УДК 004.412:519.237.5
DOI https://doi.org/10.32838/2663-5941/2019.5-1/25

Приходько С.Б.
Національний університет кораблебудування імені адмірала Макарова

Приходько Н.В.
Національний університет кораблебудування імені адмірала Макарова

Книрік К.О.
Національний університет кораблебудування імені адмірала Макарова

ТРИФАКТОРНЕ НЕЛІНІЙНЕ РЕГРЕСІЙНЕ РІВНЯННЯ ДЛЯ ОЦІНЮВАННЯ ТРУДОМІСТКОСТІ РОЗРОБКИ МОБІЛЬНИХ ЗАСТОСУНКІВ У ФАЗІ ПЛАНУВАННЯ

Метою роботи є створення множинного нелінійного рівняння регресії для оцінювання трудомісткості розробки мобільних застосунків на основі багатовимірного нормалізуючого перетворення у фазі планування за змінними, що беруться, наприклад, з такого документу, як RAD (Requirement Analysis Document). Трифакторне нелінійне регресійне рівняння для оцінювання трудомісткості розробки мобільних застосунків у фазі планування побудоване на основі нормалізації чотиривимірного негаусівського набору даних (фактична трудомісткість розробки у людино-годинах, кількості екранів, функцій та файлів мобільного застосунку) із 30 застосунків. Нормалізація цього набору даних здійснена за допомогою чотиривимірного перетворення Джонсона для сімейства S_B. Використання чотиривимірного перетворення у порівнянні з одновимірними дозволяє врахувати кореляцію між змінними, що призводить до покращення нормалізації даних, яка пов'язана з виконанням статистичної гіпотези щодо відповідності їх розподілу чотиривимірному розподілу Гауса, з подальшим підвищенням достовірності відповідного оцінювання. Виконано порівняння побудованого нелінійного рівняння з лінійним регресійним рівнянням і нелінійними регресійними рівняннями на основі десяткового логарифму і одновимірного перетворення Джонсона. Порівняно з іншими регресійними рівняннями (як лінійними, так і нелінійними) побудоване нелінійне рівняння має більше значення множинного коефіцієнту детермінації, більший відсоток передбачення, менше значення середньої величини відносної похибки та менші ширини довірчого інтервалу нелінійної регресії. Цей результат може бути пояснений найкращою багатовимірною нормалізацією і тим, що немає підстав відкидати нульову гіпотезу про те, що чотиривимірний розподіл для нормалізованих даних, який нормалізується за допомогою чотиривимірного перетворення Джонсона для сімейства S_B, є таким самим, як і чотиривимірний нормальний розподіл.

Ключові слова: *нелінійне регресійне рівняння, довірчий інтервал, інтервал передбачення, оцінювання трудомісткості, мобільний застосунок, нормалізуюче перетворення, негаусівські дані.*

Постановка проблеми. Оцінювання трудомісткості розробки застосунків є однією з важливих задач у фазі планування, яка є першою з п'яти етапів життєвого циклу розробки програмного забезпечення (далі – ПЗ) [1]. Сьогодні однією з найбільш відомих моделей для оцінювання трудомісткості є COCOMO II (COststructive COst MOdel) [2]. COCOMO II – це нелінійне рівняння регресії з параметрами, які отримані за даними метрик проєктів ПЗ. Одним із основних факторів для зазначеного рівняння є розмір ПЗ, який у фазі планування ще невідомий. Крім того, це рівняння побудоване на основі одновимірного нормалізуючого перетворення у формі десяткового логарифму, а таке перетворення не завжди дозволяє добре виконати нормалізацію даних. Все це призводить до низької достовірності оцінювання трудомісткості розробки мобільних застосунків у фазі планування.

Аналіз останніх досліджень і публікацій. Хоча розробка мобільних застосунків схожа на розробку вебзастосунків і має свої корені в більш традиційній розробці ПЗ, однак однією з критичних різниць є те, що відповідні програми часто написані спеціально для того, щоб можна було скористатися унікальними функціями, які пропонує конкретний мобільний пристрій [3]. Тому протягом останнього десятиріччя було опубліковано багато різних моделей прогнозування трудомісткості, включаючи як лінійні, так і нелінійні регресійні рівняння [4; 5].

Як відомо, є чотири основні припущення, які обґрунтовують використання рівнянь лінійної регресії, одним з яких є нормальність розподілу похибок [6]. Проте це припущення справедливе лише в окремих випадках. Отже, для прогнозування трудомісткості розробки мобільних застосунків як вибіркового середнього числа залежної випадкової величини виникає необхідність розробки відповідних нелінійних регресійних рівнянь. Застосування одновимірних нормалізуючих перетворень для побудови рівнянь і моделей нелінійної регресії не завжди призводить до задовільних результатів прогнозування (зокрема, за такими стандартними оцінками, як середня величина відносної похибки, ширина довірчого інтервалу та інтервалу передбачення [7; 8]). Це призводить до необхідності використання багатовимірних нормалізуючих перетворень.

Постановка завдання. Метою статті є побудова трифакторного рівняння нелінійної регресії та рівнянь нижньої і верхньої границь її довірчих інтервалів для оцінювання трудомісткості розробки мобільних застосунків у фазі планування залежно від кількості екранів, функцій та файлів мобільного застосунку на основі чотиривимірного нормалізуючого перетворення, що дозволить підвищити достовірність оцінювання вибіркового середнього числа залежної змінної нелінійної регресії порівняно з використанням одновимірних нормалізуючих перетворень.

Виклад основного матеріалу дослідження. Для досягнення мети статті, що сформульована вище, ми скористалися методами, наведеними в джерелі [8]. Згідно з цим джерелом спочатку виконується нормалізація багатовимірних негаусових даних за багатовимірним нормалізуючим перетворенням. У якості таких даних використані дані, що наведені у табл. 1. Ці дані отримані шляхом доповнення чотиривимірного набору даних про фактичну трудомісткість розробки Y у людино-годинах, кількості екранів X_1, функцій X_2 і файлів X_3 для 17 застосунків з джерела [9] даними для 21 мобільного застосунку (рядки з 18 до 38).

Для виявлення викидів у даних табл. 1 використовувався метод, заснований на багатовимірних нормалізуючих перетвореннях і квадраті відстані Махаланобіса MD^2 [10]. Було визначено, що немає викидів в даних табл. 1 для рівня значимості 0,005 та багатовимірного перетворення Джонсона для сімейства S_B, адже всі значення MD^2 менші, ніж величина квантиля розподілу χ^2, яка становить 14,86 для рівня значимості 0,005. Зазначимо, що оцінки параметрів багатовимірного перетворення Джонсона для сімейства S_B були знайдені за методом максимальної правдоподібності: $\hat{\gamma}_Y = 5{,}69898$, $\hat{\gamma}_1 = 0{,}524119$, $\hat{\gamma}_2 = 0{,}776179$, $\hat{\gamma}_3 = 0{,}540973$, $\hat{\eta}_Y = 2{,}40219$, $\hat{\eta}_1 = 0{,}743879$, $\hat{\eta}_2 = 0{,}79545$, $\hat{\eta}_3 = 0{,}534447$, $\hat{\varphi}_Y = -114{,}5452$, $\hat{\varphi}_2 = 1{,}6885$, $\hat{\varphi}_2 = 1{,}6885$, $\hat{\varphi}_3 = 0{,}90$, $\hat{\lambda}_Y = 3328{,}564$, $\hat{\lambda}_1 = 12{,}3743$, $\hat{\lambda}_2 = 12{,}091$ і $\hat{\lambda}_3 = 8{,}30648$. Вибіркова коваріаційна матриця $\mathbf{S}_N$ виглядає так:

$$\mathbf{S}_N = \begin{pmatrix} 1{,}0000 & 0{,}5974 & 0{,}4740 & 0{,}6625 \\ 0{,}5974 & 1{,}0000 & 0{,}8719 & 0{,}7008 \\ 0{,}4740 & 0{,}8719 & 1{,}0000 & 0{,}8166 \\ 0{,}6625 & 0{,}7008 & 0{,}8166 & 1{,}0000 \end{pmatrix}.$$

Далі для нормалізованих за багатовимірним перетворенням Джонсона сімейства S_B даних було побудовано таке лінійне регресійне рівняння:

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

де параметри рівняння (1) оцінювалися методом найменших квадратів, а їх оцінки є такими:

$$\hat{b}_0 = 0;\ \hat{b}_1 = 0{,}808152;\ \hat{b}_2 = -0{,}928296;\ \hat{b}_3 = 0{,}854262$$

Сума квадратів відхилень для рівняння (1) склала 14,867.

Потім було побудовано таке нелінійне регресійне рівняння:

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

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

Якість побудованого рівняння (2) була перевірена за множинним коефіцієнтом детермінації R^2, середньою величиною відносної помилки MMRE і відсотком прогнозування PRED(0,25), які використовуються як стандартні оцінки результатів прогнозування за допомогою регресійних моделей. Значення R^2, MMRE і PRED(0,25) дорівнюють 0,579, 0,493 і 0,526 відповідно і вказують на незадовільну якість рівняння (2) з оцінками параметрів, що були отримані згідно з даними з табл. 1 з 38 мобільних застосунків. З огляду на це для побудови нелінійного регресійного рівняння для оцінювання трудомісткості розробки мобільних застосунків у фазі планування далі було застосовано метод покращення нелінійних регресійних моделей на основі багатовимірних нормалізуючих перетворень із застосуванням квадрата відстані

Таблиця 1

Дані та межі інтервалів передбачення нелінійних регресій

№	Y	X_1	X_2	X_3	MD^2	Межі інтервалів передбачення нелінійних регресій					
						LB_1	UB_1	LB_2	UB_2	LB_5	UB_5
1	192	5	4	3	0,4	60,47	377,26	91,85	289,05	123,26	233,66
2	**272**	**5**	**4**	**3**	1,6	60,47	377,26	91,85	289,05	-	-
3	288	3	2	2	12,1	88,63	524,10	153,04	362,13	189,08	322,15
4	116	6	6	4	1,0	51,09	352,88	75,70	268,17	104,97	209,60
5	**372**	**5**	**5**	**4**	5,6	54,51	362,37	-	-	-	-
6	**504**	**9**	**8**	**6**	7,5	90,07	453,34	-	-	-	-
7	28	6	7	2	5,9	-0,71	232,51	20,09	160,96	27,83	86,22
8	**176**	**6**	**7**	**3**	3,0	18,85	277,78	37,91	204,55	-	-
9	364	10	11	9	7,3	157,35	665,23	267,93	404,24	328,16	417,98
10	120	10	10	5	3,9	48,66	363,80	73,41	272,31	97,21	204,26
11	**22**	**6**	**5**	**4**	11,9	70,82	402,23	-	-	-	-
12	224	11	6	2	6,2	73,52	447,01	111,94	322,06	124,28	243,22
13	24	2	2	1	5,7	-23,85	170,88	8,80	123,53	17,41	63,80
14	200	11	7	4	4,0	106,51	511,55	155,93	351,04	189,05	309,67
15	**160**	**6**	**6**	**7**	3,8	100,58	490,02	148,29	343,97	-	-
16	**120**	**2**	**2**	**1**	7,6	-23,85	170,88	8,80	123,53	-	-
17	**96**	**4**	**4**	**1**	9,5	-33,37	149,22	1,90	94,96	-	-
18	202	6	5	4	0,2	70,82	402,23	103,42	301,71	140,78	254,61
19	145	4	3	2	1,3	49,16	353,26	80,77	277,21	105,52	211,99
20	198	6	5	4	0,2	70,82	402,23	103,42	301,71	140,78	254,61
21	146	4	3	2	1,3	49,16	353,26	80,77	277,21	105,52	211,99
22	191	6	6	5	0,5	66,20	392,46	96,35	294,74	137,55	251,39
23	99	3	3	2	1,6	24,68	289,97	51,04	229,36	72,16	162,36
24	382	11	12	9	7,7	140,13	624,93	257,61	400,94	314,54	409,78
25	270	9	10	8	3,2	93,37	477,23	138,38	338,81	203,52	324,21
26	282	12	7	3	7,0	104,63	532,46	163,82	362,10	188,85	315,26
27	213	10	5	2	5,1	78,45	452,66	117,16	324,35	133,62	252,65
28	322	11	7	5	3,3	126,75	560,32	184,64	367,37	230,18	345,95
29	290	10	6	4	2,7	109,14	513,13	157,51	350,67	195,02	314,25
30	223	7	7	6	0,8	78,58	425,22	112,51	312,36	161,26	278,76
31	241	5	5	6	2,1	84,92	449,29	127,21	327,39	188,94	309,00
32	87	5	5	2	1,1	17,06	267,28	37,56	200,04	49,85	124,40
33	36	3	3	1	4,8	-29,04	153,63	5,31	105,65	15,99	58,72
34	216	8	7	5	0,4	77,05	418,56	108,93	307,94	146,95	262,02
35	67	5	6	2	3,0	1,39	233,20	22,41	164,98	31,85	92,77
36	115	7	7	3	1,5	30,98	306,43	50,32	228,90	65,69	152,77
37	36	2	2	1	5,4	-23,85	170,88	8,80	123,53	17,41	63,80
38	98	3	3	2	1,6	24,68	289,97	51,04	229,36	72,16	162,36

Махаланобіса та інтервалів передбачення [11]. Розповімо про суть цього методу [11]. На першому етапі, як це зазвичай робиться, початкові негаусівські дані перевіряються на наявність викидів. Якщо останні знайдено, то такі дані відкидаються. Для цього пропонується використовувати квадрат відстані Махаланобіса для нормалізованих даних. На першому етапі рівень значущості дорівнює 0,005. На другому етапі будується нелінійна регресійна модель із застосуванням відповідного методу на основі багатовимірних нормалізуючих перетворень [8]. Після цього на третьому етапі для рівня значущості, що дорівнює 0,05, визначаються межі інтервалу передбачення нелінійної регресії за методом, наведеним у джерелі [8]. На четвертому (завершальному) етапі перевіряють, чи є серед даних, за якими будувалася нелінійна регресійна модель, такі, що виходять за визначені межі інтервалу передбачення. Якщо останні знайдено, то вони відкидаються. Потім знову повторюються всі етапи для нових даних. Якщо таких викидів не було, то повторення етапів завершу-

ється, відповідна нелінійна регресійна модель побудована.

Для нелінійного регресійного рівняння (2) з оцінками параметрів, що були отримані згідно з даними з табл. 1 з 38 мобільних застосунків, виявилося, що значення Y для трьох застосунків (5, 6 і 11) виходять за визначені межі інтервалу передбачення. У табл. 1 ліва межа інтервалу передбачення, що отриманий на першій ітерації [8], позначена як LB_1, а права – як UB_1. Усього було 5 таких ітерацій, після яких залишилося 30 мобільних застосунків (1, 3, 4, 7, 9, 10, 12–14, 18–38). У табл. 1 ліва межа інтервалу передбачення, що отриманий на п'ятій ітерації [8], позначена як LB_5, а права – як UB_5.

На п'ятій ітерації для даних з 30 мобільних застосунків оцінки параметрів багатовимірного перетворення Джонсона для сімейства S_B такі: $\hat{\gamma}_Y = 0{,}58590$; $\hat{\gamma}_1 = 0{,}316749$; $\hat{\gamma}_2 = 0{,}86299$; $\hat{\gamma}_3 = 0{,}48606$; $\hat{\eta}_Y = 1{,}01714$; $\hat{\eta}_1 = 0{,}63606$; $\hat{\eta}_2 = 0{,}86557$; $\hat{\eta}_3 = 0{,}612856$; $\hat{\varphi}_2 = 1{,}5560$; $\hat{\varphi}_2 = 1{,}5560$; $\hat{\varphi}_2 = 1{,}5560$; $\hat{\varphi}_3 = 0{,}73913$; $\hat{\lambda}_Y = 500{,}266$; $\hat{\lambda}_1 = 11{,}3796$; $\hat{\lambda}_2 = 13{,}2488$; $\hat{\lambda}_3 = 8{,}52637$.

Вибіркова коваріаційна матриця $\mathbf{S}_N$ виглядає так:

$$\mathbf{S}_N = \begin{pmatrix} 1{,}0000 & 0{,}7067 & 0{,}5491 & 0{,}7808 \\ 0{,}7067 & 1{,}0000 & 0{,}8516 & 0{,}6374 \\ 0{,}5491 & 0{,}8516 & 1{,}0000 & 0{,}8161 \\ 0{,}7808 & 0{,}6374 & 0{,}8161 & 1{,}0000 \end{pmatrix}$$

На п'ятій ітерації для даних з 30 мобільних застосунків оцінки параметрів рівняння (1) є такими: $\hat{b}_0 = 0$; $\hat{b}_1 = 0{,}91484$; $\hat{b}_2 = -1{,}17495$; $\hat{b}_3 = 1{,}16023$. Сума квадратів відхилень для рівняння (1) у цьому випадку склала 1,391, що більш ніж у 10 разів менше за відповідну суму на першій ітерації.

Значення R^2, MMRE і PRED (0,25) дорівнюють 0,953, 0,133 і 0,867 відповідно і вказують на гарну якість рівняння (2) з оцінками параметрів, що були отримані згідно даними з табл. 1 з 30 мобільних застосунків.

Також за даними з табл. 1 з 30 мобільних застосунків були побудовані такі рівняння: лінійне регресійне рівняння і нелінійні регресійні рівняння за одновимірними перетвореннями у вигляді десяткового логарифму (Log10) та одновимірним перетворенням Джонсона для сімейства S_B.

Лінійне регресійне рівняння таке:

$$\hat{Y} = 40{,}250 + 28{,}973X_1 - 41{,}798X_2 + 50{,}665X_3. \quad (3)$$

Нелінійне регресійне рівняння за одновимірним перетворенням у вигляді десяткового логарифму є таким:

$$\hat{Y} = 10^{\hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}, \quad (4)$$

де $\hat{b}_0 = 1{,}73898$; $\hat{b}_1 = 1{,}6687$; $\hat{b}_2 = -2{,}1116$; $\hat{b}_3 = 1{,}30125$.

Нелінійне регресійне рівняння за одновимірним перетворенням Джонсона для сімейства S_B має вигляд (2), але з такими параметрами: $\hat{\gamma}_Y = 0{,}25204$; $\hat{\gamma}_1 = 0{,}10255$; $\hat{\gamma}_2 = 0{,}49345$; $\hat{\gamma}_3 = 0{,}61963$; $\hat{\eta}_Y = 0{,}58192$; $\hat{\eta}_2 = 0{,}63352$; $\hat{\eta}_2 = 0{,}63352$; $\hat{\eta}_3 = 0{,}58967$; $\hat{\varphi}_Y = 19{,}9286$; $\hat{\varphi}_1 = 1{,}90$; $\hat{\varphi}_2 = 1{,}81688$; $\hat{\varphi}_3 = 0{,}90$; $\hat{\lambda}_Y = 370{,}175$; $\hat{\lambda}_1 = 10{,}20$; $\hat{\lambda}_2 = 10{,}6468$; $\hat{\lambda}_3 = 8{,}6277$; $\hat{b}_0 = 0$; $\hat{b}_1 = 0{,}60292$; $\hat{b}_3 = 1{,}1148$; $\hat{b}_3 = 1{,}1148$. Значення R^2, MMRE і PRED (0,25), які дорівнюють 0,953, 0,133 і 0,867 відповідно, кращі для рівняння (2) з параметрами для багатовимірного перетворення Джонсона порівняно з одновимірним, для якого ці значення складають відповідно 0,878, 0,190 і 0,767, та з цими ж показниками для рівняння (3) – 0,838, 0,237 і 0,733, а також для рівняння (4) – 0,789, 0,206 і 0,733.

Були також визначені нижні (LB) і верхні (UB) межі довірчих інтервалів лінійної регресії та нелінійних регресій за методом, описаним у джерелі [8], на основі відповідно одновимірного і багатовимірного перетворень Джонсона та одновимірного перетворення у вигляді десяткового логарифму для рівня значущості 0,05. Ці межі наведені у таблиці 2. Для визначення нижньої і верхньої меж довірчих інтервалів побудоване відповідне рівняння:

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

де ψ_Y^{-1} – зворотне до нормалізуючого перетворення ψ_Y для залежної змінної Y; $\hat{Z}_Y$ – результат передбачення лінійного регресійного рівняння для значень компонентів вектора $\mathbf{z}_X = \{Z_1, Z_2, \ldots, Z_k\}$, що отриманий за нормалізованими даними; $\mathbf{Z}_X^+$ – матриця центрованих регресорів, яка містить значення $Z_{1_i} - \bar{Z}_1$, $Z_{2_i} - \bar{Z}_2$, ..., $Z_{k_i} - \bar{Z}_k$; $t_{\alpha/2,\nu}$ – квантиль t-розподілу Стьюдента залежно від рівня значимості $\alpha/2$ та кількості ступенів вільності ν; $S^2_{Z_Y} = \frac{1}{\nu}\sum_{i=1}^{N}\left(Z_{Y_i} - \hat{Z}_{Y_i}\right)^2$, $\nu = N - k - 1$; $\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+ - k \times k$ – матриця:

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

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

У рівнянні (5) знаки «–» та «+» відповідають рівнянням нижньої та верхньої меж довірчих інтервалів нелінійних регресій, що побудовані на

основі нормалізуючих перетварень (як одновимірних, так і багатовимірних).

Зверніть увагу, що ширина довірчого інтервалу нелінійної регресії на основі багатовимірного перетворення Джонсона менша, ніж після одновимірного перетворення Джонсона для 26 з 30 рядків даних (1, 3, 4, 7, 10, 12, 14, 18–32, 34–36, 38) та менша, ніж після одновимірного перетворення у вигляді десяткового логарифму для 25 рядків даних (1, 3, 4, 7, 10, 12, 14, 18–32, 34, 36, 38). Так само і порівняно з шириною довірчого інтервалу лінійної регресії для 30 рядків даних.

Кращі показники рівняння нелінійної регресії на основі чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B для оцінювання трудомісткості розробки мобільних застосунків у фазі планування можна пояснити кращою нормалізацією, яка перевіряється за відомими критеріями [12].

Так, якщо за критерієм на основі квадрата відстані Махаланобіса гіпотеза про нормальність багатовимірного закону розподілу нормалізованих за допомогою чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B даних для 30 застосунків з табл. 1 приймається для рівня значущості 0,005, то у разі застосування одновимірного перетворення та без нього – відкидається.

Висновки. Нами удосконалено трифакторне рівняння нелінійної регресії та рівняння нижньої і верхньої меж її довірчих інтервалів для оцінювання трудомісткості розробки мобільних застосунків у фазі планування залежно від кількості екранів,

Таблиця 2

Оцінки $\hat{Y}$ та межі довірчих інтервалів нелінійних регресій

№	Y	Одновимірне перетворення Log10			Одновимірне перетворення Джонсона			Чотиривимірне перетворення Джонсона		
		$\hat{Y}$	LB	UB	$\hat{Y}$	LB	UB	$\hat{Y}$	LB	UB
1	192	179,8	158,2	204,5	172,1	144,6	201,3	175,2	163,8	187,0
3	288	195,5	149,4	255,8	242,9	157,4	314,2	257,0	221,2	292,0
4	116	150,6	133,8	169,5	177,0	150,1	205,3	153,2	142,4	164,3
7	28	44,1	34,0	57,3	58,3	42,1	84,0	51,6	40,5	64,5
9	364	282,1	228,2	348,7	358,8	331,7	373,9	379,7	356,8	399,6
10	120	160,6	134,5	191,6	181,0	136,6	228,3	145,9	127,3	165,9
12	224	168,0	129,0	218,7	160,7	111,8	217,2	180,4	156,6	205,3
13	24	40,3	32,0	50,8	28,3	24,1	36,5	35,9	26,9	46,6
14	200	299,0	251,1	356,1	267,1	225,6	302,5	250,1	230,4	269,6
18	202	221,3	195,0	251,1	210,5	181,2	239,5	195,5	184,4	206,7
19	145	134,2	115,0	156,6	127,3	101,6	157,2	154,6	140,7	169,0
20	198	221,3	195,0	251,1	210,5	181,2	239,5	195,5	184,4	206,7
21	146	134,2	115,0	156,6	127,3	101,6	157,2	154,6	140,7	169,0
22	191	201,3	174,1	232,7	226,8	194,0	258,3	192,0	179,6	204,7
23	99	83,1	70,6	97,7	99,1	77,4	126,2	111,9	100,4	124,1
24	382	275,2	221,9	341,4	332,6	282,9	361,7	368,6	345,0	389,4
25	270	248,2	203,4	303,0	311,4	273,5	339,4	265,5	244,0	286,6
26	282	237,8	190,9	296,1	316,9	242,3	359,1	253,1	225,0	280,7
27	213	210,6	162,1	273,7	157,1	116,8	202,9	190,4	168,5	213,0
28	322	399,7	333,1	479,8	306,9	271,5	334,1	291,1	271,8	309,7
29	290	353,2	293,1	425,5	261,1	225,9	292,3	255,7	237,9	273,3
30	223	238,4	203,6	279,1	262,5	228,3	292,9	219,0	204,8	233,3
31	241	276,7	222,3	344,4	283,1	238,9	318,6	249,7	231,0	268,2
32	87	66,2	56,5	77,7	77,6	61,0	99,3	81,6	71,9	92,1
33	36	33,7	26,9	42,3	25,7	22,6	32,2	33,1	25,9	41,4
34	216	235,0	207,5	266,0	234,9	206,1	262,3	202,6	190,7	214,7
35	67	45,1	35,7	56,9	62,0	46,4	85,2	56,9	46,6	68,7
36	115	96,7	82,7	113,1	115,5	90,3	146,1	103,7	91,3	117,0
37	36	40,3	32,0	50,8	28,3	24,1	36,5	35,9	26,9	46,6
38	98	83,1	70,6	97,7	99,1	77,4	126,2	111,9	100,4	124,1

функцій та файлів мобільного застосунку на основі чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B, що дозволяє підвищити достовірність оцінювання вибіркового середнього числа залежної змінної нелінійної регресії порівняно з використанням одновимірних нормалізуючих перетворень. Порівняно з іншими регресійними рівняннями (як лінійними, так і нелінійними) побудоване рівняння має більші значення множинного коефіцієнту детермінації і відсотку передбачень та менші середні величини відносної похибки та ширини довірчого інтервалу нелінійної регресії. На прикладі удосконалення трифакторного нелінійного регресійного рівняння підтверджено працездатність методу покращення нелінійних регресійних моделей на основі багатовимірних нормалізуючих перетворень із застосуванням квадрата відстані Махаланобіса та інтервалів передбачення. Надалі планується використання інших наборів даних та нормалізуючих перетворень для побудови нелінійного регресійного рівняння для оцінювання трудомісткості розробки мобільних застосунків.

Список літератури:

1. Zhu H. Software Design Methodology: From Principles to Architectural Styles. 1st Edition. Butterworth-Heinemann : Elsevier, 2005. 368 p.

2. Boehm B.W., Abts C., Brown A.W., Chulani S., Clark B.K., Horowitz E., Madachy R., Reifer D.J., Steece B. Software Cost Estimation with COCOMO II. Upper Saddle River. NJ : Prentice Hall PTR, 2000. 544 p.

3. Rouse M. Mobile application development. URL: https://searchmicroservices.techtarget.com/definition/mobile-application-development (дата звернення: 19.07.2019).

4. Francese R., Gravino C., Risi M., Scanniello G., Tortora G. On the use of requirements measures to predict software project and product measures in the context of Android mobile apps: A preliminary study. *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.

5. Shahwaiz S.A., Malik A.A., Sabahat N.A parametric effort estimation model for mobile apps. *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.

6. Ryan T.P. Modern regression methods. 2nd Edition. New York : John Wiley & Sons, 2008. 672 p.

7. Prykhodko S., Prykhodko N., Makarova L. Building the Non-Linear Regression Equations on the Basis of Multivariate Normalizing Transformations. *Proceedings of the 2018 IEEE First International Conference on SYSTEM ANALYSIS & INTELLIGENT COMPUTING (SAIC).* (08–12 October, 2018, Kyiv, Ukraine). Kyiv, 2018. P. 48–52. DOI: 10.1109/SAIC.2018.8516742.

8. Prykhodko N.V., Prykhodko S.B. Constructing the non-linear regression models on the basis of multivariate normalizing transformations. Electronic modeling. 2018. Vol. 40. No. 6. P. 101–110. DOI: 10.15407/emodel.40.06.101.

9. Arnuphaptrairong T., Suksawasd W. An Empirical Validation of Mobile Application Effort Estimation Models. *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.

10. Prykhodko S., Prykhodko N., Makarova L., Pugachenko K. Detecting Outliers in Multivariate Non-Gaussian Data on the basis of Normalizing Transformations. *Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) «Celebrating 25 Years of IEEE Ukraine Section».* (May 29 – June 2, 2017, Kyiv, Ukraine). Kyiv, 2017. P. 846–849. DOI: 10.1109/UKRCON.2017.8100366

11. Приходько С.Б., Приходько Н.В. Метод покращення нелінійних регресійних моделей на основі багатовимірних нормалізуючих перетворень. *Прикладні науково-технічні дослідження* : матеріали III міжнар. наук.-практ. конф. м. Івано-Франківськ, 3–5 квітня 2019 р. Івано-Франківськ : Сімфоня Форте, 2019. С. 20.

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

Prykhodko S.B., Prykhodko N.V., Knyrik K.O. THREE-FACTOR NON-LINEAR REGRESSION EQUATION TO ESTIMATE THE EFFORTS OF DEVELOPMENT OF MOBILE APPLICATIONS IN A PLANNING PHASE

The goal of the work is the creation of the multiple non-linear regression equation for estimating the efforts of developing mobile applications on the basis of the multivariate normalizing transformation. A three-factor non-linear regression equation to estimate the efforts (in man-hours) of developing mobile applications is constructed on the basis of the Johnson fourvariate normalizing transformation for S_B family. This equation is built around the Requirement Analysis Document (RAD) variables (number of screens, number of functions

and number of files). Comparison of the constructed equation with the linear model and non-linear regression models based on the decimal logarithm and the Johnson univariate transformation has been performed. In comparison with other linear regression equations and non-linear regression equations based on the univariate normalizing transformations, constructed equation has a larger multiple coefficient of determination, a smaller value of the mean magnitude of relative error, a larger value of percentage of prediction and smaller widths of the confidence and prediction intervals of non-linear regression. This may be explained best multivariate normalization and the fact that there is no reason to reject the null hypothesis that the fourvariate distribution for normalized data, which normalized by the Johnson fourvariate transformation for S_B family, is the same as the fourvariate normal distribution. The practical significance of obtained results is that the software realizing the constructed equation is developed in the sci-language for Scilab. The experimental results allow to recommend the constructed model for use in practice. Prospects for further research may include the application of other multivariate normalizing transformations and data sets to construct the multiple non-linear regression equation for estimating the efforts of developing mobile applications.

Key words: *nonlinear regression equation, confidence interval, prediction interval, effort estimation, mobile application, normalizing transformation, non-Gaussian data.*