

УДК: 004.412:519.237.5

НЕЛІНІЙНЕ РЕГРЕСІЙНЕ РІВНЯННЯ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНИХ СИСТЕМ З ВІДКРИТИМ КОДОМ НА PHP

Приходько Н.В., Приходько С.Б.

THE NON-LINEAR REGRESSION EQUATION TO ESTIMATE THE SOFTWARE SIZE OF OPEN SOURCE PHP-BASED INFORMATION SYSTEMS

Prykhodko N.V., Prykhodko S.B.

Нелінійне регресійне рівняння для оцінювання розміру програмного забезпечення інформаційних систем з відкритим кодом на PHP побудовано на основі багатовимірного перетворення Джонсона для сімейства S_B . Це рівняння, в порівнянні з іншими регресійними рівняннями, має більший множинний коефіцієнт детермінації, менше значення середньої величини відносної похибки та менші ширини довірчого інтервалу та інтервалу передбачення нелінійної регресії.

Ключові слова: нелінійне регресійне рівняння, довірчий інтервал, інтервал передбачення, оцінювання розміру програмного забезпечення, інформаційна система, PHP.

Вступ. Задача оцінювання розміру програмного забезпечення (ПЗ) на ранній стадії його розробки є важливою, оскільки ця інформація використовується для прогнозування трудомісткості розробки ПЗ за допомогою таких моделей як СОСОМО II. Це призводить до необхідності розробки відповідних моделей для оцінювання розміру ПЗ, включаючи ПЗ інформаційних систем з відкритим кодом на PHP [1-3].

Постановка проблеми. У роботах [1, 2] запропоновано рівняння лінійної регресії для оцінювання розміру ПЗ інформаційних системи, що розробляються на базі певних мов програмування, включаючи PHP. Запропоновані рівняння побудовано за допомогою множинного лінійного регресійного аналізу на основі метрик діаграми класів. Проте існують чотири основні припущення, які виправдовують використання моделей лінійної регресії, одним з яких є нормальність розподілу помилок. Але це припущення справедливе лише в окремих випадках. Це призводить до необхідності використання нелінійних регресійних моделей та

рівнянь, в тому числі для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на PHP.

Аналіз останніх досліджень і публікацій. Нормалізуючі перетворення дуже часто є добрим способом побудови нелінійних регресійних моделей та рівнянь [3-9]. Однак добре відомі методи їх побудови, що засновані на одновимірних нормалізуючих перетвореннях (таких як логарифмічне і Бокса-Кокса), не враховують кореляції між випадковими змінними у разі нормалізації багатовимірних негаусових даних. Це призводить до необхідності використання багатовимірних нормалізуючих перетворень [3], які враховують цю кореляцію, для побудови нелінійних регресійних моделей та рівнянь для оцінювання розміру ПЗ інформаційних систем, в тому числі з відкритим кодом на PHP. В [3] було побудовано нелінійне регресійне рівняння для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на PHP на основі багатовимірного перетворення Джонсона для сімейства S_B . Але для побудови цього рівняння були використані дані всього з 32 інформаційних систем, кількість яких фактично межує з малою вибіркою. Тому виникає потреба у побудові відповідного рівняння з більшою кількістю даних.

Мета статті – побудова нелінійного регресійного рівняння для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на PHP на основі багатовимірного нормалізуючого перетворення, яке б в порівнянні з іншими рівняннями мало більший множинний коефіцієнт детермінації, менше значення середньої величини відносної похибки і менші ширини довірчого інтервалу та інтервалу передбачення регресії.

Виклад основного матеріалу. Як і в [3], побудову нелінійного регресійного рівняння для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на РНР здійснюємо за допомогою біективного багатовимірного нормалізуючого перетворення негаусового випадкового вектора $\mathbf{P} = \{Y, X_1, X_2, \dots, X_k\}^T$ у гаусовий випадковий вектор $\mathbf{T} = \{Z_Y, Z_1, Z_2, \dots, Z_k\}^T$

$$\mathbf{T} = \boldsymbol{\Psi}(\mathbf{P}), \tag{1}$$

і зворотного перетворення для (1)

$$\mathbf{P} = \boldsymbol{\Psi}^{-1}(\mathbf{T}). \tag{2}$$

Для цього спочатку виконуємо нормалізацію багатовимірних негаусових даних за (1). Далі за нормалізованими даними будемо лінійне регресійне рівняння

$$\hat{Z}_Y = \bar{Z}_Y + (\mathbf{Z}_X^+)^T \hat{\mathbf{b}}, \tag{3}$$

де \hat{Z}_Y – результат передбачення за лінійним регресійним рівнянням для значень компонент вектору $\mathbf{z}_X = \{Z_1, Z_2, \dots, Z_k\}$; \mathbf{Z}_X^+ – матриця центрованих регресорів $Z_{1i} - \bar{Z}_1, Z_{2i} - \bar{Z}_2, \dots, Z_{ki} - \bar{Z}_k$; $\hat{\mathbf{b}}$ – оцінка для вектору параметрів лінійного регресійного рівняння, $\mathbf{b} = \{b_1, b_2, \dots, b_k\}^T$.

Після цього за (1)-(3) будемо нелінійне регресійне рівняння

$$\hat{Y} = \Psi_Y^{-1} \left[\bar{Z}_Y + (\mathbf{Z}_X^+)^T \hat{\mathbf{b}} \right], \tag{4}$$

довірчий інтервал

$$\Psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, \nu} S_{Z_Y} \left\{ \frac{1}{N} + (\mathbf{z}_X^+)^T \left[(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right) \tag{5}$$

та інтервал передбачення нелінійної регресії

$$\Psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, \nu} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (\mathbf{z}_X^+)^T \left[(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right) \tag{6}$$

Тут $S_{Z_Y}^2 = \frac{1}{\nu} \sum_{i=1}^N (Z_{Y_i} - \hat{Z}_{Y_i})^2$, $\nu = N - k - 1$; $(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+$ – $k \times k$ матриця

$$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ = \begin{pmatrix} S_{Z_1 Z_1} & S_{Z_1 Z_2} & \dots & S_{Z_1 Z_k} \\ S_{Z_1 Z_2} & S_{Z_2 Z_2} & \dots & S_{Z_2 Z_k} \\ \dots & \dots & \dots & \dots \\ S_{Z_1 Z_k} & S_{Z_2 Z_k} & \dots & S_{Z_k Z_k} \end{pmatrix},$$

де $S_{Z_q Z_r} = \sum_{i=1}^N [Z_{q_i} - \bar{Z}_q][Z_{r_i} - \bar{Z}_r]$, $q, r = 1, 2, \dots, k$.

Для побудови нелінійного регресійного рівняння використовуємо вибірку чотиривимірних негаусових даних з [1]: фактичний розмір ПЗ в тисячах рядків коду (KLOC) Y , загальна кількість класів X_1 , загальна кількість зв'язків X_2 та середня кількість атрибутів на клас X_3 у концептуальній моделі даних з 54 інформаційних систем з відкритим кодом на РНР. Ця вибірка даних наведена в таблиці 1.

Таблиця 1

Вибірка чотиривимірних негаусових даних

№	Y	X ₁	X ₂	X ₃	№	Y	X ₁	X ₂	X ₃
1	22,599	17	7	7	28	8,947	15	5	4
2	32,243	21	13	4,524	29	9,47	8	5	5
3	16,164	13	11	7,077	30	19,381	10	10	5,8
4	83,862	35	24	6,571	31	4,219	6	2	8
5	24,22	13	9	8,077	32	65,649	34	12	7,912
6	63,929	35	19	8,029	33	12,443	9	6	6,222
7	6,697	5	5	7	34	3,856	5	2	11,6
8	55,537	25	14	8,64	35	10,225	6	4	11,333
9	55,752	39	10	9,077	36	47,734	19	13	8,053
10	62,602	30	17	7	37	4,554	4	5	8,5
11	67,111	23	22	14,957	38	16,846	9	8	6,333
12	2,552	3	1	8,333	39	33,645	16	15	5,438
13	12,17	10	5	3,7	40	6,532	6	5	6,167
14	12,757	13	9	5	41	16,006	6	9	8,667
15	5,695	7	3	8,429	42	10,507	6	4	9,333
16	11,054	9	9	3,667	43	7,607	5	5	10,2
17	29,77	17	15	3,412	44	15,875	11	7	6,455
18	11,653	9	8	8,778	45	57,643	30	23	9,533
19	6,847	5	4	3,6	46	9,452	10	4	6,6
20	13,389	7	5	11,714	47	10,81	10	6	6,5
21	14,45	12	6	16,583	48	38,781	20	14	10,1
22	4,414	6	3	3,667	49	6,118	6	3	13,333
23	2,102	3	1	3,333	50	24,227	14	10	4,643
24	42,819	20	18	3,5	51	23,457	15	9	5,8
25	4,077	4	2	9	52	22,475	11	10	12,818
26	57,408	33	14	9,242	53	8,59	6	5	7,833
27	7,428	7	3	7	54	11,724	10	3	9,6

Зазначимо, що ці 54 інформаційні системи були відібрані наступним чином. Спочатку ми поєднали дані з 32 систем, за якими будувалися регресійні рівняння в [1-3], з даними 31 системи, що використовувалися в [1] для перевірки лінійного регресійного рівняння. Потім ці об'єднані дані були перевірені на наявність викидів за [10] за допомогою квадрату відстані Махаланобісу (MD) та багатомірного перетворення Джонсона для рівня значимості 0,005. За результатами цієї перевірки

спочатку були відкинуті дані системи Webaddressbook [1] $Y = 8,363$, $X_1 = 3$, $X_2 = 2$, $X_3 = 20$ тому, що для цих даних квадрат MD дорівнює 15,79 і є більшим ніж значення квантіля розподілу χ^2 , яке становить 14,86 для рівня значимості 0,005. Далі були відкинуті дані ще 8 систем: 4 систем (Banpegex3a, Jdcms, Phpalumni, Phrcollegeex) з вибірки, за якою будувалося лінійне регресійне рівняння в [1], та 4 систем (Cocg, Phrpollster, Phrpera, Supersurf) з набору даних, що використовувався в [1] для перевірки цього рівняння.

Для нормалізації багатомірних негаусовських даних з таблиці 1 ми використовуємо одновимірне і багатомірне перетворення Джонсона для сімейства S_B . У нашому випадку перетворення Джонсона має вигляд [3]

$$T = \gamma + \eta h[\lambda^{-1}(P - \varphi)] \sim N_m(\mathbf{0}_m, \Sigma), \quad (7)$$

де $h[(y_Y, y_1, \dots, y_k)] = \{h_Y(y_Y), h_1(y_1), \dots, h_k(y_k)\}^T$; $h_i(\cdot)$ – одна з функцій

$$h = \begin{cases} \ln(y), & \text{для } S_L \text{ сімейства;} \\ \ln[y/(1-y)], & \text{для } S_B \text{ сімейства;} \\ \text{Arsh}(y), & \text{для } S_U \text{ сімейства;} \\ y & \text{для } S_N \text{ сімейства.} \end{cases} \quad (8)$$

$$\text{В (8) } y = (X - \varphi)/\lambda; \quad \text{Arsh}(y) = \ln\left(y + \sqrt{y^2 + 1}\right).$$

У нашому випадку X дорівнює Y , X_1 , X_2 або X_3 відповідно.

Параметри одновимірного і багатомірного перетворень Джонсона для сімейства S_B оцінювалися методом максимальної правдоподібності. Оцінки параметрів багатомірного перетворення Джонсона такі: $\hat{\gamma}_Y = 1,37544$, $\hat{\gamma}_1 = 1,33853$, $\hat{\gamma}_2 = 1,48264$, $\hat{\gamma}_3 = 1,20188$, $\hat{\eta}_Y = 0,69574$, $\hat{\eta}_1 = 0,79617$, $\hat{\eta}_2 = 0,98754$, $\hat{\eta}_3 = 1,0550$, $\hat{\varphi}_Y = 1,8525$, $\hat{\varphi}_1 = 2,5748$, $\hat{\varphi}_2 = 0,32355$, $\hat{\varphi}_3 = 2,5737$, $\hat{\lambda}_Y = 110,623$, $\hat{\lambda}_1 = 51,022$, $\hat{\lambda}_2 = 36,891$ і $\hat{\lambda}_3 = 18,319$. Вибіркова коваріаційна матриця S_N використовується в якості оцінки Σ

$$S_N = \begin{pmatrix} 1,0000 & 0,9539 & 0,9370 & 0,1467 \\ 0,9539 & 1,0000 & 0,8629 & 0,1467 \\ 0,9370 & 0,8629 & 1,0000 & 0,0607 \\ 0,1467 & 0,3795 & 0,0607 & 1,0000 \end{pmatrix}.$$

Оцінки параметрів для одновимірних перетворень Джонсона є такими: $\hat{\gamma}_Y = 1,05761$, $\hat{\gamma}_1 = 0,91563$, $\hat{\gamma}_2 = 0,95393$, $\hat{\gamma}_3 = 0,99748$, $\hat{\eta}_Y = 0,63815$, $\hat{\eta}_1 = 0,67954$, $\hat{\eta}_2 = 0,84111$, $\hat{\eta}_3 = 0,98946$, $\hat{\varphi}_Y = 1,8379$, $\hat{\varphi}_1 = 2,7147$, $\hat{\varphi}_2 = 0,48957$, $\hat{\varphi}_3 = 2,6251$, $\hat{\lambda}_Y = 89,201$, $\hat{\lambda}_1 = 39,930$, $\hat{\lambda}_2 = 27,843$ і $\hat{\lambda}_3 = 16,590$.

Далі за нормалізованими даними будемо лінійне регресійне рівняння (3)

$$\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3. \quad (9)$$

Параметри рівняння (9) оцінювалися методом найменших квадратів. Оцінки параметрів рівняння (9) для багатомірного перетворення Джонсона такі: $\hat{b}_0 = 0$, $\hat{b}_1 = 0,51856$, $\hat{b}_2 = 0,48715$, $\hat{b}_3 = 0,10416$, а для одновимірного перетворення Джонсона: $\hat{b}_0 = 0$, $\hat{b}_1 = 0,50321$, $\hat{b}_2 = 0,50562$, $\hat{b}_3 = 0,09039$. Сума квадратів відхилень для рівняння (9) складає 2,343 у разі одновимірного перетворення Джонсона і 2,057 для багатомірного перетворення Джонсона. Нелінійне регресійне рівняння (4) за перетворенням (7) сімейства S_B має вигляд

$$\hat{Y} = \hat{\varphi}_Y + \hat{\lambda}_Y \left[1 + e^{-(Z_Y - \hat{\gamma}_Y)/\hat{\eta}_Y} \right]^{-1}, \quad (10)$$

де $Z_j = \gamma_j + \eta_j \ln \frac{X_j - \varphi_j}{\varphi_j + \lambda_j - X_j}$, $\varphi_j < X_j < \varphi_j + \lambda_j$, $j = 1, 2, 3$.

Також було побудовано лінійне регресійне рівняння

$$\hat{Y} = -12,575 + 1,3682X_1 + 1,4529X_2 + 0,60411X_3. \quad (11)$$

Значення множинного коефіцієнту детермінації R^2 і середньої величини відносної помилки MMRE, які дорівнюють 0,9582 і 0,1501 відповідно, кращі для рівняння (10) з параметрами для багатомірного перетворення Джонсона в порівнянні з одновимірним, для якого ці значення складають відповідно 0,9576 і 0,1527. Значення відсотка прогнозування PRED(0,25), що дорівнює 0,8519, краще для рівняння (10) з параметрами для одновимірного перетворення Джонсона в порівнянні з багатомірним, для якого це значення складає 0,8333. Також значення R^2 , MMRE і PRED(0,25) кращі для нелінійного рівняння (10) в порівнянні з лінійним рівнянням (11), для якого ці значення дорівнюють відповідно 0,9508, 0,2882 і 0,7407.

Але основна перевага нелінійного рівняння (10) з параметрами для багатовимірного перетворення Джонсона полягає в менших шириних довірчого інтервалу та інтервалу передбачення. Були визначені нижні (LB) і верхні (UB) межі інтервалів передбачення нелінійних регресій за формулою (6) на основі відповідно одновимірного і багатовимірного перетворень Джонсона та лінійної регресії для рівня значимості 0,05. Ці межі для 27 інформаційних систем наведені у таблиці 2. Зауважимо, що нижня межа інтервалу передбачення лінійної регресії має від'ємні значення для 18 рядків даних: 7, 12, 15, 19, 22, 23, 25, 27, 29, 31, 34, 35, 37, 40, 42, 43, 49, 53. Всі значення нижньої межі інтервалу передбачення нелінійних регресій на основі перетворень Джонсона є додатними.

Таблиця 2

Межі інтервалів передбачення різних регресій

№	лінійна регресія		нелінійна регресія за перетворенням Джонсона			
			одномірне		багатомірне	
	LB	UB	LB	UB	LB	UB
1	15,218	34,948	12,651	33,605	13,529	32,937
2	27,852	47,703	21,704	49,737	22,282	49,425
3	15,646	35,29	14,819	37,881	15,505	36,819
4	63,743	84,558	66,610	83,757	61,496	90,304
5	13,400	32,933	13,037	34,189	13,825	33,267
6	57,555	77,979	54,775	78,365	52,816	83,973
7	-4,079	15,597	4,514	11,732	4,636	10,613
8	37,290	57,089	30,902	60,726	32,012	63,123
9	49,499	72,096	44,445	74,424	39,141	73,790
10	47,362	67,435	41,557	70,095	41,351	73,693
11	48,919	70,866	50,753	77,166	49,570	82,635
12	-11,904	7,937	2,027	2,644	2,100	2,714
13	0,647	20,566	5,993	16,775	6,244	15,426
14	11,472	31,144	11,614	31,216	12,163	29,786
15	-3,400	16,306	4,459	11,534	4,683	10,727
16	5,039	25,02	8,172	23,221	8,361	21,202
17	24,411	44,666	18,982	46,535	18,941	44,657
18	6,846	26,483	9,487	26,122	9,969	24,708
19	-7,741	12,245	3,557	8,559	3,619	7,649
20	1,364	21,322	6,315	17,546	6,648	16,339
21	11,997	33,16	11,499	32,523	12,320	31,367
22	-7,794	12,21	3,505	8,365	3,599	7,580
23	-15,102	5,094	1,969	2,421	2,020	2,459
24	32,777	53,333	26,524	57,118	25,649	55,809
25	-8,658	11,139	2,783	5,576	2,873	5,210
26	48,216	68,782	41,241	70,237	41,184	73,958
27	-4,261	15,44	4,320	11,045	4,521	10,234

Ширина інтервалу передбачення нелінійної регресії на основі багатовимірного перетворення Джонсона менша, ніж після одновимірного перетворення Джонсона для 44 рядків даних (1-3, 5, 7, 12-25, 27-31, 33-44, 46, 47, 49-54) та менша, ніж для лінійної регресії для 36 рядків даних: 1, 5, 7, 12-16, 18-23, 25, 27-31, 33-35, 37, 38, 40-44, 46, 47, 49,

50, 53, 54. Приблизно такі ж результати маємо для довірчих інтервалів за формулою (5).

Висновки. Удосконалено нелінійне регресійне рівняння для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на PHP на основі багатовимірного перетворення Джонсона для сімейства S_B . Це рівняння в порівнянні з іншими регресійними рівняннями має більший множинний коефіцієнт детермінації, менше значення середньої величини відносної похибки та менші ширини довірчого інтервалу і інтервалу передбачення для більшої кількості значень незалежних змінних. Надалі планується застосування інших даних для побудови нелінійного регресійного рівняння для оцінювання розміру ПЗ інформаційних систем з відкритим кодом на PHP.

Література

1. Tan H.B.K. Estimating LOC for information systems from their conceptual data models / H. B. K. Tan, Y. Zhao, H. Zhang // Proceedings of the 28th International Conference on Software Engineering (ICSE '06), Shanghai, China, May 20-28, 2006. – P. 321-330.
2. Tan H.B.K. Conceptual data model-based software size estimation for information systems / H. B. K. Tan, Y. Zhao, H. Zhang // Transactions on Software Engineering and Methodology. – 2009. – Vol. 19. – Issue 2. – October 2009. – Article No. 4. DOI: 10.1145/1134285.1134331
3. Prykhodko S.B. Constructing the non-linear regression equation to estimate the software size of open source PHP-based information systems / S. B. Prykhodko, N. V. Prykhodko, T. G. Smykodub, A. V. Spinov // Проблеми інформаційних технологій. – 2018. – № 1 (023). – С.118-125. – ISSN 1998-7005
4. Bates D.M. Nonlinear Regression Analysis and Its Applications / D. M. Bates, D. G. Watts. – New York: John Wiley & Sons, 1988. – 384 p.
5. Seber G.A.F. Nonlinear Regression / G. A. F. Seber, C. J. Wild. – New York: John Wiley & Sons, 1989. – 768 p.
6. Ryan T.P. Modern regression methods / T. P. Ryan. – New York: John Wiley & Sons, 1997. – 529 p.
7. Drapper N.R. Applied Regression Analysis / N. R. Drapper, H. Smith. – New York: John Wiley & Sons, 1998. – 736 p.
8. Johnson R.A. Applied Multivariate Statistical Analysis / R. A. Johnson, D. W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.
9. Chatterjee S. Handbook of Regression Analysis / S. Chatterjee, J. S. Simonoff. – New York: John Wiley & Sons, 2013. – 252 p.
10. Prykhodko S. Application of the Squared Mahalanobis Distance for Detecting Outliers in Multivariate Non-Gaussian Data / S. Prykhodko, N. Prykhodko, L. Makarova, A. Pukhalevych // Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, February 20-24, 2018. – P. 962-965. DOI: 10.1109/TCSET.2018.8336353

11. Mardia K.V. Measures of multivariate skewness and kurtosis with applications / K. V. Mardia // *Biometrika*. – 57. – 1970. – P. 519–530. DOI: 10.1093/biomet/57.3.519

References

1. Tan H.B.K. Estimating LOC for information systems from their conceptual data models / H. B. K. Tan, Y. Zhao, H. Zhang // *Proceedings of the 28th International Conference on Software Engineering (ICSE '06)*, Shanghai, China, May 20-28, 2006. – P. 321-330.
2. Tan H.B.K. Conceptual data model-based software size estimation for information systems / H. B. K. Tan, Y. Zhao, H. Zhang // *Transactions on Software Engineering and Methodology*. – 2009. – Vol. 19. – Issue 2. – October 2009. – Article No. 4. DOI: 10.1145/1134285.1134331
3. Prykhodko S.B. Constructing the non-linear regression equation to estimate the software size of open source PHP-based information systems / S. B. Prykhodko, N. V. Prykhodko, T. G. Smykodub, A. V. Spinov // *Problems of information technologies*. – 2018. – No. 1 (023). – P.118-125. – ISSN 1998-7005
4. Bates D.M. *Nonlinear Regression Analysis and Its Applications* / D. M. Bates, D. G. Watts. – New York: John Wiley & Sons, 1988. – 384 p.
5. Seber G.A.F. *Nonlinear Regression* / G. A. F. Seber, C. J. Wild. – New York: John Wiley & Sons, 1989. – 768 p.
6. Ryan T.P. *Modern regression methods* / T. P. Ryan. – New York: John Wiley & Sons, 1997. – 529 p.
7. Drapper N.R. *Applied Regression Analysis* / N. R. Drapper, H. Smith. – New York: John Wiley & Sons, 1998. – 736 p.
8. Johnson R.A. *Applied Multivariate Statistical Analysis* / R. A. Johnson, D. W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.
9. Chatterjee S. *Handbook of Regression Analysis* / S. Chatterjee, J. S. Simonoff. – New York: John Wiley & Sons, 2013. – 252 p.
10. Prykhodko S. Application of the Squared Mahalanobis Distance for Detecting Outliers in Multivariate Non-Gaussian Data / S. Prykhodko, N. Prykhodko, L. Makarova, A. Pukhalevych // *Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, Ukraine, February 20–24, 2018. – P. 962-965. DOI: 10.1109/TCSET.2018.8336353

11. Mardia K.V. Measures of multivariate skewness and kurtosis with applications / K. V. Mardia // *Biometrika*. – 57. – 1970. – P. 519–530. DOI: 10.1093/biomet/57.3.519

Приходько Н.В., Приходько С.Б. Нелинейное регрессионное уравнение для оценки размера программного обеспечения информационных систем с открытым кодом на PHP.

Нелинейное регрессионное уравнение для оценки размера программного обеспечения информационных систем с открытым кодом на PHP построено на основе многомерного преобразования Джонсона для семейства S_B . Это уравнение, по сравнению с другими регрессионными уравнениями, имеет больший множественный коэффициент детерминации, меньшее значение средней величины относительной погрешности, меньшие ширины доверительного интервала и интервала предсказания нелинейной регрессии.

Ключевые слова: нелинейное регрессионное уравнение, доверительный интервал, интервал предсказания, оценки размера программного обеспечения, информационная система, PHP.

Prykhodko N.V., Prykhodko S.B. The non-linear regression equation to estimate the software size of open source PHP-based information systems.

The non-linear regression equation to estimate the software size of open source PHP-based information systems is constructed on the basis of the Johnson multivariate transformation for S_B family. This equation, in comparison with other regression equations, has a larger multiple coefficient of determination, a smaller value of the mean magnitude of relative error, smaller widths of the confidence and prediction intervals of non-linear regression.

Keywords: nonlinear regression equation, confidence interval, prediction interval, estimation of software size, information system, PHP.

Приходько Н.В. – к.е.н., доцент, доцент кафедри фінансів Національного університету кораблебудування імені адмірала Макарова, e-mail: natalia.prykhodko@nuos.edu.ua

Приходько С.Б. – д.т.н., професор, завідувач кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова, e-mail: sergiy.prykhodko@nuos.edu.ua

Рецензент: д.е.н., проф. **Д'яченко Ю.Ю.**

Стаття подана 27.10.2018