УДК 004.942:519.25

Приходько С. Б., Макарова Л. Н.

ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА НЕЛИНЕЙНОГО УРАВНЕНИЯ РЕГРЕССИИ ВРЕМЕНИ ВОССТАНОВЛЕНИЯ РАБОТОСПОСОБНОСТИ УСТРОЙСТВ ТЕРМИНАЛЬНОЙ СЕТИ

Выполнено построение доверительного интервала нелинейного уравнения регрессии времени восстановления работоспособности устройств терминальной сети на основе как нормализирующего преобразования Джонсона, так и t-распределения Стьюдента. Проведено сравнение полученных результатов.

Ключевые слова: *доверительный интервал, нелинейная регрессия, нормализирующее преобразование, преобразование Джонсона, терминальная сеть.*

Виконано побудову довірчого інтервалу нелінійного рівняння регресії часу відновлення працездатності пристроїв термінальної мережі на основі як нормалізуючого перетворення Джонсона, так і t-розподілу Стьюдента. Проведено порівняння отриманих результатів.

Ключові слова: *довірчий інтервал, нелінійна регресія, нормалізуюче перетворення, перетворення Джонсона, термінальна мережа.*

A confidence interval of the non-linear regression equation of time restore functionality of network terminal devices based on Johnson normalizing transformation and Student t-distribution is built. Comparison of the results is made.

Key words: *confidence interval, non-linear regression, normalizing transformation, Johnson transformation, terminal network.*

Постановка проблемы. На сегодняшний день при построении информационной технологии автоматизированной системы обработки информации терминальной сети оценка времени восстановления работоспособности устройств терминальной сети играет важную роль [1]. Время восстановления работоспособности является случайной величиной (СВ), которая обычно не подчиняется нормальному закону распределения, и зависит от ряда факторов, в том числе и от расстояния между центром обслуживания и устройством терминальной сети. В этом случае оценку времени восстановления работоспособности можно осуществлять на основе соответствующей регрессионной модели [2; 3], которая будет нелинейной, т. к. указанная СВ не является гауссовской.

По эмпирическим данным возможно найти только оценку истинного уравнения регрессии, которая будет являться точечной. Более надежную оценку возможно получить с помощью доверительного интервала уравнения регрессии.

В случае нормального закона распределения СВ возможно построить линейную регрессионную модель и по ней оценить доверительный интервал уравнения регрессии. Однако, когда закон распределения СВ отличается от нормального, имеем, как правило, нелинейную регрессионную модель, для которой затруднено построение доверительного интервала уравнения регрессии без предположения о нормальности.

Поэтому проблема нахождения доверительного интервала нелинейного уравнения регрессии времени восстановления работоспособности устройств терминальной сети является актуальной.

Анализ последних исследований и публикаций. При нормальном законе распределения СВ доверительный интервал линейного уравнения регрессии возможно построить по методу, представленному в [4]. Однако данный метод не учитывает ряд особенностей эмпирической модели распределения данных, а именно: разное количество точек и разную дисперсию при фиксированных значениях аргумента.

Помимо этого подхода существует подход, основанный на применении нормализирующих преобразований Джонсона [5; 6], суть которого состоит в следующем: с помощью нормализирующего преобразования осуществить переход от исходных негауссовских СВ к гауссовским СВ, для которых построить линейное уравнение регрессии традиционным способом, найти для него доверительный интервал, а затем с помощью обратного преобразования перейти к нелинейному уравнению регрессии с доверительным интервалом.

Данный метод лишен недостатков, отмеченных у предыдущего.

Целью данной статьи является построение доверительного интервала нелинейного уравнения регрессии времени восстановления работоспособности устройств терминальной сети на основе нормализирующего преобразования Джонсона.

Изложение основного материала. В общем виде регрессионная модель может быть представлена уравнением:

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

где y – зависимая переменная, или результативный признак;

x – независимая переменная, или фактор; ε_t – случайная ошибка, или возмущение;

$f(x)$ – функция, которая определяет вид регрессионной модели – нелинейная или линейная.

Для получения линейной регрессионной модели исходные СВ необходимо нормализировать с помощью преобразования Джонсона, которое в общем виде имеет вид [7]:

$$z = \gamma + \eta h\,(x, \varphi, \lambda);\ \eta > 0;\ -\infty < \gamma < \infty;$$
$$\lambda > 0;\ -\infty < \varphi < \infty. \qquad (2)$$

Преобразование (2) имеет обратное преобразование:

$$x = \varphi + \lambda h^{-1}\,(z, \gamma, \eta);\ \eta > 0;\ -\infty < \gamma < \infty;$$
$$\lambda > 0;\ -\infty < \varphi < \infty. \qquad (3)$$

где z – нормально распределенная случайная величина с математическим ожиданием ноль и дисперсией единица;

x – случайная величина с распределением Джонсона;

γ, η, φ, λ – параметры преобразования или распределения Джонсона;

h и h^{-1} – функции определенного семейства:

$$h = \begin{cases} \ln(\tilde{x}), & x > \varphi, \quad \text{для семейства } S_L; \\ \ln[\tilde{x}/(1-\tilde{x})], & \varphi < x < \varphi + \lambda, \quad \text{для семейства } S_B; \\ \mathrm{Arsh}(\tilde{x}), & -\infty \le x \le +\infty, \quad \text{для семейства } S_U, \end{cases}$$

$$h^{-1} = \begin{cases} e^{\zeta}, & \text{для семейства } S_L; \\ 1/\left(1+e^{-\zeta}\right), & \text{для семейства } S_B; \\ \left(e^{\zeta}-e^{-\zeta}\right)/2, & \text{для семейства } S_U. \end{cases} \qquad (4)$$

Конкретное семейство распределений Джонсона выбирается исходя из значений квадрата асимметрии A^2 и эксцесса ε исходной выборки [5]. Значения неизвестных параметров распределения можно найти с помощью непараметрического метода решения задачи математического программирования, описанного в [5]:

$$\hat{\theta} = \arg\min_{\theta}\left\{A_z^2 + (\varepsilon_z - 3)^2 + m_z^2 + (D_z - 1)^2\right\} \qquad (5)$$

где θ – вектор неизвестных параметров, $\theta = \{\gamma, \eta, \varphi, \lambda\}$;

A_z – оценка асимметрии распределения z;

ε_z – оценка эксцесса распределения z;

m_z – оценка математического ожидания z;

D_z – оценка дисперсии z;

n – количество значений z в выборке.

Проверку соответствия преобразованных выборок нормальному распределению можно выполнить с помощью критериев согласия, например, χ^2 Пирсона [8].

В общем виде линейная регрессионная модель нормализированных значений СВ может быть представлена уравнением:

$$z_y = b_1 z_x + b_0, \qquad (6)$$

где b_1, b_0 – коэффициенты линейной регрессии, которые находятся методом наименьших квадратов.

Для проверки адекватности полученной линейной регрессионной модели используем критерий R^2, называемый коэффициентом детерминации [9]:

$$R^2 = 1 - \left(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2\right), \qquad (7)$$

где y_i – фактическое значение y;

$\hat{y}_i$ – расчетное значение y;

$\bar{y}$ – среднее значение y.

Достаточно качественной можно признать модель при $R^2 \ge 0{,}8$, приемлемой – при $R^2 \ge 0{,}5$.

Используя линейную регрессионную модель (6) и нормализирующее преобразование Джонсона (2) и (4), строим нелинейную регрессионную модель времени восстановления работоспособности устройств терминальной сети:

$$y = \frac{e^{c_1}(\lambda_y + \varphi_y) + \varphi_y}{1 + e^{c_1}}, \qquad (8)$$

где $c_1 = \frac{1}{\eta_y}\cdot\left(b_1 \cdot c_2 + b_0 - \gamma_y\right), c_2 = \gamma_x + \eta_x \ln\left(\frac{x - \varphi_x}{\lambda_x + \varphi_x - x}\right)$.

Алгоритм построения (1-α) % доверительного интервала линейного уравнения регрессии можно представить следующим образом:

1. При каждом фиксированном значении z_{xi} рассмотрим соответствующие ему значения z_{yi}, $i \in [0;\ m]$, как малую выборку, которую повторно нормализируем с помощью преобразования Джонсона. Для полученной выборки z_y^* найдем выборочное среднее m_{zy}^* и выборочное среднеквадратическое отклонение S_{zy}^*.

2. Используя традиционный способ на основе t-распределения Стьюдента для выборочного среднего нормальной выборки [10] найдем $\Delta z_y^* = t_{n-1} S_{zy}^* / \sqrt{m}$ и получим границы доверительного интервала выборочного среднего СВ z_y^*: $z_y^* \min = \bar{z}_y^* - \Delta z y^*$ и $z_y^* \max = \bar{z}_y^* + \Delta z y^*$.

3. По обратному преобразованию (3) и (4) для значений $z_y^* \min$ и $z_y^* \max$ получаем границы доверительного интервала выборочного среднего СВ z_y: $z_y \min$ и $z_y \max$.

4. По всем значениям $z_y \min$ строим нижнюю границу доверительного интервала уравнения регрессии с помощью квадратичного полинома вида:

$$z_y = d_2 z_x^2 + d_1 z_x + d_0, \qquad (9)$$

где d_2, d_1, d_0 – коэффициенты регрессии, которые находим методом наименьших квадратов.

5. По всем значениям z_y max строим верхнюю границу доверительного интервала уравнения регрессии по формуле (9) аналогично п. 4.

(1-α) % доверительный интервал нелинейного уравнения регрессии можно построить, используя регрессионную модель (9) для нижней и верхней границ и нормализирующее преобразование Джонсона (2) и (4):

$$y = \frac{e^{k_1}(\lambda_y + \varphi_y) + \varphi_y}{1 + e^{k_1}}, \quad (10)$$

где $k_1 = \frac{1}{\eta_y} \cdot \left(d_2 \cdot k_2^2 + d_1 \cdot k_2 + d_0 - \gamma_y\right)$, $k_2 = \gamma_x + \eta_x \ln\left(\frac{x - \varphi_x}{\lambda_x + \varphi_x - x}\right)$.

Проверку представленного метода выполним на основании данных, приведенных в [11]. Расчет велся на примере одной из выборок данных, где СВ x – расстояние от центра обслуживания до терминального устройства, м; СВ y – время восстановления работоспособности терминального устройства, мин. Законы распределения СВ x и СВ y не являются нормальными: A_x = 1,0885; ε_x = 2,9193; A_y = 1,6538; ε_y = 5,2015. Выполним нормализацию СВ x и СВ y с помощью нормализирующего преобразования Джонсона (2) и (4). Параметры преобразования найдем в результате решения задачи (5).

Для нормализации СВ *x* было выбрано семейство распределений S_B Джонсона. Параметры распределения: γ_x = 0,9676; η_x = 0,5755; φ_x = 1874,59; λ_x = 8300,00 при значении целевой функции $1{,}1319 \cdot 10^{-1}$. С доверительной вероятностью 0,95 гипотеза о соответствии преобразованной выборки нормальному закону распределения СВ принимается (χ^2_x = 6,25 при критическом значении χ^2 = 11,07).

Для нормализации СВ *y* было выбрано семейство распределений S_B Джонсона. Параметры распределения: γ_x = 1,2500; η_x = 0,5431; φ_x = 10,2939; λ_x = 2360,61 при значении целевой функции $2{,}0530 \cdot 10^{-6}$. С доверительной вероятностью 0,95 гипотеза о соответствии преобразованной выборки нормальному закону распределения СВ принимается (χ^2_y = 6,14 при критическом значении χ^2 = 9,49).

После нормализации выполняем группировку полученных значений z_y при фиксированных значениях z_{xi}, далее отбрасываем группы, в которые попало менее семи значений. Для каждой из оставшихся групп находим выборочное среднее m_{zy}, по которым и строим линейную регрессионную модель нормализированных значений $z_y(z_x)$ в соответствии с (6): $zy = 0{,}8736 z_x + 0{,}0629$, критерий R^2 составляет 0,9778.

Далее по приведенному выше алгоритму строим (1-α) % доверительный интервал линейного уравнения регрессии. Уравнение нижней границы в соответствии с (9) имеет вид $zy = 0{,}0293 z_x^2 + 0{,}988 z_x - 0{,}2975$, критерий R^2 составляет 0,978. Уравнение верхней границы в соответствии с (9) имеет вид $zy = 0{,}0936 z_x^2 + 0{,}88 z_x + 0{,}25$, критерий R^2 составляет 0,9848. Полученные значения R^2 говорят об адекватности построенных уравнений.

Сравним найденный доверительный интервал линейного уравнения регрессии с доверительным интервалом, приведенным для линейного уравнения регрессии в [5]:

$$z_y = b_1 \cdot z_x + b_0 \pm t(\alpha/2, n-2) \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(z_x - \bar{x})^2}{\sum_{i=1}^{n} (z_{xi} - \bar{x})^2}}.$$

Уравнение нижней границы имеет вид

$$z_y = 0{,}8736 z_x + 0{,}0629 - 1{,}1405 \sqrt{0{,}007 + \frac{(z_x + 0{,}0578)^2}{142{,}7807}}.$$

Уравнение верхней границы имеет вид

$$z_y = 0{,}8736 z_x + 0{,}0629 + 1{,}1405 \sqrt{0{,}007 + \frac{(z_x + 0{,}0578)^2}{142{,}7807}}.$$

Для обоих уравнений критерий R^2 составляет 0,7452, что говорит об адекватности построенных уравнений, однако метод с использованием нормализирующего преобразования Джонсона дает существенно лучший результат.

Доверительный интервал линейного уравнения регрессии нормализированного времени восстановления работоспособности устройств терминальной сети приведен на рис. 1.

В соответствии с (10) построим (1-α) % доверительный интервал нелинейного уравнения регрессии времени восстановления работоспособности устройств терминальной сети, критерий R^2 составляет 0,8082, что говорит об адекватности построенной модели. Данный доверительный интервал приведен на рис. 2.

Выводы. Впервые построен доверительный интервал нелинейного уравнения регрессии времени восстановления работоспособности устройств терминальной сети на основе нормализирующего преобразования Джонсона. В дальнейшем планируется построение информационной технологии автоматизированной системы обработки информации терминальной сети на основе полученных результатов.

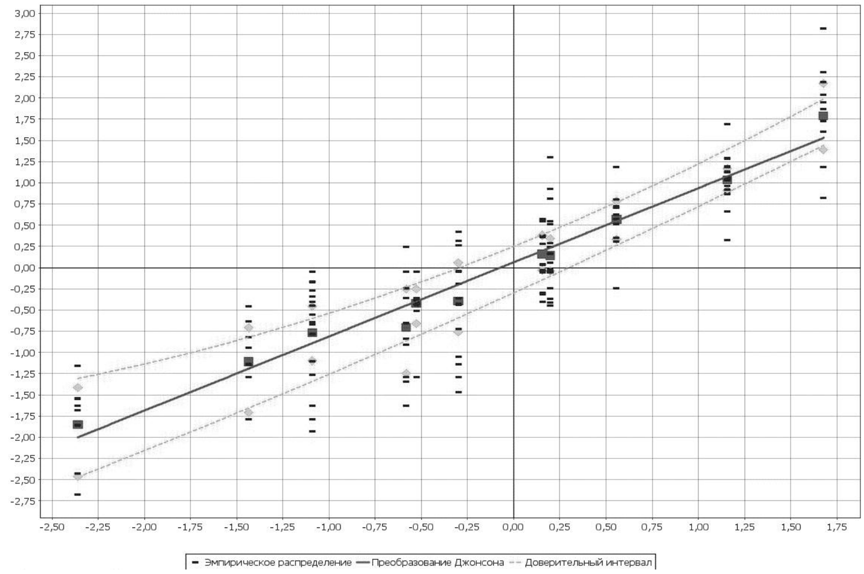

Рис. 1. Доверительный интервал линейного уравнения регрессии нормализированного времени восстановления работоспособности устройств терминальной сети

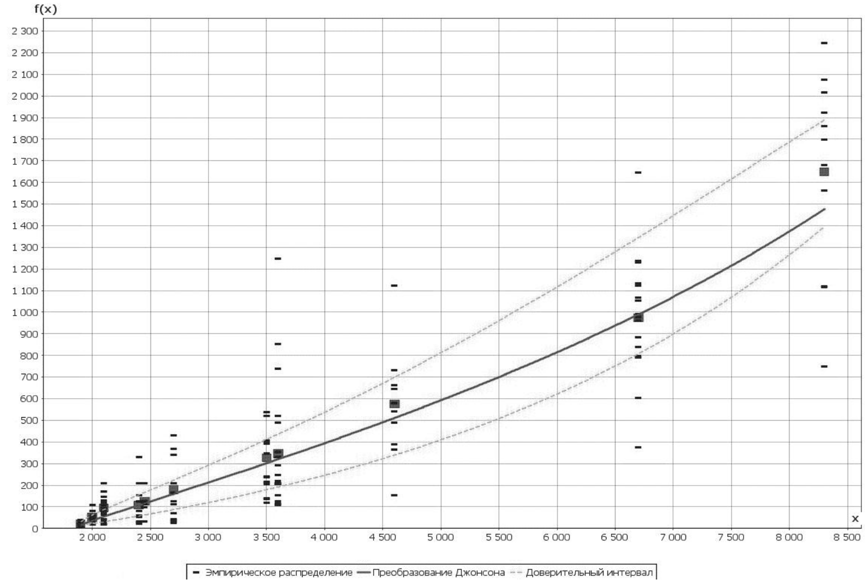

Рис. 2. Доверительный интервал нелинейного уравнения регрессии нормализированного времени восстановления работоспособности
устройств терминальной сети

ЛИТЕРАТУРА

1. Наша цель – банк в шаговой доступности [Електронний ресурс]. – Режим доступу : http://www.inpas.ru/publications/78/.
2. Грешилов А. А. Математические методы построения прогнозов [Текст] / А. А. Грешилов, В. А. Стакун, А. А. Стакун. – М. : Радио и связь, 1997. – 112 с.: ил.
3. Демиденко Е. З. Линейная и нелинейная регрессии [Текст] / Е. З. Демиденко – М. : Финансы и статистика, 1981. – 302 с., ил.
4. Yan Xin. Linear regression analysis: theory and computing [Text] / Xin Yan, Xiao Gang Su – Singapore : World Scientific Publishing Co. Pte. Ltd., 2009. – 328 p.
5. Приходько С. Б Інтервальне оцінювання статистичних моментів негаусівських випадкових величин на основі нормалізуючих перетворень [Текст] / С. Б. Приходько // Математичне моделювання : науковий журнал. – Дніпродзержинськ : ДДТУ, 2011. – № 1 (24). – С. 9–13.
6. 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.
7. Кендалл М. Теория распределений [Текст] / М. Кендалл, А. Стьюарт. – М. : Наука, 1966. – 588 с.

8. Вентцель Е. С. Теория вероятностей : [учеб. для вузов] / Е. С. Вентцель. – М. : Высш. шк., 1999. – 576 с.
9. Магнус Я. Р. Эконометрика. Начальный курс : [учебник] / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – [6-е изд., перераб. и доп.]. – М. : Дело, 2004. – 576 с.
10. Поллард Дж. Справочник по вычислительным методам статистики [Текст] / Дж. Поллард ; [пер. с англ. В. С. Занадворова ; под ред. и с предисл. Е. М. Четыркина]. – М. : Финансы и статистика, 1982. – 344 с.
11. Приходько С. Б. Определение доверительных интервалов статистических моментов времени наработки между отказами устройств терминальной сети / С. Б. Приходько, Л. Н. Макарова // Наукові праці : науково-методичний журнал. – Вип. 201. Т. 213. Комп'ютерні технології. – Миколаїв : Вид-во ЧДУ ім. Петра Могили, 2013. – С. 82–86.

 Дата надходження статті до редколегії 10.06.2014 р.

ПРИХОДЬКО Сергей Борисович – доктор технических наук, доцент, заведующий кафедрой программного обеспечения автоматизированных систем, Национальный университет кораблестроения имени адмирала Макарова.

Круг научных интересов: математическое моделирование стохастических процессов в информационных технологиях

МАКАРОВА Лидия Николаевна – соискатель, Национальный университет кораблестроения имени адмирала Макарова.

Круг научных интересов: математическое моделирование стохастических процессов в информационных технологиях