

Макарова Л.М.

Національний університет кораблебудування імені адмірала Макарова

Латанська Л.О.

Національний університет кораблебудування імені адмірала Макарова

Пухалевич А.В.

Національний університет кораблебудування імені адмірала Макарова

Приходько К.С.

Національний університет кораблебудування імені адмірала Макарова

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ ТРИВАЛОСТІ ВИКОНАННЯ РОБІТ З РОЗРОБКИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ З ВИКОРИСТАННЯМ МЕТОДОЛОГІЇ AGILE

У роботі побудовано однофакторну нелінійну регресійну модель для оцінювання тривалості виконання робіт із розробки програмного забезпечення з використанням методології Agile шляхом застосування одновимірного нормалізуючого перетворення Джонсона сім'ї S_B , що дозволило підвищити достовірність оцінювання тривалості виконання робіт у порівнянні з існуючими моделями. Використання нормалізуючого перетворення Джонсона сім'ї S_B дозволило зменшити довірчий інтервал та інтервал прогнозування нелінійного рівняння регресії в порівнянні з одновимірним нормалізуючим перетворенням на основі десятичного логарифма й лінійної моделі.

Проблема достовірного оцінювання необхідних трудових витрат на розробку є однією з ключових проблем під час управління проектами розробки програмного забезпечення. У статті розглянуті питання оцінювання тривалості виконання робіт із розробки програмного забезпечення з використанням гнучкої методології розробки Agile, яка останнім часом отримує все більшу популярність у розробників і має певні особливості, унаслідок чого важко отримати в достатньому обсязі необхідну інформацію для оцінювання тривалості розробки й використовувати найпоширеніші методи для оцінювання тривалості проєктів із розробки програмного забезпечення.

Гарні результати, які дає застосування нелінійних регресійних моделей з використанням методу нормалізуючих перетворень для рішення задач оцінювання тривалості й трудомісткості програмних проєктів, дозволили вжити зазначений підхід для побудови нелінійної регресійної моделі для оцінювання тривалості виконання робіт із розробки програмного забезпечення з використанням методології Agile.

Також було розроблено програмне забезпечення для прогнозування часу виконання робіт на основі побудованої нелінійної регресійної моделі, що дозволило скоротити час проведення відповідних розрахунків.

Ключові слова: регресійна модель, нелінійна регресія, нормалізуюче перетворення Джонсона, інтервал прогнозування, методологія Agile.

Постановка проблеми. Нині розробка програмного забезпечення (далі – ПЗ) ведеться з використанням різних мов програмування, платформ та методологій в умовах обмеження часу й ресурсів і потребує достовірного оцінювання необхідних трудових витрат на розробку, що є однією з ключових проблем при управлінні проєктами розробки ПЗ.

Існуючі на сьогодні моделі оцінювання трудомісткості виконання робіт використовують різні фактори, такі, як: кількість рядків коду, кількість

функціональних точок, кількість сторінок документації, досвід розробників та інші [1; 2]. На основі оцінки трудомісткості можна отримати оцінку безпосередньо тривалості виконання робіт.

Основними методами оцінювання тривалості виконання робіт з розробки ПЗ є експертні, навчально-орієнтовані, алгоритмічні, математичні, статистичні та змішані методи [3]. Найбільш використовувані методи, такі як COCOMO, COCOMO II, PERT та ISBSG, належать до алгоритмічних або математично-статистичних [4].

Все більшу популярність у розробників ПЗ отримує методологія Agile, яка має певні особливості, в наслідок чого важко отримати інформацію для оцінювання тривалості розробки на основі деяких з перелічених факторів в достатньому обсязі [5; 6]. Також не враховується час на виправлення знайдених помилок. Тому необхідність побудови нелінійної регресійної моделі для оцінювання тривалості виконання робіт з розробки ПЗ із використанням методології Agile є актуальною. Така модель дозволить підвищити достовірність оцінювання часу розробки ПЗ.

Аналіз останніх досліджень і публікацій. Методологія Agile відноситься до сімейства гнучкої методології розробки ПЗ. Її основою є ітеративна розробка. Ця методологія має певні особливості [7]:

- розробка базується на коротких циклах – ітераціях, які зазвичай тривають один-два тижні та мають вигляд програмного проекту в мініатюрі;
- неможливість одразу сформулювати вичерпний та актуальний набір варіантів використання через постійні зміни вимог до ПЗ;
- необхідність швидко й чітко визначити, скільки часу й ресурсів потрібно використати для розробки в умовах коротких та інтенсивних ітерацій, щоб мати можливість якомога скоріше надати бізнесу інформацію про стан розробки ПЗ.

Унаслідок цих особливостей методології Agile, використання найбільш поширених методів для оцінювання тривалості проектів із розробки ПЗ не завжди дозволяє достовірно оцінити трудомісткість і час розробки ПЗ.

Під час побудови регресійних моделей для оцінювання тривалості виконання робіт із розробки ПЗ виникає ряд труднощів:

- випадкові величини, що входять до регресійної моделі, здебільшого не підпорядковуються гаусівському закону розподілу, тому маємо нелінійну регресію;
- залежна випадкова величина залежить від двох і більше факторів одночасно, тому маємо множинну регресію.

Для побудови нелінійних регресійних моделей існує кілька методів:

- метод простого перебору полягає в завданні різних видів рівняння регресії та вибору найкращого наближення за певним критерієм. Недоліки методу – велика кількість обчислень і неможливість отримання найкращого рішення – роблять такий метод неефективним [8].
- метод лінеаризуючих перетворень полягає у переході від нелінійної регресії до лінійної завдяки заміні вихідних змінних і коефіцієнтів. Недоліки методу – не завжди можливо підібрати таку заміну [9–11] та спрощення регресійної моделі й втрата частини інформації, що пов'язана з нелінійністю.
- метод нормалізуючих перетворень полягає у пошуку перетворень, за допомогою яких можна здій-

снити перехід від вихідних негаусівських випадкових величин до гаусівських. Для отриманих гаусівських випадкових величин будують лінійне рівняння регресії, яке далі перетворюють на нелінійне рівняння за допомогою зворотних нормалізуючих перетворень [10].

Використання методу нормалізуючих перетворень для рішення задач оцінювання тривалості та трудомісткості програмних проектів дає гарні результати [12; 13]. Це дозволяє застосувати такий метод для побудови нелінійної регресійної моделі для оцінювання тривалості виконання робіт з розробки ПЗ з використанням методології Agile.

Постановка завдання. Метою роботи є побудова однофакторної нелінійної регресійної моделі для оцінювання тривалості виконання робіт із розробки ПЗ з використанням методології Agile шляхом застосування одновимірного нормалізуючого перетворення Джонсона, що дозволило підвищити достовірність оцінювання тривалості виконання робіт у порівнянні з існуючими моделями.

Виклад основного матеріалу дослідження. Регресійна модель у загальному вигляді може бути представлена таким рівнянням:

$$y = \bar{y} + \varepsilon_t = f(x) + \varepsilon_t, \quad (1)$$

де y – залежна змінна (результативна ознака); $f(x)$ – функція, що визначає вид регресійної моделі (лінійна або нелінійна); x – незалежна змінна (фактор); ε_t – випадкова помилка (збурення).

У разі гаусівського закону розподілу випадкових величин регресійна модель є лінійною і може бути представлена таким рівнянням:

$$z_y = b_1 z_x + b_0 + \varepsilon, \quad (2)$$

де b_1, b_0 – коефіцієнти лінійної регресії, які знаходяться методом найменших квадратів; ε – випадкова величина, що розподілена за гаусівським законом, $\varepsilon \sim N(0, 1)$.

Для оцінювання точності рівняння регресії намагаються знайти його довірчий інтервал. У разі гаусівського закону розподілу випадкових величин довірчий інтервал лінійного рівняння регресії може бути представлений таким рівнянням [14]:

$$y = \hat{y} \pm t_{(\alpha/2, n-2)} \cdot S \cdot \sqrt{1 + \frac{(x - \bar{x})^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (3)$$

де \hat{y} – розрахункове значення y за рівнянням регресії; $t_{(\alpha/2, n-2)}$ – квантіль t -розподілу Стюдента; α – рівень значущості; n – кількість випадкових величин;

$$S = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}.$$

Аналогічно може бути представлений таким рівнянням інтервал прогнозування лінійного рівняння регресії:

$$y = \hat{y} \pm t_{(\alpha/2, n-2)} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (4)$$

Як було показано в [12], розподіл емпіричних даних тривалості й трудомісткості розробки ПЗ відрізняється від гаусівського закону розподілу випадкових величин. Тому застосуємо метод нормалізуючих перетворень для побудови нелінійної регресійної моделі для оцінювання тривалості виконання робіт з розробки ПЗ із використанням методології Agile.

Як нормалізуюче перетворення можна використовувати перетворення на основі логарифма (десятькового або натурального), Бокса-Кокса, Джонсона. У роботі використовується чотирьох-параметричне нормалізуюче перетворення Джонсона, яке дозволяє врахувати особливості емпіричних даних, зокрема асиметрію та ексцес.

У загальному вигляді нормалізуюче перетворення Джонсона може бути представлено таким рівнянням [15]:

$$z = \gamma + \eta h(x, \phi, \lambda); \quad \eta > 0; -\infty < \gamma < \infty; \lambda > 0; -\infty < \phi < \infty; \quad (5)$$

де z – нормована нормально розподілена випадкова величина; $\gamma, \eta, \phi, \lambda$ – параметри перетворення Джонсона; x – випадкова величина, яка нормалізується; h – функція певної сім'ї Джонсона:

$$S_L \quad h_1(x, \phi, \lambda) = \ln(\tilde{x}), \quad x > \phi;$$

$$S_B \quad h_2(x, \phi, \lambda) = \ln\left(\frac{\tilde{x}}{1 - \tilde{x}}\right), \quad \phi < x < \phi + \lambda; \quad (6)$$

$$S_U \quad h_3(x, \phi, \lambda) = \text{Arsh}(\tilde{x}), \quad -\infty \leq x \leq +\infty;$$

$$\text{де } \tilde{x} = \frac{x - \phi}{\lambda}.$$

Перетворення, зворотне до (5), може бути представлено таким рівнянням:

$$x = \phi + \lambda h^{-1}(z, \gamma, \eta);$$

$$\eta > 0; -\infty < \gamma < \infty; \lambda > 0; -\infty < \phi < \infty; \quad (7)$$

де x – випадкова величина з розподілом Джонсона; $\gamma, \eta, \phi, \lambda$ – параметри перетворення Джонсона; z – нормально розподілена випадкова величина; h^{-1} – функції певної сім'ї Джонсона:

$$S_L \quad h_1^{-1}(z, \gamma, \eta) = e^z,$$

$$S_B \quad h_2^{-1}(z, \gamma, \eta) = \frac{1}{1 + e^{-z}}, \quad (8)$$

$$S_U \quad h_3^{-1}(z, \gamma, \eta) = \frac{e^z - e^{-z}}{2},$$

$$\text{де } \zeta = \frac{z - \gamma}{\eta}.$$

Для вибору конкретної сім'ї розподілу Джонсона скористаємося аналітичною залежністю ексцесу від квадрату асиметрії, яка наведена в [16].

Для перевірки якості побудованого рівняння регресії використаємо коефіцієнт детермінації R^2 , середню величину відносної похибки $MMRE$ та рівень прогнозування $Pred(0,25)$.

Для побудови нелінійної регресійної моделі були використані такі дані про тривалість виконання робіт з розробки ПЗ з використанням мето-

дології Agile: кількість $n=886$, випадкова величина X – трудомісткість (людино-годин), випадкова величина Y – тривалість (годин).

Перевірка емпіричних даних на викиди проводилася за допомогою квадрата відстані Махалано-біса, методика виконання якої наведена, зокрема, в [17], після чого скоригована вибірка склала $n=761$.

Ймовірнісні характеристики скоригованої вибірки емпіричних даних наведені в таблиці 1.

Таблиця 1

Характеристики вибірки емпіричних даних

Параметр	ВВ X	ВВ Y
\hat{m}	179,0394	21,3459
\hat{D}	207312,45	116,2418
$\hat{\sigma}$	455,3158	10,7816
\hat{A}	6,0339	0,9018
$\hat{\varepsilon}$	50,0488	2,8680

Сім'я розподілів Джонсона для випадкових величин X та Y була обрана у відповідності з [16] за значенням оцінок A^2 та ε – сім'я розподілів S_B .

Нормалізація випадкових величин X та Y була виконана у відповідності з (5) для сім'ї S_B Джонсона (6). Знайдені параметри перетворення Джонсона наведені в таблиці 2.

Таблиця 2

Параметри перетворення Джонсона

Параметр	ВВ X	ВВ Y
γ	3,1023	0,9954
η	0,5724	0,7311
ϕ	0,4538	7,9401
λ	10925,91	50,1021

У результаті нормалізації отримали нормовані нормально розподілені випадкові величини Z_X та Z_Y .

Лінійна регресійна модель для нормалізованих даних була побудована згідно з (2) та представлена таким рівнянням:

$$Z_y = 0,1213Z_x + 0,0087 + \varepsilon.$$

Нелінійна регресійна модель була побудована з використанням лінійної регресійної моделі та зворотного нормалізуючого перетворення (7) з відповідною сім'єю S_B Джонсона (8) та представлена таким рівнянням:

$$y = \frac{e^c(\lambda_y + \phi_y) + \phi_y}{1 + e^c},$$

$$\text{де } c = \frac{1}{\eta_y} \cdot \left(b_1 \left[\gamma_x + \eta_x \ln\left(\frac{x - \phi_x}{\lambda_x + \phi_x - x}\right) \right] + b_0 - \gamma_y + \varepsilon \right).$$

Остаточно нелінійна регресійна модель має такий вигляд:

$$y = \frac{e^c \cdot 6,5544 + 0,05}{1 + e^c},$$

де $c = 0,1352 \cdot \ln\left(\frac{x - 0,4538}{6,9582 - x}\right) - 3,6031 + \varepsilon$.

Порівняємо побудовану нелінійну регресійну модель на основі нормалізуючого перетворення Джонсона з такими моделями: нелінійна регресійна модель на основі десяткового логарифма й лінійна модель, побудована без використання нормалізації в припущенні про нормальність емпіричних даних, за допомогою коефіцієнту детермінації R^2 , середньої величини відносної похибки $MMRE$ та рівня прогнозування $Pred(0,25)$, які наведені в таблиці 3.

Таблиця 3

Оцінка якості регресійних моделей

	Нелінійна модель на основі нормалізуючого перетворення Джонсона	Нелінійна модель на основі десяткового логарифма	Лінійна модель
R_2	0,7460	0,4146	0,3779
$MMRE$	0,2063	0,3473	0,5108
$Pred(0,25)$	0,7698	0,5765	0,3088

Як видно із наведених даних, модель, побудована на основі нормалізуючого перетворення Джонсона, має кращі значення параметрів оцінки якості регресійних моделей: значення рівня прогнозування $Pred(0,25)$ перевищує 0,75, а середня величина відносної похибки $MMRE$ менша за 0,25. Однак значення коефіцієнту детермінації R^2 , хоча й більше, ніж 0,6, та менше за 0,8. Це свідчить про необхідність застосування двовимірного нормалізуючого перетворення Джонсона для врахування взаємного впливу випадкових величин X та Y .

95% довірчий інтервал нелінійного рівняння регресії було побудовано згідно з (3) з використанням 95% довірчого інтервалу лінійного рівняння регресії та зворотного нормалізуючого перетворення (7) з відповідною сім'єю S_B Джонсона (8):

$$y_1 = \frac{e^{k_1(\lambda_y + \phi_y)} + \phi_y}{1 + e^{k_1}},$$

де:

$$k_1 = \frac{1}{\eta_y} \cdot \left(b_1 \cdot z_x + b_0 - \gamma_y \pm t_{(\alpha/2, n-2)} \cdot S_{z_y} \cdot \sqrt{\frac{1}{n} + \frac{(z_x - \bar{z}_x)^2}{\sum_{i=1}^n (z_{xi} - \bar{z}_x)^2}} \right),$$

$$z_x = \gamma_x + \eta_x \ln\left(\frac{x - \phi_x}{\lambda_x + \phi_x - x}\right).$$

На рисунку 1 представлені рівняння регресії та відповідні довірчі інтервали, побудовані: на основі нормалізуючого перетворення Джонсона (червоний колір), на основі десяткового логарифма (чорний колір), без використання нормалізації (сірий колір). Останні два представлені для порівняння з нелінійною моделлю на основі нормалізуючого перетворення Джонсона. Представлено збільшений фрагмент рисунку, $Y \in [12; 24]$, в якому знаходяться рівняння регресії та довірчі інтервали.

95% інтервал прогнозування нелінійного рівняння регресії було побудовано згідно з (4) з використанням 95% інтервалу прогнозування лінійного рівняння регресії та зворотного нормалізуючого перетворення (7) з відповідною сім'єю S_B Джонсона (8) аналогічно 95% довірчому інтервалу.

На рисунку 2 представлені рівняння регресії та відповідні інтервали прогнозування, побудовані: на основі нормалізуючого перетворення Джонсона (червоний колір), на основі десяткового логарифма (чорний колір), без використання нормалізації (сірий колір). Останні два також представлені для порівняння з нелінійною моделлю на основі нормалізуючого перетворення Джонсона.

З рисунку 2 видно, що:

- після перевірки емпіричних даних на викиди й коригування вихідної вибірки, дані, що залишилися, не містять викидів (тобто знаходяться в межах інтервалу прогнозування);
- усі значення нижньої границі інтервалу прогнозування для лінійної моделі від'ємні, в той час коли нижні границі інтервалу прогнозування для обох нелінійних моделей – на основі нормалізуючого перетворення Джонсона й на основі десяткового логарифма – більші ніж нуль;
- довжини інтервалів прогнозування, отримані для лінійної моделі й моделі на основі десяткового логарифма, більші відповідних довжин, отриманих для моделі на основі нормалізуючого перетворення Джонсона.

Висновки. В результаті роботи було побудовано однофакторну нелінійну регресійну модель для оцінювання тривалості виконання робіт з розробки ПЗ з використанням методології Agile шляхом застосування одновимірного нормалізуючого перетворення Джонсона, що дозволило підвищити достовірність оцінювання тривалості виконання робіт у порівнянні з існуючими моделями.

Також було розроблено ПЗ для прогнозування часу виконання робіт на основі побудованої нелінійної регресійної моделі, що дозволило скоротити час проведення відповідних розрахунків.

В подальшому планується застосування двовимірного нормалізуючого перетворення Джонсона для побудови нелінійної регресійної моделі для оцінювання тривалості виконання робіт із розробки ПЗ із використанням методології Agile.

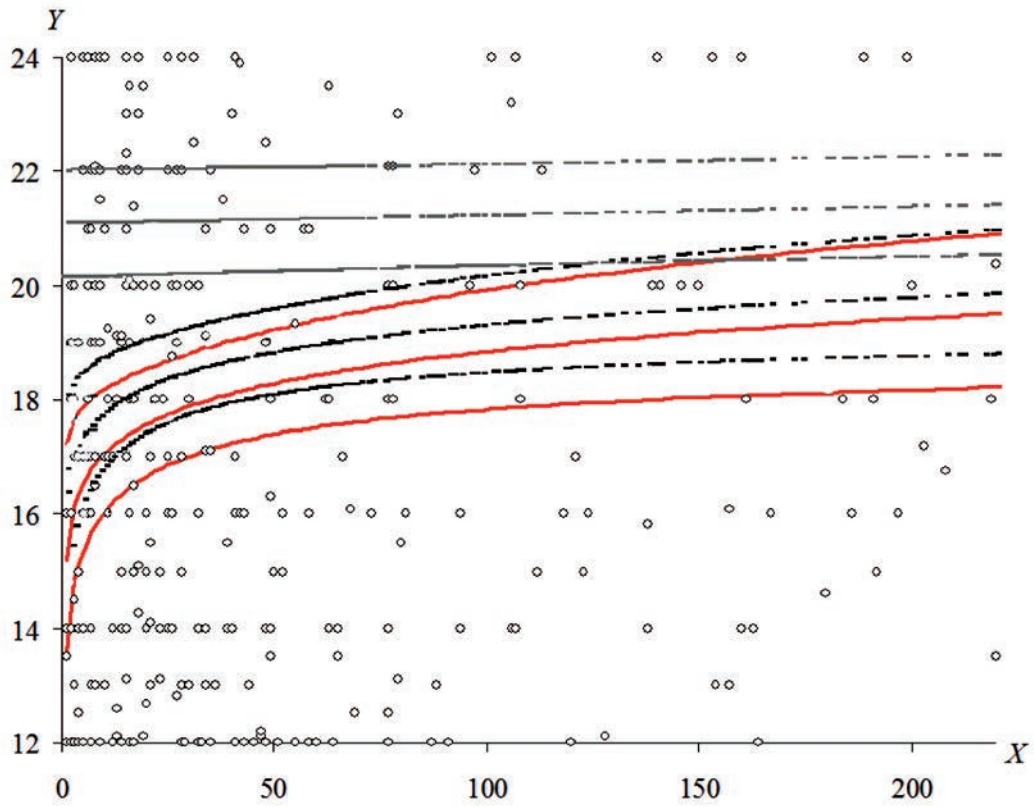


Рис. 1. Рівняння регресії та довірчі інтервали

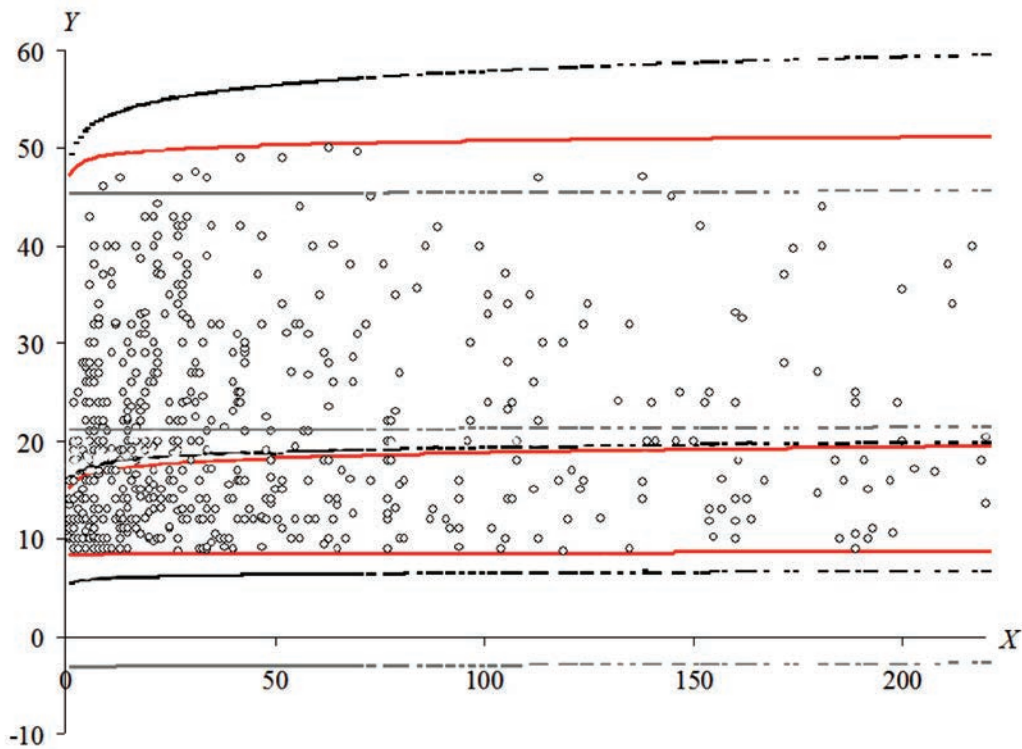


Рис. 2. Рівняння регресії та інтервали прогнозування

Список літератури:

1. Голованова М.А., Надин Е.В. Оценка трудоемкости работ на ранних стадиях создания программного обеспечения. *Системы обработки информации*. 2014. № 8 (124). С. 151–156.
2. Lazic L., Dokic I., Milinkovic S. Challenges in Estimating Software Testing Effort. *INFOTEH-JAHORINA : Zbornik radova XIII međunarodni naučno-stručni simpozijum, Jahorina, 19–21 mart 2014*. Sarajevo, 2014. P. 637–642.
3. Temnenco V. Software Estimation, Enterprise-Wide. URL: <http://www.ibm.com/developerworks/rational/library/jun07/temnenco/index.html> (дата звернення: 22.09.2020).
4. Bates D.M., Watts D.G. *Nonlinear Regression Analysis and Its Applications*. Wiley, 1988. 384 p.
5. Полезные метрики в Agile: Метрики производительности. URL: <http://it-study.by/philippov-agile-metrics> (дата звернення: 25.09.2020).
6. Agile Testing Process. URL: <http://www.slideshare.net/ExigenServices/agile-testing-process-26580032> (дата звернення: 25.09.2020).
7. Agile-манифест разработки программного обеспечения. URL: <http://agilemanifesto.org/iso/ru/manifesto.html> (дата звернення: 25.09.2020).
8. Дрейпер Н., Смит Г. Прикладной регрессионный анализ : В 2-х кн. Москва : Финансы и статистика, 1987. Кн. 2. 351 с.
9. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики : Учебник для вузов : В 2 т. Москва : ЮНИТИ-ДАНА, 2001. Т. 1 : Теория вероятностей и прикладная статистика. 656 с.
10. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. Москва : ФИЗМАТЛИТ, 2006. 816 с.
11. Chatterjee S., Simonoff J.S. *Handbook of Regression Analysis*. Wiley, 2012. 240 p.
12. Приходько С.Б., Пухалевич А.В. Інтервальне оцінювання статистичних моментів часу затримок виконання програмних проєктів на основі перетворення Джонсона. *Збірник наукових праць НУК*. 2010. № 2 (431). С. 118–124.
13. Prykhodko S., Prykhodko N., Makarova L. Estimating the Software Size of Open-Source PHP-Based Systems Using Non-Linear Regression Analysis. *Advanced Computer Information Technologies (ACIT-2018) : Proceedings of International Conference, Ceske Budejovice, June 1–3, 2018*. CEUR Workshop Proceedings, 2019. Vol. 2300. P. 199–202.
14. Xin Yan, Xiao Gang Su. *Linear regression analysis: theory and computing*. Singapore : World Scientific Publishing Co. Pte. Ltd., 2009. 328p.
15. Кендалл М., Стюарт А. Теория распределений / под ред. А.Н. Колмогорова. Москва : Наука. Гл. ред. физ.-мат. лит., 1966. 588с.
16. Приходько С.Б., Макарова Л.Н., Приходько А.С. Аналитическая зависимость для выбора семейства распределений Джонсона. *Проблеми інформаційних технологій*. 2016. № 02 (020). С. 105–110.
17. Prykhodko S., Prykhodko N., Makarova L., Pukhalevych A. Outlier Detection in Non-Linear Regression Analysis Based on the Normalizing Transformations. *Proceedings of the 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv–Slavske, February 25–29, 2020*. Lviv, 2020. P. 407–410. DOI: <https://doi.org/10.1109/TCSET49122.2020.235464>

Makarova L.M., Latanska L.O., Pukhalevych A.V., Prykhodko K.S. NON-LINEAR REGRESSION MODEL FOR DURATION ESTIMATING EXECUTION OF WORKS FOR SOFTWARE DEVELOPMENT USING AGILE METHODOLOGY

A univariate nonlinear regression model for estimating the duration of software development using Agile methodology based on the Johnson's univariate normalizing transform of the SB family was built in the paper. It allowed increased the reliability of estimating the duration of work compared to existing models. The use of the Johnson's normalizing transform of the SB family allowed to reduce the confidence interval and the prediction interval of the nonlinear regression equation in comparison with the univariate normalization transformation based on the decimal logarithm and the linear model.

The problem of reliably estimating the required labor costs for development is one of the key problems in managing software development projects. The article discussed the issues of estimating the duration of software development using a flexible development methodology Agile, which has recently become increasingly popular with developers and has certain features, making it difficult to obtain sufficient information to estimate the duration of development and use the most common methods for estimating the duration of software development projects.

The good results of nonlinear regression models using the method of normalizing transformations to solve problems of estimating the duration and effort of software projects, allowed to apply this approach to build a nonlinear regression model to estimate the duration of software development using Agile methodology.

Software to predict the time of execution of works on the basis of the built nonlinear regression model, which allowed to reduce the time of the corresponding calculations was also developed.

Key words: regression model, non-linear regression, Johnson's normalizing transformation, prediction interval, Agile methodology.