

УДОСКОНАЛЕННЯ ТРЬОХФАКТОРНОЇ НЕЛІНІЙНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ ТРУДОМІСТКОСТІ РОЗРОБКИ МОБІЛЬНИХ ЗАСТОСУНКІВ

УДК 004.412:519.237.5

DOI: <https://doi.org/10.35546/2313-0687.2019.26.21-31>

Сергій Приходько,

д.т.н., професор, завідувач кафедри програмного забезпечення автоматизованих систем,
Національний університет кораблебудування імені адмірала Макарова, Миколаїв, Україна,

E-mail: sergiy.prykhodko@nuos.edu.ua, 0000-0002-2325-018X

Наталія Приходько,

к.е.н., доцент, доцент кафедри фінансів,
Національний університет кораблебудування імені адмірала Макарова, Миколаїв, Україна,

E-mail: natalia.prykhodko@nuos.edu.ua, 0000-0002-3554-7183

Катерина Книрик,

аспірант кафедри програмного забезпечення автоматизованих систем,
Національний університет кораблебудування імені адмірала Макарова, Миколаїв, Україна,

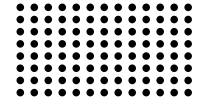
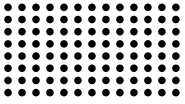
E-mail: katrin010692@gmail.com, 0000-0001-8434-4035

Анотація. Метою статті є удосконалення трьохфакторної нелінійної регресійної моделі для оцінювання трудомісткості розробки мобільних застосунків на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B .

Методи дослідження. Модель та інтервали передбачення трьохфакторної нелінійної регресії для оцінювання трудомісткості розробки мобільних застосунків побудовані на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B для негаусівських даних за допомогою відповідних методів множинного нелінійного регресійного аналізу. Методи побудови моделей, довірчих інтервалів та інтервалів передбачення нелінійних регресій засновані на множинному нелінійному регресійному аналізі з використанням багатовимірних нормалізуючих перетворень. Ці методи дозволяють враховувати кореляцію між випадковими величинами у разі нормалізації багатовимірних негаусівських даних. Загалом, це призводить до збільшення множинного коефіцієнту детермінації та відсотку передбачених значень, зменшення середньої величини відносної похибки, ширин довірчих інтервалів і інтервалів передбачення в порівнянні з лінійними моделями та нелінійними моделями, побудованими з використанням одновимірних нормалізуючих перетворень.

Основні результати дослідження. Здійснено порівняння удосконаленої моделі з моделями лінійної регресії та нелінійних регресій на основі десяткового логарифму та одновимірного перетворення Джонсона сімейства S_B .

Наукова новизна. Удосконалено нелінійну регресійну модель та рівняння інтервалу передбачення нелінійної регресії для оцінювання трудомісткості розробки мобільних застосунків в залежності від кількості екранів,



функцій та файлів мобільного застосунку на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B , що дозволяє підвищити достовірність оцінювання трудомісткості розробки мобільних застосунків у порівнянні з використанням одновимірних нормалізуючих перетворень. Модель, що побудовано, в порівнянні з іншими регресійними моделями (як лінійними, так і нелінійними), має більші значення множинного коефіцієнту детермінації і відсотка прогнозування, менше значення середньої величини відносної похибки та менші ширини інтервалу передбачення нелінійної регресії.

Практична значимість. Практична значимість отриманих результатів полягає в тому, що програмне забезпечення, що реалізує побудовану модель, розроблено Sci-мовою для системи Scilab.

Ключові слова: нелінійна регресійна модель, оцінювання трудомісткості, мобільний застосунок, нормалізуюче перетворення, негаусівські дані.

Постановка проблеми. Оцінювання трудомісткості розробки програмного забезпечення (ПЗ) є однією з важливих задач на початкових етапах його створення. Сьогодні однією з найбільш відомих моделей для оцінювання трудомісткості розробки ПЗ є COSOMO II (COststructive COst MOdel) – нелінійне регресійне рівняння, яке залежить в першу чергу від розміру ПЗ. Але на початкових етапах розробки розмір ПЗ ще невідомий. Крім того, зазначене рівняння побудовано на основі одновимірного нормалізуючого перетворення у вигляді десяткового логарифму. А таке перетворення не завжди дозволяє добре виконати нормалізацію даних. Все це веде до низької достовірності оцінювання трудомісткості розробки будь-якого ПЗ, у тому числі і мобільних застосунків.

Аналіз останніх досліджень і публікацій. Хоча розробка мобільних застосунків схожа на розробку веб-застосунків і має свої корені в більш традиційній розробці ПЗ, однак, однією з суттєвих різниць є те, що відповідні застосунки часто створюють під конкретний мобільний пристрій для того, щоб скористатися його унікальними можливостями [1]. Тому протягом останнього десятиріччя було опубліковано багато різних моделей для оцінювання трудомісткості розробки мобільних застосунків, включаючи регресійні, як лінійні, так і нелінійні [2-5]. Саме регресійні моделі описують трудомісткість як випадкову величину в залежності від певних факторів. А враховуючи те, що розподіл трудомісткості як правило не є гаусівським [4, 5], потрібно використовувати саме нелінійні регресійні моделі, а їх побудову вести на основі багатовимірних нормалізуючих перетворень [5, 6]. Саме тому у [5] була запропонована трьохфакторна нелінійна регресійна модель для оцінювання трудомісткості розробки мобільних застосунків

в залежності від кількості екранів, функцій та файлів мобільного застосунку на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B . Але при її побудові на останній ітерації відносна точність оцінок параметрів складала 1%, що могло вплинути на якість моделі в першу чергу щодо збільшення ширин довірчого інтервалу нелінійної регресії. Тому виникає необхідність в удосконаленні цієї моделі.

Мета дослідження полягає в удосконаленні трьохфакторної нелінійної регресійної моделі для оцінювання трудомісткості розробки мобільних застосунків на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B , що дозволяє підвищити достовірність оцінювання трудомісткості розробки мобільних застосунків у порівнянні з іншими регресійними моделями.

Виклад матеріалу дослідження. Для досягнення мети дослідження, як і в [5], ми скористалися методами наведеними в [6]. Згідно з [6] спочатку виконується нормалізація багатовимірних негаусівських даних за багатовимірним нормалізуючим перетворенням. У якості таких даних використані дані, що наведені у табл. 1 [4, 5]. Ці дані отримані шляхом доповнення чотиривимірного набору даних про фактичну трудомісткість розробки Y у людино-годинах, кількість екранів X_1 , функцій X_2 і файлів X_3 для 17 застосунків з [7] ще даними для 21 мобільного застосунку (рядки з 18 по 38).

Як і в [4, 5] для виявлення викидів у даних табл. 1 використовувався метод, оснований на багатовимірних нормалізуючих перетвореннях і квадраті відстані Махаланобіса MD^2 . Також як і в [4, 5] було отримано, що немає викидів в даних табл. 1 для рівня значимості 0,005 та чотиривимірного перетворення Джонсона для сімейства S_B тому, що всі значення MD^2 менше ніж величина квантіля розподілу χ^2 , яка становить 14,86

Таблиця 1 – Дані та границі інтервалів передбачення нелінійних регресій

№	Y	X ₁	X ₂	X ₃	Границі інтервалів передбачення нелінійних регресій					
					LB ₁	UB ₁	LB ₂	UB ₂	LB ₆	UB ₆
1	192	5	4	3	60,47	377,26	91,85	289,05	141,59	216,58
2	272	5	4	3	60,47	377,26	91,85	289,05	-	-
3	288	3	2	2	88,63	524,10	153,04	362,13	233,55	323,15
4	116	6	6	4	51,09	352,88	75,70	268,17	115,56	185,70
5	372	5	5	4	54,51	362,37	-	-	-	-
6	504	9	8	6	90,07	453,34	-	-	-	-
7	28	6	7	2	-0,71	232,51	20,09	160,96	25,13	64,56
8	176	6	7	3	18,85	277,78	37,91	204,55	-	-
9	364	10	11	9	157,35	665,23	267,93	404,24	342,22	409,18
10	120	10	10	5	48,66	363,80	73,41	272,31	110,82	183,41
11	22	6	5	4	70,82	402,23	-	-	-	-
12	224	11	6	2	73,52	447,01	111,94	322,06	171,73	256,38
13	24	2	2	1	-23,85	170,88	8,80	123,53	15,40	49,06
14	200	11	7	4	106,51	511,55	155,93	351,04	-	-
15	160	6	6	7	100,58	490,02	148,29	343,97	-	-
16	120	2	2	1	-23,85	170,88	8,80	123,53	-	-
17	96	4	4	1	-33,37	149,22	1,90	94,96	-	-
18	202	6	5	4	70,82	402,23	103,42	301,71	160,44	237,61
19	145	4	3	2	49,16	353,26	80,77	277,21	124,75	197,83
20	198	6	5	4	70,82	402,23	103,42	301,71	160,44	237,61
21	146	4	3	2	49,16	353,26	80,77	277,21	124,75	197,83
22	191	6	6	5	66,20	392,46	96,35	294,74	151,10	227,57
23	99	3	3	2	24,68	289,97	51,04	229,36	76,40	136,18
24	382	11	12	9	140,13	624,93	257,61	400,94	326,63	397,31
25	270	9	10	8	93,37	477,23	138,38	338,81	218,51	301,14
26	282	12	7	3	104,63	532,46	163,82	362,10	246,66	331,09
27	213	10	5	2	78,45	452,66	117,16	324,35	181,01	264,80
28	322	11	7	5	126,75	560,32	184,64	367,37	278,36	354,89
29	290	10	6	4	109,14	513,13	157,51	350,67	239,89	320,15
30	223	7	7	6	78,58	425,22	112,51	312,36	177,74	257,11
31	241	5	5	6	84,92	449,29	127,21	327,39	204,64	286,26
32	87	5	5	2	17,06	267,28	37,56	200,04	53,04	103,79
33	36	3	3	1	-29,04	153,63	5,31	105,65	15,20	47,36
34	216	8	7	5	77,05	418,56	108,93	307,94	168,47	246,62
35	67	5	6	2	1,39	233,20	22,41	164,98	29,01	69,61
36	115	7	7	3	30,98	306,43	50,32	228,90	72,42	131,49
37	36	2	2	1	-23,85	170,88	8,80	123,53	15,40	49,06
38	98	3	3	2	24,68	289,97	51,04	229,36	76,40	136,18

для рівня значущості 0,005. Як і в [4, 5] оцінки параметрів багатовимірного перетворення Джонсона для сімейства S_B були знайдені за методом максимальної правдоподібності: $\hat{\gamma}_Y = 5,69898$, $\hat{\gamma}_1 = 0,524119$, $\hat{\gamma}_2 = 0,776179$, $\hat{\gamma}_3 = 0,540973$, $\hat{\eta}_Y = 2,40219$, $\hat{\eta}_1 = 0,743879$, $\hat{\eta}_2 = 0,79545$, $\hat{\eta}_3 = 0,534447$, $\hat{\phi}_Y = -114,5452$, $\hat{\phi}_1 = 1,7242$, $\hat{\phi}_2 = 1,6885$, $\hat{\phi}_3 = 0,90$, $\hat{\lambda}_Y = 3328,564$, $\hat{\lambda}_1 = 12,3743$, $\hat{\lambda}_2 = 12,091$ і $\hat{\lambda}_3 = 8,30648$.

Далі для нормалізованих за чотиривимірним перетворенням Джонсона сімейства SB даних будуємо лінійну регресійну модель [6]

$$Z_Y = \hat{Z}_Y + \varepsilon = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3 + \varepsilon, \quad (1)$$

де ε – випадкова величина з розподілом Гаусу, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Параметри моделі (1) оцінювалися методом найменших квадратів, а їх оцінки є такими як і в [5]: $\hat{b}_0 = 0$, $\hat{b}_1 = 0,808152$, $\hat{b}_2 = -0,928296$, $\hat{b}_3 = 0,854262$. Сума квадратів відхилень для моделі (1) склала 14,867.

За (1) та чотиривимірним перетворенням Джонсона сімейства S_B будуємо нелінійну регресійну модель для оцінювання трудомісткості розробки мобільних застосунків [5, 6]

$$Y = \hat{\phi}_Y + \hat{\lambda}_Y \left[1 + e^{-\left(\hat{Z}_Y + \varepsilon - \hat{\gamma}_Y\right) / \hat{\eta}_Y} \right]^{-1}, \quad (2)$$

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

Якість побудованої моделі (2) була перевірена за множинним коефіцієнтом детермінації R^2 , середньою

$$\hat{Y}_{PI} = \psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, \nu} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (\mathbf{z}_X^+)^T \left[(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right), \quad (3)$$

де ψ_Y – перша компонента багатовимірного нормалізуючого перетворення $\mathbf{T} = \psi(\mathbf{P})$ негаусівського випадкового вектору $\mathbf{P} = \{Y, X_1, X_2, \dots, X_k\}^T$ в гаусівський випадковий вектор $\mathbf{T} = \{Z_Y, Z_1, Z_2, \dots, Z_k\}^T$; $t_{\alpha/2, \nu}$ – квантіль t-розподілу Стюдента з кількістю ступенів вільності ν та рівнем значущості $\alpha/2$; \mathbf{Z}_X^+ – матриця центрованих регресорів, яка у нашому

випадку містить значення $Z_{1_i} - \bar{Z}_1$, $Z_{2_i} - \bar{Z}_2$, $Z_{3_i} - \bar{Z}_3$; $\mathbf{z}_X^+ = \{Z_{1_i} - \bar{Z}_1, Z_{2_i} - \bar{Z}_2, Z_{3_i} - \bar{Z}_3\}^T$; $S_{Z_Y}^2 = \frac{1}{\nu} \sum_{i=1}^N (Z_{Y_i} - \hat{Z}_{Y_i})^2$, $\nu = N - k - 1$; $(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ - k \times k$ матриця

величиною відносної помилки MMRE і відсотком прогнозування PRED(0,25), які використовуються в якості стандартних оцінок результатів передбачення за допомогою регресійних моделей і в програмній інженерії [8, 9]. Значення R^2 , MMRE і PRED(0,25) дорівнюють 0,579, 0,493 і 0,526 відповідно. Вони вказують на незадовільну якість рівняння (2) з оцінками параметрів, що були отримані за даними з табл.1 з 38 мобільних застосунків. Тому як і в [5] для побудови нелінійної регресійної моделі для оцінювання трудомісткості розробки мобільних застосунків далі було застосовано метод покращення нелінійних регресійних моделей на основі багатовимірних нормалізуючих перетворень із застосуванням інтервалів передбачення. Суть цього методу полягає у наступному. Спочатку на першому етапі, як це зазвичай робиться, початкові негаусівські дані перевіряються на наявність викидів і, якщо останні знайдено, то вони відкидаються. На першому етапі рівень значущості приймається 0,005. Далі на другому етапі будується нелінійна регресійна модель із застосуванням відповідного методу на основі багатовимірних нормалізуючих перетворень [6]. Після цього на третьому етапі для рівня значущості, що дорівнює 0,05, визначаються границі інтервалу передбачення нелінійної регресії за методом, наведеним в [6]. І на останньому четвертому етапі перевіряють, чи є серед даних, за якими будувалася регресійна модель такі, що виходять за визначені границі інтервалу передбачення. Та, якщо останні знайдено, то вони відкидаються і ми повторюємо всі етапи для нових даних. Якщо таких викидів не було, то повторення етапів припиняється, відповідна нелінійна регресійна модель побудована.

Для визначення нижньої і верхньої границь інтервалів передбачення нелінійних регресій побудовано відповідні рівняння за [6]

$$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ = \begin{pmatrix} S_{Z_1Z_1} & S_{Z_1Z_2} & \dots & S_{Z_1Z_k} \\ S_{Z_1Z_2} & S_{Z_2Z_2} & \dots & S_{Z_2Z_k} \\ \dots & \dots & \dots & \dots \\ S_{Z_1Z_k} & S_{Z_2Z_k} & \dots & S_{Z_kZ_k} \end{pmatrix}, \quad (4)$$

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

У нашому випадку $k=3$.

Для нелінійної регресійної моделі (2) з оцінками параметрів, що були отримані за даними табл. 1 з 38 застосунків для рівня значущості 0,05 виявилось: значення Y для трьох застосунків (5, 6 і 11) виходять за визначені границі інтервалу передбачення. У табл. 1 ліві границі інтервалу передбачення, що отриманий на першій і другій ітераціях за (3), позначені як LB_1 і LB_2 , а праві – як UB_1 і UB_2 відповідно. Всього було 6 таких ітерацій, після яких залишилося 29 застосунків (1, 3, 4, 7, 9, 10, 12, 13, 18-38). У табл. 1 ліва границя інтервалу передбачення, що отриманий на шостій ітерації за (3), позначена як LB_6 , а права – як UB_6 . Зазначимо, що отримані на п'ятій ітерації ширини інтервалів передбачення у порівнянні з тими, що наведені у [5], виявилися меншими на 10,0–22,5%. За рахунок цього на п'ятій ітерації ще один застосунок 14 був вилучений. Зазначимо, що у [5] п'ята ітерація була останньою. У нашому випадку, що розглядається, на шостій ітерації викидів не було, відповідна нелінійна регресійна модель побудована. На шостій ітерації для даних з 29 мобільних застосунків оцінки параметрів чотиривимірного перетворення Джонсона сімейства S_B такі: $\hat{\gamma}_Y = 0,638164$, $\hat{\gamma}_1 = 0,387413$, $\hat{\gamma}_2 = 0,840380$, $\hat{\gamma}_3 = 0,477514$, $\hat{\eta}_Y = 1,12311$, $\hat{\eta}_1 = 0,659463$, $\hat{\eta}_2 = 0,831303$, $\hat{\eta}_3 = 0,632614$, $\hat{\phi}_Y = -28,4433$, $\hat{\phi}_1 = 1,82645$, $\hat{\phi}_2 = 1,59121$, $\hat{\phi}_3 = 0,657479$, $\hat{\lambda}_Y = 543,161$, $\hat{\lambda}_1 = 11,5548$, $\lambda_2 = 12,9939$ і $\hat{\lambda}_3 = 8,63368$, оцінки параметрів рівняння (1) наступні: $\hat{b}_0 = 0$, $\hat{b}_1 = 1,17702$, $\hat{b}_2 = -1,43269$, $\hat{b}_3 = 1,18398$, а матриця (4) для визначення нижньої і верхньої границь інтервалів передбачення нелінійної регресії за (3) є такою:

$$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ = \begin{pmatrix} 29,0 & 24,8 & 18,7 \\ 24,8 & 29,0 & 23,6 \\ 18,7 & 23,6 & 29,0 \end{pmatrix}. \quad (5)$$

Зазначимо, сума квадратів відхилень для моделі (1) на шостій ітерації склала 0,6189, що у 24 рази менше за

відповідну суму на першій ітерації. Сума квадратів відхилень для моделі (2) на шостій ітерації склала 4669,6, що у 42 рази менше за відповідну суму на першій ітерації. Значення R^2 , MMRE і PRED(0,25) дорівнюють 0,984, 0,103 і 0,862 відповідно і вказують на добру якість моделі (2) з оцінками параметрів, що були отримані на шостій ітерації за даними табл. 1 з 29 мобільних застосунків.

Для порівняння нелінійної регресійної моделі (2), яка отримана на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B , з іншими моделями було побудовано лінійну регресійну модель та нелінійні регресійні моделі із застосуванням одновимірних нормалізуючих перетворень у вигляді десяткового логарифму та Джонсона сімейства S_B з оцінками параметрів, що були також знайдені за даними табл. 1 з 29 мобільних застосунків. Лінійна регресійна модель для оцінювання трудомісткості розробки мобільних застосунків має вигляд

$$Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \varepsilon, \quad (6)$$

де оцінки параметрів такі: $\hat{b}_0 = 36,8484$, $\hat{b}_1 = 31,3019$, $\hat{b}_2 = -43,4205$, $\hat{b}_3 = 50,5983$.

Сума квадратів відхилень для моделі (6) склала 40 908,4, що майже у 9 разів більше за відповідну суму для моделі (2), яка отримана на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B . Значення R^2 , MMRE і PRED(0,25) для моделі (6) дорівнюють 0,857, 0,228 і 0,793 відповідно. Хоча ці значення вказують на непогану якість моделі (6) з оцінками параметрів, що були отримані за даними табл. 1 з 29 мобільних застосунків, але вони гірші за відповідні значення для моделі (2), яка побудована на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B . Крім того була перевірена нульова гіпотеза про нормальність закону розподілу випадкової величини ε моделі (6) за критерієм Пірсона χ^2 . Для вибірки значень випадкової величини ε значення χ^2 , яке дорівнює 23,37, більше за $\chi_{кр}^2$, що становить 7,81 для 3 ступенів вільності та 0,05 рівня значущості. Тобто, цю гіпотезу про нормальність розподілу випадкової величини ε потрібно відкинути. Це свідчить про відсутність теоретичного обґрунтування можливості використання моделі лінійної регресії (6) і призводить до необхідності побудови нелінійної регресійної моделі для оцінювання трудомісткості розробки мобільних застосунків.

Для побудови нелінійної регресійної моделі для оцінювання трудомісткості розробки мобільних застосунків негаусівськи дані табл. 1 з 29 застосунків ми нормалізуємо за одновимірним перетворенням у формі десяткового логарифму. Далі для нормалізованих даних будуємо лінійну регресійну модель (1), параметри якої оцінювалися методом найменших квадратів, а їх оцінки є такими: $\hat{b}_0 = 1,72206$, $\hat{b}_1 = 1,75435$, $\hat{b}_2 = -2,16953$, $\hat{b}_3 = 1,29794$. Сума квадратів відхилень для моделі (1) у цьому разі склала 0,29135. Після чого за (1) та перетворенням у вигляді десяткового логарифму будуємо нелінійну регресійну модель для оцінювання трудомісткості розробки мобільних застосунків

$$Y = 10^{\hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}. \quad (7)$$

Сума квадратів відхилень для моделі (7) склала 53 624,7, що понад 11 разів більше за відповідну суму для моделі (2), яка отримана на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B . Значення R^2 , MMRE і PRED(0,25) для моделі (7) дорівнюють 0,812, 0,198 і 0,690 відповідно. Значення PRED(0,25) вказує на незадовільну якість моделі (6) з оцінками параметрів, що були отримані за даними табл. 1 з 29 мобільних застосунків. Значення всіх трьох показників гірші за відповідні значення для моделі (2), яка побудована на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B .

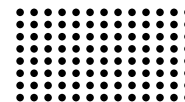
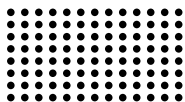
Нелінійну регресійну модель (2), яка отримана на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B , ми також порівняли з нелінійною регресійною моделлю із застосуванням одновимірного нормалізуючого перетворення Джонсона сімейства S_B з оцінками параметрів, що були знайдені за даними табл. 1 з 29 мобільних застосунків. Для даних з 29 застосунків оцінки параметрів одновимірних перетворень Джонсона сімейства S_B такі: $\hat{\gamma}_Y = 0,250621$, $\hat{\gamma}_1 = 0,147151$, $\hat{\gamma}_2 = 0,471451$, $\hat{\gamma}_3 = 0,605927$, $\hat{\eta}_Y = 0,548155$, $\hat{\eta}_1 = 0,519404$, $\hat{\eta}_2 = 0,558891$, $\hat{\eta}_3 = 0,575457$, $\hat{\phi}_Y = 21,1791$, $\hat{\phi}_1 = 1,90$, $\hat{\phi}_2 = 1,90$, $\hat{\phi}_3 = 0,90$, $\hat{\lambda}_Y = 367,0776$, $\hat{\lambda}_1 = 10,2064$, $\hat{\lambda}_2 = 10,3804$ і $\hat{\lambda}_3 = 8,58177$, а оцінки параметрів рівняння (1) наступні: $\hat{b}_0 = 0$, $\hat{b}_1 = 1,19747$, $\hat{b}_2 = -1,43924$, $\hat{b}_3 = 1,22071$. Сума квадратів відхилень для моделі (2) з параметрами, що отримані на основі одновимірного нормалізуючого перетворення Джонсона сімейства S_B ,

склала 39 265,6, що понад 8 разів більше за відповідну суму для моделі (2) з параметрами, які отримані на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B . Значення R^2 , MMRE і PRED(0,25) для моделі (2) на основі одновимірного нормалізуючого перетворення Джонсона сімейства S_B дорівнюють 0,863, 0,188 і 0,690 відповідно. Значення PRED(0,25) вказує на незадовільну якість моделі (2) на основі одновимірного нормалізуючого перетворення Джонсона сімейства S_B . Значення всіх трьох показників гірші за відповідні значення для моделі (2), яка побудована на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B .

Зазначимо, що основна перевага моделі (2) на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B у порівнянні з лінійною моделлю (6) та нелінійними регресійними моделями із застосуванням одновимірних нормалізуючих перетворень у вигляді десяткового логарифму (Log10) та Джонсона сімейства S_B полягає у менших шириних інтервалу передбачення нелінійної регресії трудомісткості розробки мобільних застосунків практично для всіх даних. Границі інтервалів передбачення нелінійних регресій трудомісткості розробки 29 мобільних застосунків з табл. 1 наведені у табл. 2 для лінійної моделі (6), нелінійних моделей (7) та (2) для двох перетворень Джонсона сімейства S_B : одновимірного і чотиривимірного.

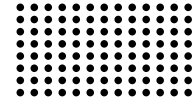
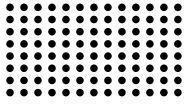
Данні табл. 2 вказують на те, що модель (2) з відповідними параметрами для чотиривимірного перетворення Джонсона сімейства S_B у порівнянні з моделями (6) та (7) має значно менші ширини інтервалу передбачення (від 97 % до 444 % у порівнянні з лінійною моделлю та від 22 % до 527 % у порівнянні з нелінійною моделлю на основі Log10) для всіх 29 мобільних застосунків. Також з табл. 2 можна побачити, що ширини довірчого інтервалу нелінійної регресії на основі чотиривимірного перетворення Джонсона сім'ї S_B менші, ніж для одновимірного перетворення Джонсона для 28 застосунків від 16% до 329% (окрім застосунку 33, для якого ширина менше на 9%).

Кращі показники оцінювання трудомісткості розробки мобільних застосунків за моделлю нелінійної регресії на основі чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B можна в першу чергу пояснити кращою нормалізацією, яка перевірялася за відомими критеріями [10]. Так, якщо за критерієм на



Таблиця 2 – Границі інтервалів передбачення лінійної та нелінійних регресій

№	Y	Лінійної регресії		Одновимірне перетворення Log10		Одновимірне перетворення Джонсона		Чотиривимірне перетворення Джонсона	
		LB	UB	LB	UB	LB	UB	LB	UB
1	192	85,9	257,1	107,8	309,2	44,4	340,6	141,6	216,6
3	288	56,8	233,5	111,5	351,7	136,9	384,7	233,6	323,2
4	116	81,5	251,6	89,7	256,0	40,7	332,4	115,6	185,7
7	28	-72,1	115,9	24,9	78,1	23,0	162,5	25,1	64,6
9	364	234,9	420,4	164,5	495,7	181,8	385,1	342,2	409,2
10	120	77,1	260,2	95,4	281,0	41,5	340,4	110,8	183,4
12	224	128,7	314,9	100,3	317,6	68,2	370,7	171,7	256,4
13	24	-24,9	151,3	22,6	69,1	21,7	79,7	15,4	49,1
18	202	124,4	295,5	132,9	381,0	56,0	356,2	160,4	237,6
19	145	46,7	219,3	79,8	232,0	35,6	317,2	124,8	197,8
20	198	124,4	295,5	132,9	381,0	56,0	356,2	160,4	237,6
21	146	46,7	219,3	79,8	232,0	35,6	317,2	124,8	197,8
22	191	131,1	303,1	119,0	344,2	55,3	356,3	151,1	227,6
23	99	15,0	188,4	48,1	140,4	27,3	256,1	76,4	136,2
24	382	221,9	409,1	160,7	485,7	71,6	375,0	326,6	397,3
25	270	198,8	379,5	144,9	432,9	98,2	376,9	218,5	301,1
26	282	167,2	353,5	144,1	440,2	290,2	387,7	246,7	331,1
27	213	142,1	325,8	126,0	398,9	61,6	364,9	181,0	264,8
28	322	239,9	420,5	243,4	723,3	197,5	384,9	278,4	354,9
29	290	202,4	381,1	215,0	640,8	124,4	379,7	239,9	320,2
30	223	168,7	342,5	141,0	410,4	72,7	368,0	177,7	257,1
31	241	187,9	371,8	159,0	481,3	82,6	373,3	204,6	286,3
32	87	-9,4	164,3	38,9	113,4	25,2	219,4	53,0	103,8
33	36	-36,5	138,6	19,1	58,3	21,4	50,7	15,2	47,4
34	216	151,0	321,6	141,8	406,2	67,3	364,3	168,5	246,6
35	67	-56,6	124,7	25,6	78,3	23,2	166,8	29,0	69,6
36	115	15,7	192,0	57,4	166,9	29,7	282,7	72,4	131,5
37	36	-24,9	151,3	22,6	69,1	21,7	79,7	15,4	49,1
38	98	15,0	188,4	48,1	140,4	27,3	256,1	76,4	136,2



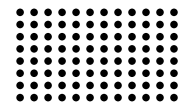
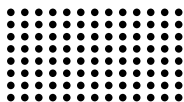
основі квадрата відстані Махаланобіса гіпотеза про нормальність багатовимірного закону розподілу нормалізованих за допомогою чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B даних для 29 застосунків з табл. 1 приймається для рівня значущості 0,05, то у випадку застосування одновимірних перетворень – відкидається. Також про негаусівський розподіл чотиривимірних даних з табл. 1 свідчить оцінка багатовимірного ексцесу β_2 , яка визначалася за [11]. Відомо, що для m -вимірного нормального розподілу $\beta_2 = m(m + 2)$. У нашому випадку $\beta_2 = 24$. Для чотиривимірних даних табл. 1 з 29 застосунків, що нормалізовані за допомогою одновимірного перетворення Джонсона сім'ї S_B , оцінка β_2 дорівнює 34,82, що на 45% перевищує теоретичне значення. У разі застосування чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B для зазначених даних оцінка β_2 дорівнює 25,09, що всього на 4,6% перевищує теоретичне значення.

Висновки. Удосконалено нелінійну регресійну модель та рівняння інтервалу передбачення нелінійної регресії для оцінювання трудомісткості розробки мобільних застосунків в залежності від кількості

екранів, функцій та файлів мобільного застосунку на основі чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B , що дозволяє підвищити достовірність оцінювання трудомісткості розробки мобільних застосунків у порівнянні з використанням одновимірних нормалізуючих перетворень. Модель, що побудовано, в порівнянні з іншими регресійними моделями (як лінійними, так і нелінійними), має більші значення множинного коефіцієнту детермінації і відсотка прогнозування, менше значення середньої величини відносної похибки та менші ширини інтервалу передбачення нелінійної регресії. На прикладі удосконалення трьохфакторної нелінійної регресійної моделі для оцінювання трудомісткості розробки мобільних застосунків підтверджено працездатність методу покращення нелінійних регресійних моделей на основі багатовимірних нормалізуючих перетворень із застосуванням квадрата відстані Махаланобіса та інтервалів передбачення. В майбутньому планується використання інших наборів даних для побудови нелінійної регресійної моделі для оцінювання трудомісткості розробки мобільних застосунків.

СПИСОК ЛІТЕРАТУРИ:

1. Rouse, M. (2019) *Mobile application development*. Retrieved December 19, 2019, from <https://searchmicroservices.techtarget.com/definition/mobile-application-development>
2. Francese, R., Gravino, C., Risi, M., Scanniello, G., & Tortora, G. (2015) On the use of requirements measures to predict software project and product measures in the context of Android mobile apps: A preliminary study. In *Proceedings of the 41st Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2015)*. (August 26–28, 2015, Funchal, Portugal) (pp. 357–364). Funchal. <https://doi.org/10.1109/SEAA.2015.22>
3. Shahwaiz, S.A., Malik, A.A., & Sabahat, N. (2016) A parametric effort estimation model for mobile apps. In *Proceedings of the 19th International Multi-Topic Conference (INMIC 2016)* (pp. 1–6). Islamabad. <https://doi.org/10.1109/INMIC.2016.7840114>
4. Приходько, С.Б., Приходько, Н.В., & Книрик, К.О. (2019) Трьохфакторне нелінійне регресійне рівняння для оцінювання трудомісткості розробки мобільних застосунків у фазі планування. *Науковий журнал «Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки»*, 30(69), № 5, 154–160. <https://doi.org/10.32838/2663-5941/2019.5-1/25>
5. Prykhodko, S., Prykhodko, N., Knyrik, K., & Pukhalevych, A. (2019) Mathematical Modeling of Effort of Mobile Application Development in a Planning Phase. In *Proceedings of the 1st International Workshop on Information-Communication Technologies & Embedded Systems, CEUR Workshop Proceedings*, Vol. 2516 (pp. 96–105.). Mykolaiv, CEUR-WS.org
6. Prykhodko, N.V., & Prykhodko, S.B. (2018) Constructing the non-linear regression models on the basis of multivariate normalizing transformations. *Electronic modeling*, 6(40), 101–110. <https://doi.org/10.15407/emodel.40.06.101>
7. Arnuphaptrairong, T., & Suksawasat, W. (2017) An Empirical Validation of Mobile Application Effort Estimation Models. In *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2017)* (pp. 697–701). Hong Kong.
8. Foss, T., Stensrud, E., Kitchenham, B., & Myrtveit, I. (2003) A simulation study of the model evaluation criterion MMRE. *IEEE Transactions on software engineering*, 11(29), 985–995.
9. Port, D., & Korte, M. (2008) Comparative studies of the model evaluation criterions MMRE and PRED in software cost estimation research. In *Proceedings of the 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (pp. 51–60). ACM, New York.
10. Olkin, I., Sampson, A.R. (2001) Multivariate Analysis: Overview. In N.J. Smelser, & P.B. Baltes (Eds.) 1st edn., *International encyclopedia of social & behavioral sciences* (pp. 10240–10247). Elsevier, Pergamon.
11. Mardia, K.V. (1970) Measures of multivariate skewness and kurtosis with applications. *Biometrika*, Vol. 57, 519–530. <https://doi.org/10.1093/biomet/57.3.519>



IMPROVING THE THREE-FACTOR NON-LINEAR REGRESSION MODEL TO ESTIMATE MOBILE APPLICATION DEVELOPMENT EFFORTS

Sergiy Prykhodko,

Dr. Sc., Professor, Head of Department of Software of Automated Systems,
Admiral Makarov National University of Shipbuilding, Nikolayev, Ukraine,
e-mail: sergiy.prykhodko@nuos.edu.ua, 0000-0002-2325-018X

Natalia Prykhodko,

PhD, Assistant Professor, Assistant Professor of Finance Department,
Admiral Makarov National University of Shipbuilding, Nikolayev, Ukraine,
e-mail: natalia.prykhodko@nuos.edu.ua, 0000-0002-3554-7183

Kateryna Knyrik,

Postgraduate Student of Department of Software of Automated Systems,
Admiral Makarov National University of Shipbuilding, Nikolayev, Ukraine,
e-mail: katrin010692@gmail.com, 0000-0001-8434-4035

Abstract. The goal of the article is to improve the three-factor non-linear regression model for estimating the efforts of development of mobile applications based on the Johnson four-variate normalizing transformation for S_B family.

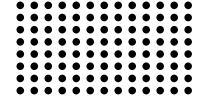
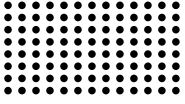
Research methods. The model and prediction intervals of the three-factor non-linear regression model to estimate mobile application development efforts are constructed on the basis of the Johnson four-variate normalizing transformation for non-Gaussian data with the help of appropriate methods of multiple non-linear regression analysis. The methods based on the multiple non-linear regression analysis using the multivariate normalizing transformations to build the models, confidence and prediction intervals of multiple non-linear regressions are used. The methods allow to take into account the correlation between random variables in the case of normalization of multivariate non-Gaussian data. In general, this leads to an increase of a multiple coefficient of determination and percentage of prediction, a reduction of the mean magnitude of relative error, the widths of the confidence and prediction intervals in comparison with the linear models and nonlinear models constructed by univariate normalizing transformations.

Main results of research. Comparison of the improved model with the linear model and non-linear regression models based on the decimal logarithm and the Johnson univariate transformation for S_B family has been performed.

The scientific novelty of obtained results is that the three-factor non-linear regression model and prediction interval equation of nonlinear regression to estimate mobile application development efforts are improved based on the number of screens, features, and mobile application files using the Johnson four-variate transformation for S_B family. This allows us to increase the confidence of estimating the mobile application development efforts compared to the use of univariate normalizing transformations. The improved model, in comparison with other regression models (both linear and non-linear), has a larger multiple coefficient of determination, a larger value of percentage of prediction, a smaller value of the mean magnitude of relative error, and smaller widths of the prediction intervals of multiple non-linear regression.

The practical significance. The practical significance of obtained results is that the software realizing the improved model is developed in the sci-language for Scilab.

Keywords: *non-linear regression model, effort estimation, mobile application, normalizing transformation, non-Gaussian data.*



СОВЕРШЕНСТВОВАНИЕ ТРЕХФАКТОРНОЙ НЕЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ ДЛЯ ОЦЕНКИ ТРУДОЕМКОСТИ РАЗРАБОТКИ МОБИЛЬНЫХ ПРИЛОЖЕНИЙ

Сергей Приходько,

д.т.н., профессор, заведующий кафедрой программного обеспечения автоматизированных систем,
Национальный университет кораблестроения имени адмирала Макарова, Николаев, Украина,
e-mail: sergiy.prykhodko@nuos.edu.ua, 0000-0002-2325-018X

Наталья Приходько,

к.э.н., доцент, доцент кафедры финансов,
Национальный университет кораблестроения имени адмирала Макарова, Николаев, Украина,
e-mail: natalia.prykhodko@nuos.edu.ua, 0000-0002-3554-7183

Екатерина Кнырик,

аспирант кафедры программного обеспечения автоматизированных систем,
Национальный университет кораблестроения имени адмирала Макарова, Николаев, Украина,
e-mail: katrin010692@gmail.com, 0000-0001-8434-4035

Аннотация. Целью статьи является совершенствование трехфакторной нелинейной регрессионной модели для оценки трудоемкости разработки мобильных приложений на основе четырехмерного нормализующего преобразования Джонсона семейства S_B .

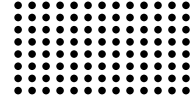
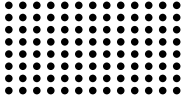
Методы исследования. Модель и интервалы предсказания трехфакторной нелинейной регрессии для оценки трудоемкости разработки мобильных приложений построены на основе четырехмерного нормализующего преобразования Джонсона семейства S_B для негауссовских данных с помощью соответствующих методов множественного нелинейного регрессионного анализа. Методы построения моделей, доверительных интервалов и интервалов предсказания нелинейных регрессий основаны на множественном нелинейном регрессионном анализе с использованием многомерных нормализующих преобразований. Эти методы позволяют учитывать корреляцию между случайными величинами в случае нормализации многомерных негауссовских данных. В общем, это приводит к увеличению множественного коэффициента детерминации и проценту предсказанных значений, уменьшению средней величины относительной погрешности ширин доверительных интервалов и интервалов предсказания по сравнению с линейными моделями и нелинейными моделями, построенными с использованием одномерных нормализующих преобразований.

Основные результаты исследования. Проведено сравнение усовершенствованной модели с моделями линейной регрессии и нелинейных регрессий на основе десятичного логарифма и одномерного преобразования Джонсона семейства S_B .

Научная новизна. Усовершенствовано нелинейную регрессионную модель и уравнения интервала предсказания нелинейной регрессии для оценки трудоемкости разработки мобильных приложений в зависимости от количества экранов, функций мобильного приложения на основе четырехмерного нормализующего преобразования Джонсона семейства S_B , что позволяет повысить достоверность оценки трудоемкости разработки мобильных приложений по сравнению с использованием одномерных нормализующих преобразований. Построенная модель по сравнению с другими регрессионными моделями (как линейными, так и нелинейными), имеет большие значения множественного коэффициента детерминации и процента прогнозирования, меньшее значение средней величины относительной погрешности и меньшие ширины интервала предсказания нелинейной регрессии.

Практическая значимость. Практическая значимость полученных результатов заключается в том, что программное обеспечение, реализующее построенную модель, разработаны Sci-языком для системы Scilab.

Ключевые слова: нелинейная регрессионная модель, оценка трудоемкости, мобильное приложение, нормализующее преобразования, негауссовские данные.



REFERENCES:

1. Rouse, M. Mobile application development / M. Rouse // URL: <https://searchmicroservices.techtarget.com/definition/mobile-application-development> (дата звернення: 19.12.2019).
2. Francese, R. On the use of requirements measures to predict software project and product measures in the context of Android mobile apps: A preliminary study / R. Francese, C. Gravino, M. Risi, G. Scanniello, G. Tortora // Proceedings of the 41st Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2015). (August 26–28, 2015, Funchal, Portugal). Funchal, 2015. – P. 357–364. – DOI: 10.1109/SEAA.2015.22
3. Shahwaiz, S.A. A parametric effort estimation model for mobile apps / S.A. Shahwaiz, A.A. Malik, N. Sabahat // Proceedings of the 19th International Multi-Topic Conference (INMIC 2016). (December 5–6, 2016, Islamabad, Pakistan). Islamabad, 2016. – P. 1–6. – DOI: 10.1109/INMIC.2016.7840114
4. Приходько, С.Б. Трьохфакторне нелінійне регресійне рівняння для оцінювання трудомісткості розробки мобільних застосунків у фазі планування / С.Б. Приходько, Н.В. Приходько, К.О. Книрик // Науковий журнал «Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки», Том 30 (69), № 5, 2019. – С. 154–160. – DOI: 10.32838/2663-5941/2019.5-1/25
5. Prykhodko, S. Mathematical Modeling of Effort of Mobile Application Development in a Planning Phase / Sergiy Prykhodko, Natalia Prykhodko, Kateryna Knyrik, Andrii Pukhalevych // Proceedings of the 1st International Workshop on Information-Communication Technologies & Embedded Systems, November, 14–15, 2019, Mykolaiv, Ukraine. CEUR Workshop Proceedings. – 2019. – Vol. 2516. – Mykolaiv, Ukraine. CEUR-WS.org – P. 96–105. – ISSN 1613–0073.
6. Prykhodko, N.V. Constructing the non-linear regression models on the basis of multivariate normalizing transformations / N.V. Prykhodko, S.B. Prykhodko // Electronic modeling, Vol. 40, No. 6, 2018. – P. 101–110. DOI: 10.15407/emodel.40.06.101
7. Arnuphaptrairong, T. An Empirical Validation of Mobile Application Effort Estimation Models / T. Arnuphaptrairong, W. Suksawasd // Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2017). (March 15–17, 2017, Hong Kong, China). Hong Kong, 2017. – P. 697–701.
8. Foss, T. A simulation study of the model evaluation criterion MMRE / T. Foss, E. Stensrud, B. Kitchenham, I. Myrtveit // IEEE Transactions on software engineering, 11(29), 2003. – P. 985–995.
9. Port, D. Comparative studies of the model evaluation criterions MMRE and PRED in software cost estimation research / D. Port, M. Korte // Proceedings of the 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. – ACM, New York, 2008. – P. 51–60.
10. Olkin, I. Multivariate Analysis: Overview / I. Olkin, A. R. Sampson // International encyclopedia of social & behavioral sciences / N.J. Smelser, P.B. Baltes (eds.) 1st edn. – Elsevier, Pergamon, 2001. – P. 10240–10247.
11. Mardia, K.V. Measures of multivariate skewness and kurtosis with applications / K.V. Mardia // Biometrika, Vol. 57, 1970. – P. 519–530. – DOI: 10.1093/biomet/57.3.519