data [Text] / S. B. Prykhodko, N. V. Prykhodko, O. O. Kudin, T. G. Smykodub // Problems of information technologies. – 2017. – № 1 (021). – P. 134-138. – ISSN 1998-7005
4. Stanfield, P.M. Multivariate input modeling with Johnson distributions [Text] / P.M. Stanfield, J.R. Wilson, G.A. Mirka, N.F. Glasscock, J.P. Psihogios and J.R. Davis // Proceedings of the 28th Winter simulation conference WSC'96, December 8-11, 1996, Coronado, CA, USA, ed. S.Andradyttir, K.J.Healy, D.H.Withers, and B.L.Nelson, IEEE Computer Society Washington, DC, USA, P. 1457-1464.
5. Приходько, С.Б. Методичні вказівки до виконання лабораторних робіт з дисципліни «Обробка експериментальних даних на ЕОМ» [Текст]. – Миколаїв: НУК, 2005. – 52 с.

Sergiy PRYKHODKO, Natalia PRYKHODKO, Oleh KUDIN, Dmytro KOSHOVY
Mykolayiv

CONSTRUCTING THE EQUATION OF TRANSFORMED PREDICTION ELLIPSE FOR DATA OF THE DURATION OF THE DEVELOPMENT OF 3D-MODELS OF SHIP HULL SECTIONS

In the article for the first time the equation of transformed prediction ellipse for data of the duration of the development of 3D-models of ship hull sections is constructed. An example of applying the equation to the determination of outliers in these bivariate non-Gaussian data is given.

Keywords: *equation of transformed prediction ellipse; S_B family of Johnson's transformation; bivariate non-Gaussian data; duration; 3D-model of ship hull section.*

Сергей ПРИХОДЬКО, Наталья ПРИХОДЬКО, Олег КУДИН, Дмитрий КОШЕВОЙ
Николаев

ПОСТРОЕНИЕ УРАВНЕНИЯ ТРАНСФОРМИРОВАННОГО ЭЛЛИПСА ПРЕДСКАЗАНИЯ ПО ДАННЫМ ДЛИТЕЛЬНОСТИ РАЗРАБОТКИ 3D-МОДЕЛЕЙ СЕКЦИЙ КОРПУСОВ СУДОВ

В статье впервые построено уравнение трансформированного эллипса предсказания для двумерных негауссовских данных продолжительности разработки 3D-моделей секций корпусов судов в зависимости от массы секций на основе преобразования Джонсона семьи S_B . Приводится пример применения этого уравнения для определения выбросов в указанных двумерных негауссовских данных.

Ключевые слова: *уравнение трансформированного эллипса предсказания; преобразования Джонсона на семейства S_B ; двумерные негауссовские данные; продолжительность; 3D-модель секции корпуса судна.*

Стаття надійшла до редколегії 23.10.2017