

<http://dx.doi.org/10.15589/jnn20140312>

УДК 005.8

П 77

## DEVELOPING THE NONLINEAR REGRESSION MODELS OF THE SOFTWARE PROJECT DURATION BASED ON THE JOHNSON TRANSFORMATION

### РОЗРОБКА НЕЛІНІЙНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ ТРИВАЛОСТІ ПРОГРАМНИХ ПРОЕКТІВ НА ОСНОВІ ПЕРЕТВОРЕННЯ ДЖОНСОНА

Serhii B. Prykhodko

[sergiy.prykhodko@nuos.edu.ua](mailto:sergiy.prykhodko@nuos.edu.ua)

ORCID: 0000-0002-2325-018X

Andrii V. Pukhalevych

[andrii.pukhalevych@nuos.edu.ua](mailto:andrii.pukhalevych@nuos.edu.ua)

ORCID: 0000-0002-8827-3251

С. Б. Приходько,

д-р техн. наук, доц.;

А. В. Пухалевич,

старш. лаборант

National University of Shipbuilding, Mykolayiv

Національний університет кораблебудування, м. Миколаїв

**Abstract.** The regression models of software project duration based on project effort were considered. The aim of the research is developing of non-linear regression models of software project duration based on project effort using Johnson transformation for PC (personal computer) platform and MF (mainframe) platform. Empirical data from ISBSG dataset were analyzed using mathematical statistic methods. Software project duration non-linear regression models based on project effort were developed basing on Johnson transformation using regression analyze methods. Regression model developing approach based on Johnson transformation vs. approach based on common logarithm transformation was compared. Non-linear regression models of software project duration based on Johnson transformation have better characteristics than the models based on common logarithm transformation. Therefore, usage of regression models based on Johnson transformation for software project duration estimation was recommended.

**Keywords:** software project duration, time management, Johnson transform.

**Анотація.** На основі нормалізуючого перетворення Джонсона із сім'ї  $S_B$  побудовано нелінійні регресійні моделі тривалості програмних проєктів залежно від трудомісткості. Виконано порівняння побудованих моделей з моделями на основі логарифмічного перетворення. Рекомендовано використання регресійних моделей на основі перетворення Джонсона для оцінювання тривалості програмних проєктів.

**Ключові слова:** тривалість програмних проєктів, управління часом, перетворення Джонсона.

**Аннотация.** На основе нормализующего преобразования Джонсона из семейства  $S_B$  построены регрессионные модели длительности программных проектов в зависимости от трудоемкости. Выполнено сравнение построенных моделей с моделями на основе логарифмического преобразования. Рекомендовано использование моделей на основе преобразования Джонсона для оценивания длительности программных проектов.

**Ключевые слова:** длительность программных проектов, управление временем, преобразование Джонсона.

#### REFERENCES

- [1] Prykhodko S.B., Makarova L.N. *Analiticheskaya zavisimost dlya vybora raspredeleniya Dzhonsona semeystva SL* [Analytical dependence for selecting Johnson distribution of SL family]. *Vestnik KhNTU* [Bulletin of HNTU], 2012, issue 45, pp. 101–104.
- [2] Prykhodko S.B. *Intervalne otsiniuvannya statystychnykh momentiv nehausivskykh vypadkovykh velychyn na osnovi normalizuiuchykh peretvoren* [Interval estimation of statistical moments non-gaussian random variables based on normalizing transformations]. *Matematychni modeliuvannia – Mathematical modeling*, 2011, issue 25, pp. 9–13.
- [3] Prykhodko S.B., Pukhalevych A.V. *Rozrobka neliniinoi rehresiinoi modeli tryvalosti prohramnykh proektiv na osnovi normalizuiuchoho peretvorennya Dzhonsona* [Developing non-linear regression model the duration of software project duration based on Johnson transformation]. *Radioelektronni i kompiuterni systemy – Radio electronic and computer systems*, 2012, issue 56, pp. 90–93.
- [4] *CHAOS Report. The Standish Group International, Inc.*, 2009. Available at: [http://www.standishgroup.com/newsroom/chaos\\_2009.php](http://www.standishgroup.com/newsroom/chaos_2009.php).

- [5] Boehm B.W. *Software engineering economics*. Englewood Cliffs, NJ, Prentice Hall, 1981. 768 p.
- [6] Johnson R.A., Wichern D.W. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 2007. 800 p.
- [7] Oligny S., Bourque P., Fournier B. Exploring the relation between effort and duration in software engineering projects. Proc. of the *World Computer Congress*, Aug. 2000. – pp. 175–178.

### ПОСТАНОВКА ПРОБЛЕМИ

У галузі управління проектами зі створення програмного забезпечення існує проблема низької достовірності оцінювання тривалості програмних проектів [7]. Відомо, що в результаті неефективного управління 68 % програмних проектів не виконуються в заданий термін [4].

### АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У літературі для оцінювання тривалості програмних проектів пропонується використовувати параметричні моделі, які базуються на трудомісткості робіт. Серед цих моделей найбільш розповсюдженою є модель COCOMO [5] – нелінійна регресійна модель тривалості програмних проектів залежно від їх трудомісткості, яка була побудована для різних платформ: персональних комп'ютерів (PC), мейнфреймів (MF). У роботі [7] для цих платформ на основі емпіричних даних International Software Benchmarking Standards Group (ISBSG) було побудовано більш нові регресійні моделі.

Розподіл тривалості і трудомісткості програмних проектів не є нормальним [3], тому неможливо побудувати адекватну лінійну регресійну модель. Через це виникає необхідність побудови нелінійних регресійних моделей з використанням нормалізуючих перетворень. Моделі COCOMO та ISBSG були побудовані з використанням логарифмічного перетворення. Проте це перетворення не завжди дозволяє добре нормалізувати емпіричні дані [3], тому виникає необхідність використовувати інші

нормалізуючі перетворення. Одним з них є перетворення Джонсона [6]. У роботі [3] на основі емпіричних даних ISBSG було побудовано регресійну модель з використанням перетворення Джонсона, але без урахування платформи, для якої розроблюється програмне забезпечення. Хоча врахування платформи дозволяє більш точно оцінювати тривалість програмних проектів.

**МЕТА РОБОТИ** полягає у побудові нелінійних регресійних моделей тривалості програмних проектів залежно від їх трудомісткості з використанням перетворення Джонсона для платформ PC та MF.

### ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Побудова нелінійних регресійних моделей тривалості програмних проектів залежно від їх трудомісткості виконується в 3 етапи. Спочатку емпіричні дані нормалізуються з використанням нормалізуючого перетворення; далі за цими даними будується лінійна регресійна модель; на останньому етапі з лінійної моделі, застосовуючи перетворення, зворотно до нормалізуючого, будується нелінійна регресійна модель.

Для побудови регресійних моделей було використано емпіричні дані ISBSG. З поміж 789 проектів було відібрано проекти, які задовольняли наступні вимоги [3]: проекти не були віднесені до групи за низькою точністю даних; були відомі тривалість і трудомісткість проекту; відома платформа, для якої розробляється програмне забезпечення. Статистичні характеристики вибірок тривалості і трудомісткості для платформ PC та MF показані в табл. 1.

Таблиця 1. Статистичні характеристики вибірки ISBSG

Параметр	PC		MF	
	Тривалість, міс.	Трудомісткість, люд.-год	Тривалість, міс.	Трудомісткість, люд.-год
Кількість значень	52	52	215	215
Мінімальне значення	2	170	1	97
Максимальне значення	30	14520	84	104690
Середнє значення	9,25	2348,77	11,78	6418,31
Середньоквадратичне відхилення	5,58	3232,77	10,61	12441,63
Асиметрія	1,41	2,47	2,74	4,04
Експес	5,68	9,44	14,48	24,70
$\chi^2 / \chi_{кр}^2 (\alpha = 0,05)$	28,08 / 9,49	156,66 / 9,49	10 <sup>7</sup> / 12,59	10 <sup>10</sup> / 12,59

Нехай випадкова величина  $t$  – тривалість програмних проектів; випадкова величина  $e$  – трудомісткість. Гіпотези про нормальний закон розподілу величин  $t$  та  $e$  були відкинуті за критерієм

узгодженості Пірсона. Значення асиметрії та експесу величин  $t$  і  $e$  також значно відрізняються від відповідних значень для нормального закону розподілу.

Вибір сім'ї нормалізуючого перетворення Джонсона робиться по діаграмі Джонсона за значеннями асиметрії та ексцесу [1]. Для нормалізації емпіричних значень тривалості та трудомісткості програмних проєктів було вибрано перетворення Джонсона сім'ї  $S_B$  [6]:

$$z = \gamma + \eta \ln \left( \frac{x - \varphi}{\lambda + \varphi - x} \right), \quad (1)$$

де  $x$  – випадкова величина, розподіл якої відрізняється від нормального;  $z$  – нормалізована (гаусівська) випадкова величина;  $\gamma, \eta, \varphi, \lambda$  – параметри перетворення Джонсона,  $\varphi < \gamma < \varphi + \lambda; \eta > 0; \lambda > 0; -\infty < \varphi < \infty; -\infty < \gamma < \infty$ .

Оцінки параметрів  $\hat{\theta}$  перетворення (1) знаходимо шляхом розв'язання наступної задачі математичного програмування [2]:

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

Таблиця 2. Оцінки параметрів перетворення Джонсона

Платформа	Тривалість ( $t$ )				Трудомісткість ( $e$ )			
	$\gamma_t$	$\eta_t$	$\varphi_t$	$\lambda_t$	$\gamma_e$	$\eta_e$	$\varphi_e$	$\lambda_e$
PC	2,390	1,303	0,433	54,747	1,612	0,539	158,853	18248
MF	9,221	1,305	-0,255	10513	3,731	0,628	81,462	753187

Таблиця 3. Статистичні характеристики нормалізованих вибірок

Параметр	PC		MF	
	$z_t$	$z_e$	$z_t$	$z_e$
Кількість значень	49	49	215	215
Мінімальне значення	-2,20	-2,38	-2,57	-3,04
Максимальне значення	2,60	2,32	2,93	2,57
Середнє значення	0	0	0	0
Середньоквадратичне відхилення	1,01	1,01	1	1
Асиметрія	0	0	0	0
Ексцес	3,12	3,12	3,03	3,03
$\chi^2 / \chi_{кр}^2$ ( $\alpha = 0,05$ )	7,85 / 9,49	2,09 / 9,49	5,95 / 12,59	4,37 / 12,59

Гіпотези про нормальний закон розподілу величин  $z_t$  та  $z_e$  були підтвержені за критерієм узгодженості Пірсона. Значення асиметрії та ексцесу величин  $z_t$  та  $z_e$  приблизно дорівнюють відповідним значенням для нормального закону розподілу.

Після нормалізації було побудовано за нормалізованими випадковими величинами  $z_t$  та  $z_e$  лінійну регресійну модель:

$$z_t = a + bz_e, \quad (3)$$

Таблиця 4. Характеристики лінійних регресійних моделей

Параметр	PC		MF	
	Перетворення Джонсона	Логарифмічне перетворення	Перетворення Джонсона	Логарифмічне перетворення
Кількість значень	49	49	205	205
$R^2$	0,235	0,220	0,46	0,45
$F$	14,450	13,270	173,47	168,75
Константа $a$	-0,051	0,140	-0,04	-0,31
Коефіцієнт $b$	0,492	0,242	0,69	0,37

де  $\theta$  – вектор невідомих параметрів,  $\theta = \{\gamma, \eta, \varphi, \lambda\}$ ;  $\bar{z}$  – математичне сподівання нормалізованої випадкової величини  $z$ ,  $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$ ;  $S_z^2$  – незміщена оцінка

дисперсії величини  $z$ ,  $S_z^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2$ ;  $A_z$  – незміщена оцінка асиметрії величини  $z$ ;  $\varepsilon_z$  – незміщена оцінка ексцесу величини  $z$ .

Визначені за виразом (2) оцінки параметрів наведені в табл. 2.

Значення величин  $t$  та  $e$  були нормалізовані за перетворенням (1) з параметрами, вказаними в табл. 2. Було отримано нормалізовані випадкові величини  $z_t$  та  $z_e$ . Статистичні характеристики величин  $z_t$  та  $z_e$  для кожної платформи показані в табл. 3.

де  $a$  – константа;  $b$  – коефіцієнт нелінійної регресійної моделі.

Після побудови лінійної моделі (3) частина проєктів була вилучена через високі значення зовнішніх студентизованих залишків та плеча, після чого лінійна модель була побудована повторно. Крім того, аналогічна лінійна регресійна модель була побудована за величинами, нормалізованими з використанням логарифмічного перетворення, для того щоб порівняти два підходи. Характеристики побудованих лінійних регресійних моделей наведені в табл. 4.

Значення коефіцієнтів  $R^2$  для лінійних регресійних моделей, побудованих за даними, нормалізованими перетворенням Джонсона, є вищими для всіх платформ. Використання перетворення Джонсона дозволяє краще нормалізувати емпіричні дані, ніж застосування тільки логарифмічного перетворення, тому якість відповідних лінійних регресійних моделей є вищою.

Для логарифмічного перетворення регресійні моделі є такими:  $\log_{10} t = 0,140 + 0,242 \log_{10} e$  для платформи РС і  $\log_{10} t = -0,31 + 0,37 \log_{10} e$  для платформи MF. Або в традиційній формі:

$$\begin{aligned} t &= 1,38e^{0,242} \text{ для платформи РС;} \\ t &= 0,48e^{0,37} \text{ для платформи MF.} \end{aligned} \quad (4)$$

Нелінійна регресійна модель на основі перетворення Джонсона

$$t = \frac{\varphi_t + (\lambda_t + \varphi_t)c}{1 + c}, \quad (5)$$

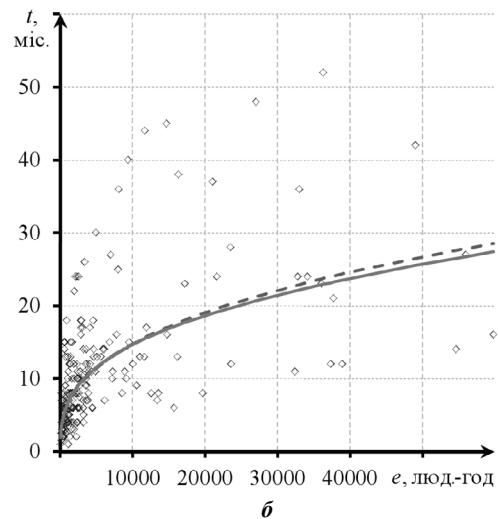
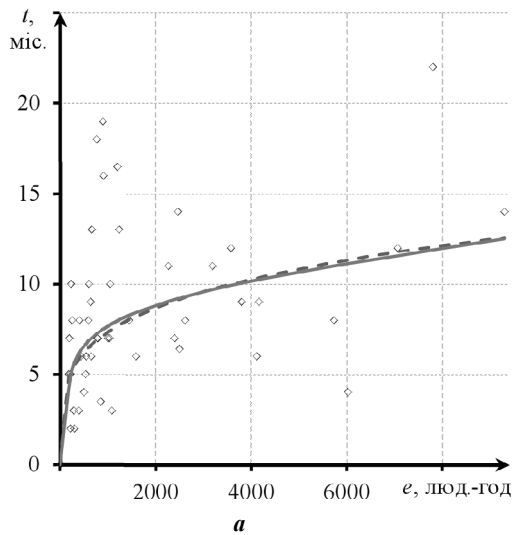


Рис. 1. Нелінійні регресійні моделі тривалості програмних проектів: а – платформа РС; б – платформа MF

Таблиця 5. Характеристики нелінійних регресійних моделей

Параметр	РС		MF	
	Перетворення Джонсона	Логарифмічне перетворення	Перетворення Джонсона	Логарифмічне перетворення
$R^2$	0,37	0,36	0,16	0,13
Відносна позиція	1	2	1	2
Сума квадратів залишків	11046	11059	845	873
Відносна позиція	1	2	1	2
Середнє ВЗ*	-0,19	-0,17	-0,14	-0,13
Відносна позиція	2	1	2	1
Середнє модуль ВЗ*	0,50	0,50	0,43	0,43
Відносна позиція	1	1	1	1
MRE Pred(0,25)**	0,37	0,36	0,47	0,39
Відносна позиція	1	2	1	2
RMS (рівень значень 25 % та вище)	0,77	0,77	0,62	0,61
Відносна позиція	1	1	2	1

\* ВЗ – відносні залишки.

\*\* Відсоток значень, передбачених з рівнем 25 % і вище ( $|B3| < 0,25$ ).

## ВИСНОВКИ

Побудовано нелінійні регресійні моделі тривалості програмних проектів залежно від їх трудомісткості з використанням перетворення Джонсона для платформ РС та MF. Порівняння побудованих моделей з моделями на основі логарифмічного перетворення

показало, що для оцінювання тривалості програмних проектів краще використовувати нелінійні регресійні на основі перетворення Джонсона. В подальших дослідженнях планується побудова нелінійної регресійної моделі тривалості програмних проектів залежно від їх трудомісткості з використанням перетворення Джонсона для платформи mid-range.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] **Приходько, С. Б.** Аналитическая зависимость для выбора распределения Джонсона семейства SL [Текст] / С. Б. Приходько, Л. Н. Макарова // Вестник ХНТУ. – Херсон : ХНТУ, 2012. – № 2 (45). – С. 101–104.
- [2] **Приходько, С. Б.** Інтервальне оцінювання статистичних моментів негаусівських випадкових величин на основі нормалізуючих перетворень [Текст] / С. Б. Приходько // Математичне моделювання : науковий журнал – Дніпродзержинськ, 2011. – № 1 (24). – С. 9–13.
- [3] **Приходько, С. Б.** Розробка нелінійної регресійної моделі тривалості програмних проектів на основі нормалізуючого перетворення Джонсона [Текст] / С. Б. Приходько, А. В. Пухалевич // Радіоелектронні і комп'ютерні системи. – Харків, 2012. – № 4 (56) – С. 90–93.
- [4] CHAOS Report [Електронний ресурс] // The Standish Group International, Inc. – 2009. – Режим доступу: <http://www.standishgroup.com>.
- [5] **Boehm, B. W.** Software engineering economics [Текст] / B. W. Boehm. – Englewood Cliffs, NJ : Prentice Hall, 1981. – 768 p.
- [6] **Johnson, R. A.** Applied Multivariate Statistical Analysis [Текст] / R. A. Johnson, D. W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.
- [7] **Oligny, S.** Exploring the relation between effort and duration in software engineering projects [Текст] / S. Oligny, P. Bourque, A. Abran, B. Fournier // In proc. of the World Computer Congress, Aug. 2000. – P. 175–178.

---

© С. Б. Приходько, А. В. Пухалевич

Надійшла до редколегії 24.02.2014

Статтю рекомендує до друку член редколегії ЗНП НУК  
д-р техн. наук, проф. *І. І. Коваленко*