

ТРЬОХ-ФАКТОРНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ КІЛЬКОСТІ ДЕФЕКТІВ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Кузьмічов І.С.¹, Пухалевич А.В. к.т.н.²

^{1,2} Національний університет кораблебудування імені адмірала Макарова

^{1,2} Україна, Миколаїв

² andrii.pukhalevych@nuos.edu.ua

Анотація. Модель, довірчі інтервали та інтервали передбачення трьохфакторної нелінійної регресії для оцінювання кількості дефектів в програмному забезпеченні побудовані на основі нормалізуючого перетворення у вигляді десяткового логарифму. Проведено порівняння побудованої моделі з моделлю лінійної регресії.

Ключові слова: оцінювання дефектів програмного забезпечення; нелінійна регресійна модель; довірчий інтервал; інтервал прогнозування; нормалізуюче перетворення.

Необхідність прогнозування дефектів програмного забезпечення (ПЗ) є важливою на сьогодні як частина ініціатив щодо вдосконалення процесів виробництва ПЗ, його якості та надійності. За останні 10 років було опубліковано сотні різних моделей прогнозування дефектів, включаючи і регресійні моделі [1]. У [1] запропоновано кілька лінійних рівнянь регресії для оцінювання кількості дефектів ПЗ, розробленого в Java, PHP та .NET, включаючи всі дефекти та функціональні дефекти. Однак, по-перше, результат прогнозування за рівнянням регресії є середнім значенням залежної випадкової величини. У рівнянні регресії відсутня складова, що описує випадкову помилку. Результатом моделі регресії прогнозування є значення залежної випадкової величини, оскільки в регресійній моделі є випадкова складова. По-друге, є чотири основні припущення, які обґрунтовують використання лінійних регресійних моделей, одне з яких – це нормальність розподілу помилок. Але це припущення справедливе лише в окремих випадках. Тому для оцінювання кількості дефектів як значення залежної випадкової величини існує необхідність розробки відповідних нелінійних регресійних моделей [2]. А враховуючи на те, що розподіл кількості дефектів не є гаусівським, потрібно використовувати саме нелінійні регресійні моделі, а їх побудову вести на основі нормалізуючих перетворень [3].

В роботі для оцінювання кількості дефектів ПЗ пропонується використовувати трьохфакторну нелінійну регресійну модель, що побудована на основі нормалізуючого перетворення у вигляді десяткового логарифму за даними з [2]:

$$Y = 10^{\varepsilon_x + \hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}, \quad (1)$$

де ε_x – гаусівська випадкова величина, $\varepsilon_x \sim N(0, \sigma_x)$; X_1 – фактичний розмір ПЗ в тисячах рядків коду; X_2 – загальні зусилля перед тестуванням в днях; X_3 – кількість тестових випадків; $\hat{b}_0 = -0.3354$; $\hat{b}_1 = 0.5185$; $\hat{b}_2 = -0.4248$; $\hat{b}_3 = 0.6723$.

Також було побудовано трьохфакторну лінійну регресійну модель за даними з [2]

$$Y = -1.047 + 0.7918X_1 - 0.0488X_2 + 0.0356X_3 + \varepsilon_x. \quad (2)$$

Для порівняння моделей (1) та (2) були визначені значення множинного коефіцієнту детермінації R^2 , середньої величини відносної помилки MMRE та відсотка прогнозування PRED(0,25). Значення R^2 краще для моделі (2) – 0,8237 у порівнянні з 0,6640 для моделі (1).

Значення MMRE та PRED (0,25), навпаки, кращі для моделі (1) – 0,2304 і 0,6571 відповідно у порівнянні з 0,3617 і 0,5143 для моделі (2).

Також було побудовано довірчі інтервали та інтервали передбачення трьохфакторних лінійної і нелінійної регресій. Ширини довірчих інтервалів та інтервалів передбачення виявилися меншими для нелінійної регресії у порівнянні з лінійною для більшої кількості даних з [2].

Висновки.

Удосконалено трьохфакторну нелінійну регресійну модель нелінійної регресії для оцінювання кількості дефектів ПЗ в залежності від фактичного розміру ПЗ, загальних зусиль перед тестуванням, кількості тестових випадків на основі нормалізуючого перетворення у вигляді десяткового логарифму, що дозволяє підвищити достовірність оцінювання залежної змінної нелінійної регресії у порівнянні з лінійною регресійною моделлю. Модель, що побудована, в порівнянні з лінійною регресійною моделлю, має більше значення відсотку передбачень та менші середні величини відносної похибки і ширини інтервалу передбачення нелінійної регресії. В подальшому планується використання інших наборів даних для побудови нелінійної регресійної моделі для оцінювання кількості дефектів ПЗ.

REFERENCES

- [1] Suffian, M.D.M., & Ibrahim, S. (2012). A prediction model for system testing defects using regression analysis. *International Journal of Soft Computing and Software Engineering*, 2, (7), 69–78. DOI: 10.7321/jsce.v2.n7.6
- [2] Prykhodko, S., Prykhodko, N., Makarova, L., Prykhodko, K., Pukhalevych, A., & Smykodub, T. (2019). *A Non-Linear Regression Model to Estimate the Number of Defects for Software Testing Phase*. Proceedings of the 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON), July 2 - 6, 2019, Lviv, Ukraine. DOI: 10.1109/UKRCON.2019.8880027
- [3] Prykhodko, N.V., & Prykhodko, S.B. (2018). Constructing the non-linear regression models on the basis of multivariate normalizing transformations. *Electronic modeling*, 40, (6), 101–110. DOI: 10.15407/emodel.40.06.101

Kuzmichov I.S., Pukhalevych A.V.

A three-factor regression model to estimate the software defect number

Annotation. The model, confidence and prediction intervals of three-factor non-linear regression for estimating the number of defects in software are constructed on the basis of the normalizing transformation in the decimal logarithm form. Comparison of the constructed model with the linear regression model is performed.

Keywords: software defect estimation; non-linear regression model; confidence interval; prediction interval; normalizing transformation.

Кузьмичов И.С., Пухалевич А.В.

Трехфакторная модель регрессии для оценки числа дефектов программного обеспечения

Аннотация. Модель, доверительные интервалы и интервалы прогнозирования трехфакторной нелинейной регрессии для оценки числа дефектов в программном обеспечении построены на основе нормализующего преобразования в виде десятичного логарифма. Проведено сравнение построенной модели с моделью линейной регрессии.

Ключевые слова: оценка дефектов программного обеспечения, нелинейная регрессионная модель, доверительный интервал, интервал прогнозирования, нормализующее преобразование.