

УДК 005.8

С.Б. ПРИХОДЬКО, А.В. ПУХАЛЕВИЧ

*Національний університет кораблебудування ім. адмірала Макарова, Україна***РОЗРОБКА НЕЛІНІЙНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ ТРИВАЛОСТІ ПРОГРАМНИХ ПРОЕКТІВ НА ОСНОВІ НОРМАЛІЗУЮЧОГО ПЕРЕТВОРЕННЯ ДЖОНСОНА**

В роботі розроблено нелінійну регресійну модель тривалості програмних проектів в залежності від зусиль на їх виконання на основі нормалізуючого перетворення Джонсона із сім'ї S_U . Було виконано порівняння логарифмічного перетворення та перетворення Джонсона для нормалізації емпіричних даних тривалості робіт в програмних проектах та зусиль на виконання цих проектів за критерієм Пірсона. Показано, що в порівнянні з регресійними моделями COCOMO та ISBSG, побудованими на основі логарифмічного перетворення, розроблена регресійна модель є кращою за сумою квадратів відхилень між тривалістю за моделлю та емпіричними даними.

Ключові слова: програмний проект, розробка програмного забезпечення, нелінійна регресійна модель, тривалість програмних проектів, нормалізуюче перетворення Джонсона.

Вступ

На сьогоднішній день при оцінюванні тривалості проектів з розробки програмного забезпечення часто використовуються нелінійні регресійні моделі тривалості програмних проектів в залежності від зусиль на їх виконання, такі як COCOMO, ISBSG [1, 2]. Побудова вказаних регресійних моделей здійснювалася на основі нормалізації емпіричних даних за допомогою логарифмічного перетворення. Але логарифмічне перетворення не завжди дозволяє добре нормалізувати емпіричні дані, що приводе до необхідності використовувати інші перетворення.

Аналіз останніх досліджень показує, що існують інші нормалізуючі перетворення, які дозволяють краще виконати нормалізацію емпіричних даних. Одним з таких є перетворення Джонсона [3, 4]. Перетворення Джонсона залежить від чотирьох параметрів, що дозволяє будувати регресійні моделі для широкого спектру емпіричних даних за методом побудови нелінійних рівнянь регресії на основі нормалізуючих перетворень, який був запропонований в [5].

Тому метою роботи є побудова нелінійної регресійної моделі тривалості програмних проектів в залежності від зусиль на їх виконання на основі перетворення Джонсона.

Для побудови регресійної моделі було використано дані по 348 проектах з розробки програмного забезпечення з бази даних International Software Benchmarking Standards Group (ISBSG) [2] (рис. 1).

Нехай x – це емпіричні значення зусиль на виконання проектів; y – емпіричні значення тривало-

сті проектів. Розподіл величин x та y не є нормальним, асиметрія та ексцес значно відрізняються від нормального закону: $A_x = 4,87$, $\varepsilon_x = 34,66$, $A_y = 3,11$, $\varepsilon_y = 17,92$.

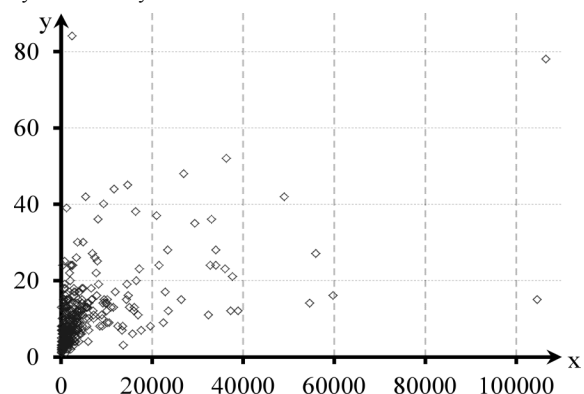


Рис. 1. Емпіричні значення тривалості робіт та затрачених зусиль в програмних проектах: x – емпіричні значення зусиль на виконання проектів, людино-годин; y – емпіричні значення тривалості проектів, місяців

Для побудови регресійної моделі використаємо нормалізуюче перетворення Джонсона, яке дозволяє здійснювати перехід до випадкової величини з нормальним розподілом.

Нормалізацію випадкової величини x за перетворенням Джонсона виконуємо за

$$z = \gamma + \eta q(x, \varphi, \lambda); \quad (1)$$

де q – функція, яка залежить від обраної сім'ї розподілу Джонсона;

γ , η , λ і φ – параметри перетворення, які знаходяться за емпіричним розподілом [3] або викорис-

товуючи непараметричний підхід [4], $\eta > 0$; $-\infty < \gamma < \infty$; $\lambda > 0$; $-\infty < \varphi < \infty$.

Сім'я перетворення Джонсона підбирається за значеннями оцінок асиметрії та ексцесу [3, 4].

Для величин x та y оцінки параметрів перетворення знаходимо в результаті рішення наступної задачі [4]:

$$\hat{\theta} = \arg \min_{\theta} \left\{ A_z^2 + (\varepsilon_z - 3)^2 + \bar{z}^2 + (S_z^2 - 1)^2 \right\},$$

де $\hat{\theta}$ – оцінка вектору невідомих параметрів;

θ – вектор невідомих параметрів, $\theta = \{\gamma, \eta, \lambda, \varphi\}$;

\bar{z} – вибіркове середнє гаусівської випадкової величини z , $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$;

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i;$$

S_z^2 – вибіркова дисперсія гаусівської випадкової

$$\text{величини } z, S_z^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2.$$

За значеннями оцінок асиметрії та ексцесу для нормалізації величини x було вибрано перетворення Джонсона сім'ї S_U :

$$z_x = \gamma_x + \eta_x \operatorname{Arsh} \left(\frac{x - \varphi_x}{\lambda_x} \right), \quad (2)$$

де z_x – гаусівська випадкова величина зусиль на виконання проектів, $-\infty \leq x \leq +\infty$;

$$\operatorname{Arsh} \left(\frac{x - \varphi_x}{\lambda_x} \right) = \ln \left[\frac{x - \varphi_x}{\lambda_x} + \sqrt{\left(\frac{x - \varphi_x}{\lambda_x} \right)^2 + 1} \right].$$

Значення параметрів перетворення Джонсона були такими: $\gamma_x = -6,496$, $\eta_x = 0,632$, $\lambda_x = 0,134$ і $\varphi_x = 81,721$. Використовуючи перетворення (2) з вказаними параметрами були отримані значення гаусівської випадкової величини z_x .

При нормалізації за перетворенням Джонсона значення критерію Пірсона $\chi^2 = 4,68$. При використанні для нормалізації логарифмічного перетворення, значення критерію Пірсона складає $\chi^2 = 11,69$. Критичне значення критерію Пірсона $\chi_{кр}^2 = 12,26$ ($\nu = 9 - 2 - 1 = 6$, $\alpha = 0,05$). Отримане значення критерію Пірсона менше ніж критичне для обох використаних перетворень.

За значеннями оцінок асиметрії та ексцесу для нормалізації величини y було вибрано перетворення Джонсона сім'ї S_B :

$$z_y = \gamma_y + \eta_y \ln \left(\frac{y - \varphi_y}{\lambda_y + \varphi_y - y} \right), \quad (3)$$

де z_y – гаусівська випадкова величина тривалості програмних проектів, $\varphi_y \leq y \leq \varphi_y + \lambda_y$.

Значення параметрів перетворення Джонсона були такими: $\gamma_y = 15,905$, $\eta_y = 1,237$, $\lambda_y = 3352506$ і $\varphi_y = -0,106$. Використовуючи перетворення (3) з вказаними параметрами були отримані значення величини z_y . Гістограма випадкової величини z_y і її теоретичний розподіл у вигляді щільності ймовірності $f(z)$ наведені на рис. 2.

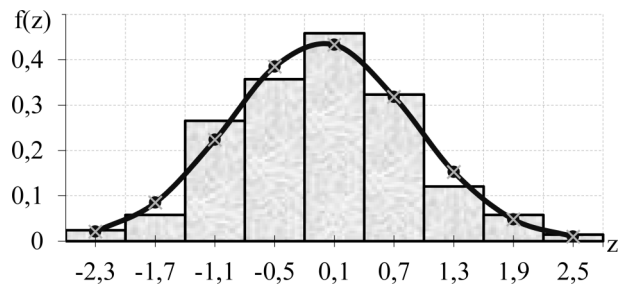


Рис. 2. Розподіл нормалізованих за перетворенням Джонсона значень тривалості програмних проектів: \square – емпіричний розподіл нормалізованої величини; \bullet – нормальний розподіл

Для нормалізованих за перетворенням Джонсона значень тривалості програмних проектів z_y значення критерію Пірсона $\chi^2 = 6,54$. При використанні для нормалізації логарифмічного перетворення, значення критерію Пірсона складає $\chi^2 = 14,05$, гістограма отриманої випадкової величини z_y і її теоретичний розподіл у вигляді щільності ймовірності $f(z)$ наведені на рис. 3.

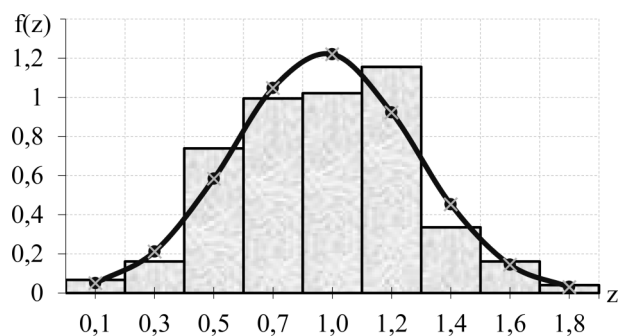


Рис. 3. Розподіл нормалізованої за логарифмічним перетворенням тривалості програмних проектів: \square – емпіричний розподіл нормалізованої величини; \bullet – нормальний розподіл

Критичне значення критерію Пірсона $\chi_{кр}^2 = 12,59$ ($\nu = 9 - 2 - 1 = 6$, $\alpha = 0,05$). При викорис-

танні перетворення Джонсона значення критерію Пірсона менше ніж критичне; при використанні логарифмічного перетворення значення критерію Пірсона більше ніж критичне, тобто використання логарифмічного перетворення не дозволяє досить добре виконати нормалізацію цих емпіричних даних.

Так як для величин z_x та z_y нормалізує перетворення Джонсона дає менше значення χ^2 , ніж логарифмічне перетворення, а при нормалізації величини y за логарифмічним перетворенням значення критерію Пірсона перевищує критичне значення $\chi^2 > \chi_{кр}^2$, то для нормалізації величин x та y краще використовувати нормалізує перетворення Джонсона.

Використовуючи перетворення (2) зі знайденими параметрами нормалізуємо емпіричні значення зусиль на виконання програмних проектів.

Після нормалізації виконуємо групування нормалізованих значень затрачених зусиль z_x за однаковими значеннями тривалості проекту, а потім відкидаємо групи, в які попало менше 7 значень затрачених зусиль. Для кожної з груп, що залишилися, знаходимо оцінку математичного сподівання зусиль на виконання.

Використовуючи отримані значення математичного сподівання будемо лінійну регресійну модель нормалізованих зусиль на виконання програмних проектів в залежності від значень їх тривалості:

$$z_x = ay + b, \quad (4)$$

де a та b – постійні коефіцієнти, які знаходяться за методом найменших квадратів. Побудована лінійна регресійна модель показана на рис. 4.

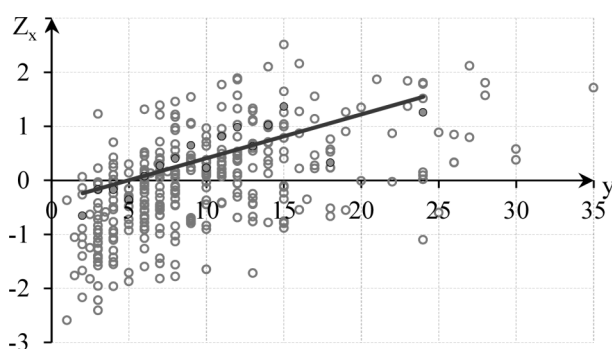


Рис. 4. Лінійна регресійна модель нормалізованих зусиль на виконання програмних проектів в залежності від їх тривалості: — — лінія регресії; ○ — емпіричні дані; ● — оцінка математичного сподівання в групі

Використовуючи лінійну регресійну модель (4) та нормалізує перетворення Джонсона (2), будемо нелінійну регресійну модель тривалості програ-

мних проектів в залежності від зусиль на їх виконання:

$$y = \frac{1}{a} \left[\gamma_x + \eta_x \operatorname{Arsh} \left(\frac{x - \varphi_x}{\lambda_x} \right) - b \right]. \quad (5)$$

Побудована нелінійна регресійна модель та існуючі регресійні моделі (COCOMO і ISBSG) показані на рис. 5.

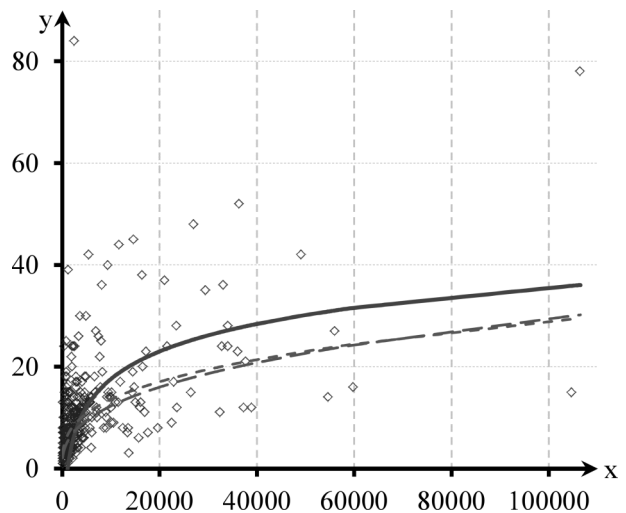


Рис. 5. Регресійна модель тривалості програмних проектів в залежності від зусиль на їх виконання: x – емпіричні значення зусиль на виконання проектів, людино-годин; y – емпіричні значення тривалості проектів, місяців; — — побудована регресійна модель; - - - - COCOMO; ····· ISBSG

Для порівняння регресійних моделей використовуємо такий критерій як сума квадратів відхилень (СКВ) між тривалістю за моделлю та емпіричними даними, $СКВ = \sum_{i=1}^n (y_i - y_x)^2$. Для розробленої регресійної моделі $СКВ = 25605$; для моделі ISBSG $СКВ = 26214$; для моделі COCOMO $СКВ = 28464$.

Розроблена регресійна модель краще описує емпіричні дані за вказаним критерієм.

Висновки

Отримала подальший розвиток нелінійна регресійна модель тривалості програмних проектів в залежності від зусиль на їх виконання за рахунок застосування нормалізує перетворення Джонсона із сім'ї S_U , що дозволяє краще описувати емпіричні дані.

В подальших дослідженнях планується будувати довірчі інтервали регресії тривалості про-

грамних проектів в залежності від зусиль на їх виконання.

Література

1. Boehm, B.W. *Software engineering economics* [Text] / B.W. Boehm. – Prentice-Hall, 1981. – 320 p.
2. *Developing Project Duration Models in Software Engineering* [Text] / P. Bourque, S. Oigny, A. Abran, B. Fourrier // *Journal of Computer Science and Technology*. – Boston, 2007. – № 3 (22). – P. 348-357.
3. Приходько, С.Б. *Інтервальне оцінювання параметрів стохастичних диференціальних систем на основі модифікації узагальненого методу моментів* [Текст] / С.Б. Приходько // *Матеріали XIII*

Міжнарод. конф. з автоматичного управління (Автоматика-2006), (Вінниця, 25 – 28 вересня 2006 р.) – Вінниця: УНІВЕРСУМ-Вінниця, 2007. – С. 69-75.

4. Приходько, С.Б. *Інтервальне оцінювання статистичних моментів негаусівських випадкових величин на основі нормалізуючих перетворень* [Текст] / С.Б. Приходько // *Математичне моделювання*. – 2011. – № 1 (24). – С. 9-13.

5. Приходько, С.Б. *Метод побудови нелінійних рівнянь регресії на основі нормалізуючих перетворень* [Текст] / С.Б. Приходько // *Тези доповідей міждерж. наук.-методич. конф. “Проблеми математичного моделювання” (Дніпродзержинськ, 13-15 червня 2012 р.)*. – Дніпродзержинськ: ДДТУ, 2012. – С. 31-33.

Надійшла до редакції 13.11.2012

Рецензент: д-р техн. наук, проф., заслужений діяч науки і техніки України, директор інституту комп'ютерних та інженерно-технологічних наук, зав. каф. інформаційних управляючих систем та технологій К.В. Кошкін, Національний університет кораблебудування ім. адмірала Макарова, Миколаїв, Україна.

РАЗРАБОТКА НЕЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ ДЛИТЕЛЬНОСТИ ПРОГРАММНЫХ ПРОЕКТОВ НА ОСНОВЕ НОРМАЛИЗИРУЮЩЕГО ПРЕОБРАЗОВАНИЯ ДЖОНСОНА

С.Б. Приходько, А.В. Пухалевич

В работе разработана нелинейная регрессионная модель длительности программных проектов в зависимости от усилий на их выполнение на основе нормализующего преобразования Джонсона из семи S_U . Было выполнено сравнение логарифмического преобразования и преобразования Джонсона для нормализации эмпирических данных длительности работ в программных проектах и усилий на выполнение этих проектов по критерию Пирсона. Показано, что в сравнении с регрессионными моделями COCOMO и ISBSG, построенными на основе логарифмического преобразования, разработанная регрессионная модель лучше по сумме квадратов отклонений между длительностью по модели и эмпирическими данными.

Ключевые слова: программный проект, разработка программного обеспечения, нелинейная регрессионная модель, длительность программных проектов, нормализующее преобразование Джонсона.

DEVELOPMENT OF NON-LINEAR REGRESSION MODEL OF SOFTWARE PROJECT DURATION ON THE BASIS OF JOHNSON NORMALIZING TRANSFORMATION

S.B. Prikhodko, A.V. Pukhalevich

Nonlinear regression model of software projects duration in the depending on effort based on Johnson normalizing transformation from S_U family was developed. Comparison of logarithmic transformation and Johnson transformation for normalization of empirical data of software projects duration and development effort using Pearson criteria was done. It was shown that in comparison with regression models COCOMO and ISBSG, built on the basis of the logarithmic transformation, new developed regression model is better (using sum of squared deviations between the model predicted duration and empirical data).

Key words: software project, software development, non-linear regression model, software project duration, Johnson normalizing transformation.

Приходько Сергій Борисович – канд. техн. наук, доц., зав. каф. програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна, e-mail: sprikhodko@nuos.edu.ua.

Пухалевич Андрій Володимирович – аспірант каф. програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна, e-mail: a.puhalevich@gmail.com.