

# ПОСТРОЕНИЕ НЕЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ ВРЕМЕНИ ВОССТАНОВЛЕНИЯ РАБОТОСПОСОБНОСТИ УСТРОЙСТВ ТЕРМИНАЛЬНОЙ СЕТИ

УДК 004.942:519.25

## ПРИХОДЬКО Сергей Борисович

д.т.н., доцент, заведующий кафедрой программного обеспечения автоматизированных систем,  
Национальный университет кораблестроения имени адмирала Макарова.

**Научные интересы:** математическое моделирование случайных процессов в информационных технологиях.

## МАКАРОВА Лидия Николаевна

соискатель, Национальный университет кораблестроения имени адмирала Макарова.

**Научные интересы:** математическое моделирование случайных процессов в информационных технологиях.

### ПОСТАНОВКА ПРОБЛЕМЫ

Оценка времени восстановления работоспособности устройств терминальной сети играет важную роль в практических задачах ее управления [1]. Время восстановления работоспособности является случайной величиной (СВ), которая как правило не подчиняется нормальному закону распределения, и зависит от ряда факторов: модели конкретного устройства, вида отказа в обслуживании, сезона года, 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].

Однако при построении подобных регрессионных моделей возникает ряд сложностей:

– если СВ, входящие в регрессионную модель, не подчиняются нормальному закону распределения, это приводит к нелинейной регрессии;

– если зависимая СВ зависит одновременно от двух и более факторов, что обычно и бывает при решении практических задач, получаем множественную регрессию.

### АНАЛИЗ ПОСЛЕДНИХ ИССЛЕДОВАНИЙ И ПУБЛИКАЦИЙ

Для построения нелинейных регрессионных моделей существует несколько методов: метод простого перебора, методы линеаризирующих и нормализующих преобразований [4].

Метод простого перебора, или метод всех возможных регрессий, требует задания различных видов уравнения регрессии и выбора наилучшего приближения из заданных по определенному критерию. Это требует достаточно большого количества вычислений и не всегда приводит к получению наилучшего решения. Все эти недостатки делает данный метод неэффективным [4].

Целью линеаризирующих преобразований является переход от нелинейной регрессии к линейной путем замены исходных переменных и коэффициентов, однако не всегда возможно подобрать такую замену [5, 6].

Суть метода нормализующих преобразований состоит в следующем. Сначала ищут преобразования, с помощью которых удастся осуществить переход от исходных негауссовских СВ к гауссовским СВ. Для преобразованных гауссовских СВ строят линейное уравне-

ние регрессии, которое далее преобразуют в нелинейную модель с помощью ранее найденных нормализующих преобразований [6].

Простейшим нормализующим преобразованием является логарифмирование исходных СВ. Лучшие результаты дает подход на основе нормализующего преобразования Джонсона [7, 8].

**Целью статьи** является построение нелинейной регрессионной модели времени восстановления работоспособности устройств терминальной сети на основе нормализующего преобразования Джонсона.

### ИЗЛОЖЕНИЕ ОСНОВНОГО МАТЕРИАЛА

В общем виде регрессионная модель может быть представлена уравнением:

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

где  $y$  – зависимая переменная, или результативный признак,  $x$  – независимая переменная, или фактор,  $\varepsilon_t$  – случайная ошибка, или возмущение,  $f(x)$  – функция, которая определяет вид регрессионной модели: нелинейная или линейная.

Нелинейные регрессионные модели могут быть двух видов:

– нелинейные относительно независимых переменных, но линейные по оцениваемым параметрам, например, полиномы различных степеней, гиперболические функции;

– нелинейные по оцениваемым параметрам, например, степенная, показательная, экспоненциальная функции.

В случае нелинейной регрессионной модели значения параметров находят с помощью методов градиентного спуска или при помощи замены переменной и сведения нелинейной регрессионной модели к линейной, в которой присутствует строго линейная связь между переменными.

Поскольку СВ, подчиняющиеся нормальному закону распределения, гарантировано дадут линейное уравнение регрессии [6], необходимо нормализовать исходные СВ с помощью нормализующего преобразования Джонсона, которое в общем виде имеет вид [9]:

$$z = \gamma + \eta q(x, \phi, \lambda); \quad (2)$$

$$-\infty < \gamma < \infty; \eta > 0; -\infty < \phi < \infty; \lambda > 0,$$

где  $q$  – нелинейная функция;  $\gamma, \eta, \phi, \lambda$  – параметры преобразования, причем  $\gamma$  и  $\eta$  – параметры формы,  $\phi$  – параметр смещения,  $\lambda$  – параметр масштаба;  $x$  – СВ, которая нормализуется;  $z$  – нормированная нормально распределенная СВ с нулевым математическим ожидание.

Джонсон предложил три семейства функций  $q$ :

$$q_1(x, \phi, \lambda) = \ln\left(\frac{x - \phi}{\lambda}\right), x > \phi;$$

$$q_2(x, \phi, \lambda) = \ln\left(\frac{x - \phi}{\lambda + \phi - x}\right), \phi < x < \phi + \lambda; \quad (3)$$

$$q_3(x, \phi, \lambda) = \operatorname{Arsh}\left(\frac{x - \phi}{\lambda}\right), -\infty \leq x \leq +\infty.$$

Функция  $q_1$  относится к семейству  $S_L$ , функция  $q_2$  – семейству  $S_B$ , функция  $q_3$  – семейству  $S_U$ . Каждому семейству соответствует своя функция плотности вероятности. Конкретное семейство преобразования (распределения) Джонсона выбирается исходя из значений квадрата асимметрии  $A^2$  и эксцесса  $\varepsilon$  исходной выборки [10]. Значения неизвестных параметров преобразования Джонсона в общем случае можно найти с помощью непараметрического метода путем решения задачи математического программирования [11]:

$$\hat{\theta} = \arg \min_{\theta} \left\{ A_z^2 + (\varepsilon_z - 3)^2 + m_z^2 + (D_z - 1)^2 \right\}, \quad (4)$$

где  $\theta$  – вектор неизвестных параметров,  $\theta = \{\gamma, \eta, \phi, \lambda\}$ ,  $A_z = \frac{1}{n\sigma_z^3} \sum_{i=1}^n (z_i - m_z)^3$  – оценка асимметрии распределения  $z$ ,  $\varepsilon_z = \frac{1}{n\sigma_z^4} \sum_{i=1}^n (z_i - m_z)^4$  – оценка эксцесса распределения  $z$ ,  $m_z = \frac{1}{n} \sum_{i=1}^n z_i$  – оценка математического ожидания  $z$ ,  $D_z = \sigma_z^2 = \frac{1}{n} \sum_{i=1}^n (z_i - m_z)^2$  – оценка дисперсии  $z$ ,  $n$  – количество значений  $z$  в выборке.

Проверку соответствия преобразованных выборок нормальному распределению можно выполнить с помощью критериев согласия, например  $\chi^2$  Пирсона или Колмогорова-Смирнова [12].

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

$$z_y = b_1 z_x + b_0, \quad (5)$$

где  $b_1, b_0$  – коэффициенты линейной регрессии.

Параметры линейной регрессионной модели (5) находятся известным методом наименьших квадратов.

Для проверки адекватности полученной линейной регрессионной модели используем критерий  $R^2$ , называемый коэффициентом детерминации [13]:

$$R^2 = 1 - \left( \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right), \quad (6)$$

где:  $y_i$  – фактическое значение  $y$ ,  $\hat{y}_i$  – расчетное значение  $y$ ,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  – среднее значение  $y$ .

Чем ближе  $R^2$  к единице, тем выше качество модели. При  $R^2=1$  линия регрессии точно соответствует всем наблюдениям. Достаточно качественной можно признать модель при  $R^2 \geq 0,8$ , приемлемой – при  $R^2 \geq 0,5$ .

Используя линейную регрессионную модель (5) и нормализующее преобразование Джонсона (2) и (3), строим нелинейную регрессионную модель времени восстановления работоспособности устройств терминальной сети:

$$y = \frac{e^c (\lambda_y + \phi_y) + \phi_y}{1 + e^c}, \quad (7)$$

где:

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

Для проверки адекватности полученной нелинейной регрессионной модели кроме критерия  $R^2$ , используем также сумму квадратов отклонений  $S_y$ :

$$S_y = \sum_{i=1}^n [y_i - f(x_i)]^2, \quad (8)$$

где:  $y_i$  – фактическое значение СВ  $y$ ,  $f(x_i)$  – расчетное значение СВ  $y$  согласно уравнению (7).

Для построения нелинейной регрессионной модели в качестве исходных данных были рассмотрены 28106 случаев отказа в обслуживании 278 устройств терминальной сети за период 36 месяцев, приведенные в работах [14, 15]. Пусть 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:  $A_x=1,0885$ ;  $\varepsilon_x=2,9193$ ;  $A_y=1,6538$ ;  $\varepsilon_y=5,2015$ . Выполним нормализацию СВ  $x$  и СВ  $y$  с помощью нормализующего преобразования Джонсона (2) и (3). Параметры преобразования найдем в результате решения задачи (4).

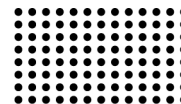
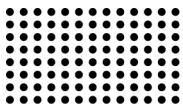
**Для нормализации СВ  $x$**  исходя из значений квадрата асимметрии  $A_x^2$  и эксцесса  $\varepsilon_x$  исходной выборки, было выбрано семейство распределений  $S_B$  Джонсона. Параметры распределения:  $\gamma_x=0,9676$ ;  $\eta_x=0,5755$ ;  $\phi_x=1874,59$ ;  $\lambda_x=8300,00$  при значении целевой функции  $1,1319 \cdot 10^{-1}$ .

Для проверки соответствия преобразованного распределения нормальному закону применялся критерий согласия  $\chi^2$  Пирсона:  $\chi^2_x=6,25$  при критическом значении  $\chi^2=11,07$ . С доверительной вероятностью 0,95 гипотеза о соответствии преобразованной выборки нормальному закону распределения СВ принимается.

Эмпирическое распределение СВ  $x$  приведено на рис. 1а, а ее нормализованное распределение  $z_x$  – на рис. 1б.

**Для нормализации СВ  $y$**  исходя из значений квадрата асимметрии  $A_y^2$  и эксцесса  $\varepsilon_y$  исходной выборки, было выбрано семейство распределений  $S_B$  Джонсона. Параметры распределения:  $\gamma_y=1,2500$ ;  $\eta_y=0,5431$ ;  $\phi_y=10,2939$ ;  $\lambda_y=2360,61$  при значении целевой функции  $2,0530 \cdot 10^{-6}$ .

Для проверки соответствия преобразованного распределения нормальному закону применялся критерий согласия  $\chi^2$  Пирсона:  $\chi^2_y=6,14$  при критическом значении  $\chi^2=9,49$ . С доверительной вероятностью 0,95



гіпотеза о соответствии преобразованной выборки нормальному закону распределения СВ принимается.

Эмпирическое распределение СВ у приведено на рис. 2а, а ее нормализованное распределение  $z_y$  – на рис. 2б.

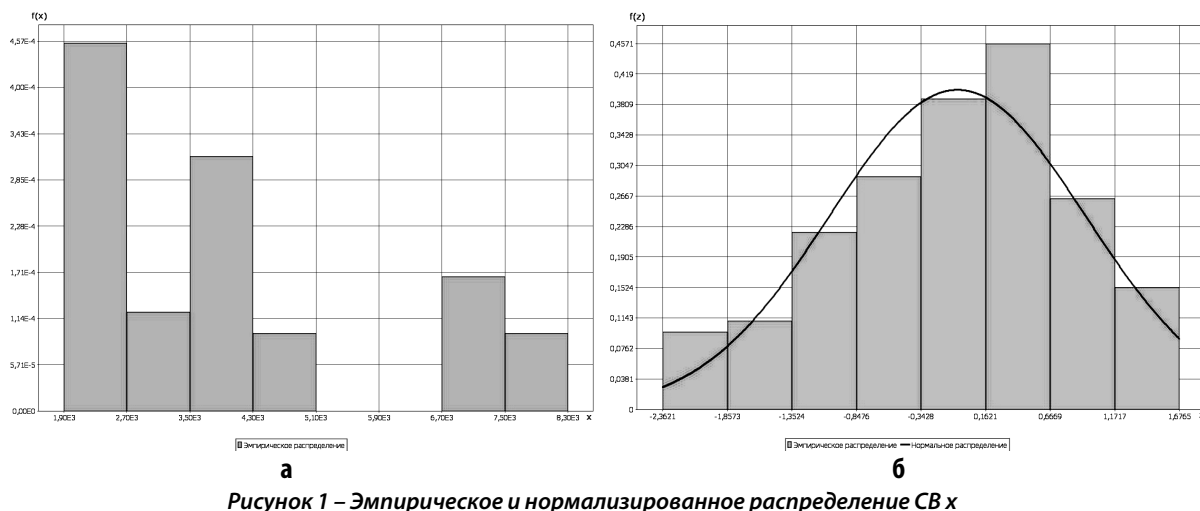


Рисунок 1 – Эмпирическое и нормализованное распределение СВ x

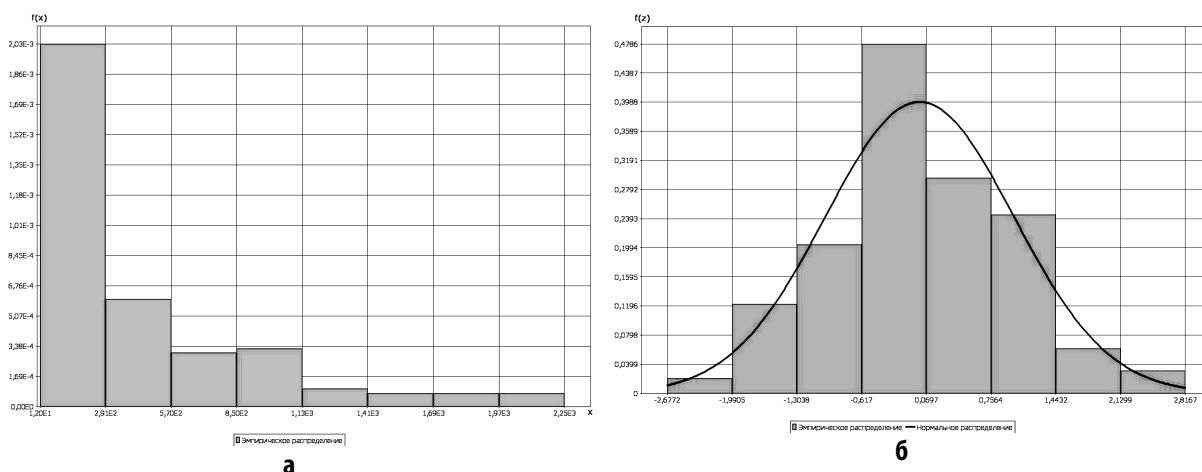


Рисунок 2 – Эмпирическое и нормализованное распределение СВ y

После нормализации выполняем группировку нормализованных значений  $z_y$  при одинаковых фиксированных значениях  $z_x$ , далее отбрасываем группы, в которые попало менее семи значений. Для каждой из оставшихся групп находим выборочное среднее  $m_{z_y}$ , по которым и строим линейную регрессионную модель нормализованных значений  $z_y(z_x)$  в соответствии с (5). Коэффициенты линейного уравнения регрессии находим с помощью метода наименьших квадратов:  $b_1=0,8765$ ;  $b_0=0,0495$ . Линейная регрессионная модель нормализованного времени восстановления работоспособности устройств терминальной сети приведена на рис. 3.

Сравним построенную линейную регрессионную модель с моделью на основе логарифмического преобразования, для построения которой использовался десятичный логарифм. Уравнение регрессии данной модели согласно (5) имеет вид:  $\lg y_i = b_1 \cdot \lg x_i + b_0$ .

Для сравнения построенных линейных регрессионных моделей используем критерий  $R^2$ , найденный по формуле (6): для модели на основе нормализации Джонсона  $R^2=0,7676$ , для модели на основе логарифмического преобразования  $R^2=0,7407$ . Полученные значения говорят об адекватности построенных моделей, однако нормализующее преобразование Джонсона дает лучший результат.

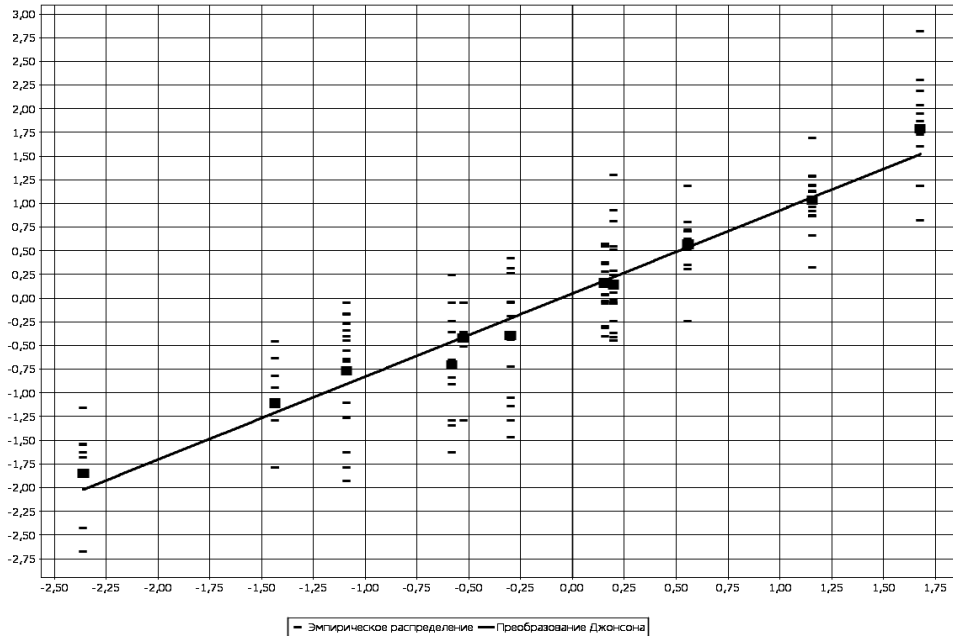


Рисунок 3 – Линейная регрессионная модель нормализованного времени восстановления работоспособности устройств терминальной сети

В соответствии с (7), построим нелинейную регрессионную модель времени восстановления работоспособности устройств терминальной сети, которая приведена на рис. 4.

Сравним построенную нелинейную регрессионную модель с моделью на основе логарифмического преоб-

разования, для построения которой использовался десятичный логарифм. Уравнение регрессии данной модели имеет вид:  $y = c \cdot x^{b_1}$ , где  $c = 10^{b_0}$ . Данная модель также приведена на рис. 4.

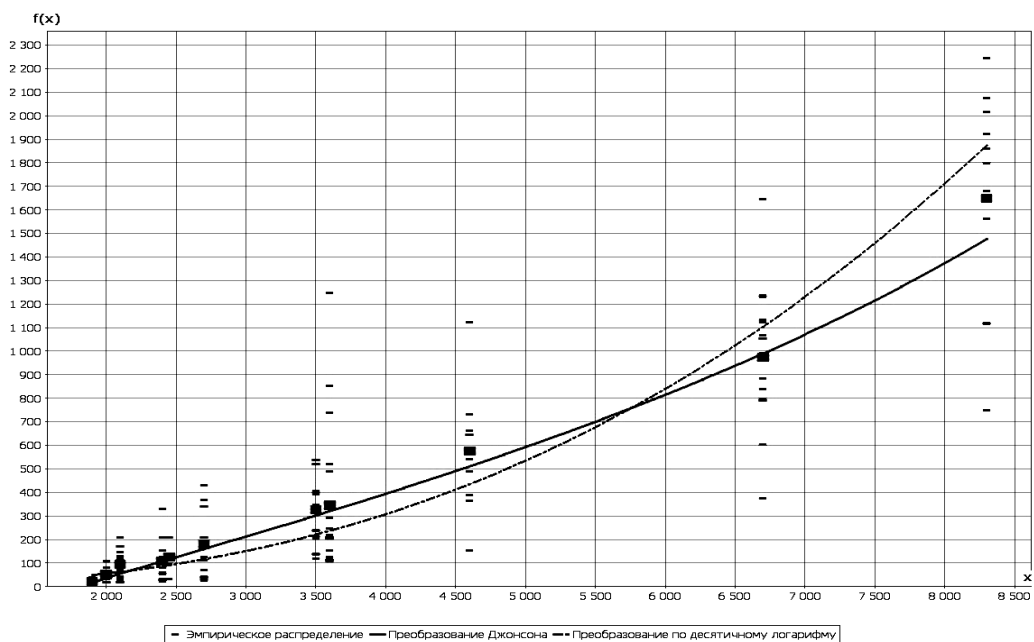
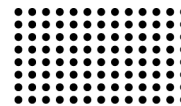
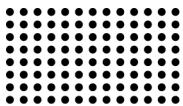


Рисунок 4 – Нелинейные регрессионные модели времени восстановления работоспособности устройств терминальной сети



Для сравнения построенных нелинейных регрессионных моделей используем такие критерии, как  $R^2$ , найденный по формуле (6), и сумму квадратов отклонений  $S_y$ , найденную по формуле (8). Для модели на основе нормализации Джонсона:  $R^2=0,8082$ ,  $S_y=6,9180 \cdot 10^6$ . Для модели на основе логарифмического преобразования  $R^2=0,7759$ ,  $S_y=8,0851 \cdot 10^6$ . Анализ критерия  $R^2$  говорит об адекватности построенных моделей, однако нормализующее преобразование Джонсона дает лучший результат. Из анализа критерия  $S_y$  следует, что нелинейная регрессионная модель на основе преобразования Джонсона более адекватно описывает исходные эмпирические данные, чем нели-

нейная регрессионная модель на основе логарифмического преобразования.

### ВЫВОДЫ

Впервые построена нелинейная регрессионная модель времени восстановления работоспособности устройств терминальной сети на основе нормализующего преобразования Джонсона. Для выполнения расчетов было доработано соответствующее программное обеспечение на языке программирования Java.

В дальнейшем планируется разработка информационной технологии автоматизированной системы обработки информации терминальной сети на основе построенной модели.

### ЛИТЕРАТУРА:

1. Nasha cel' – bank v shagovoyj dostupnosti [Elektronnyj resurs] / Rezhim dostupa: <http://www.inpas.ru/publications/78/>
2. Greshilov, A.A. Matematicheskie metody postroeniya prognozov [Tekst] / A.A. Greshilov, V.A. Stakun, A.A. Stakun – M.: Radio i svjaz', 1997. – 112 s.
3. Demidenko, E.Z. Lineijnaja i nelineijnaja regressii [Tekst] / E.Z. Demidenko – M.: Finansy i statistika, 1981. – 302 s.
4. Dreijper, N. Prikladnoj regressionnyj analiz: V 2-kh kn. Kn. 2 / Per. s angl. – 2-e izd., pererab. i dop. [Tekst] / N. Dreijper, G. Smit – M.: Finansy i statistika, 1987. – 351 s.
5. Ajjvazjan, S.A. Prikladnaja statistika. Osnovy ehkonometriki: Uchebnik dlja vuzov: V 2 t. 2-e izd., ispr. – T. 1: Teorija verojatnostej i prikladnaja statistika [Tekst] / S.A. Ajjvazjan, V.S. Mkhitarjan – M.: JuNITI-DANA, 2001. – 656 s.
6. Kobzar', A.I. Prikladnaja matematicheskaja statistika. Dlja inzhenerov i nauchnykh rabotnikov [Tekst] / A.I. Kobzar' – M.: FIZMATLIT, 2006. – 816 s.
7. Prikhod'ko, S.B. Metod pobudovi nelinejnykh rivnjan' regresii na osnovi normalizujuchikh peretvoren' [Tekst] / S.B. Prikhod'ko // Tezi dopovidej mizhderzh. nauk.-metodich. konf. "Problemi matematichnogo modeljuvannja" (Dniprodzerzhins'k, 13-15 chervnja 2012 r.). – Dniprodzerzhins'k: DDTU, 2012. – S.31-33.
8. Prikhod'ko, S.B. Rozrobka nelinejnoj regresijnoj modeli trivalosti programnykh proektiv na osnovi normalizujuchogo peretvoren'na Dzhonsona [Tekst] / S.B. Prikhod'ko, A.V. Pukhalevich // Radioelektronni i komp'juterni sistemi. – Kharkiv: "KhAI", 2012. – №4 (56). – S.90-93.
9. Kendall, M. Teorija raspredelenij [Tekst] / M. Kendall, A. St'juart – M.: Nauka, 1966. – 588 s.
10. Kovalenko, I.I. Suchasni metodi statistichnogo analizu danikh: Navchal'nij posibnik [Tekst] / I.I. Kovalenko, S.B. Prikhod'ko, L.O. Latans'ka. – Mikolaiv: NUK, 2011. – 192 s.
11. Prikhod'ko, S.B. Interval'ne ocinjuvannja statistichnykh momentiv negausiv's'kikh vipadkovikh velichin na osnovi normalizujuchikh peretvoren' [Tekst] / S.B. Prikhod'ko // Matematichne modeljuvannja: naukovij zhurnal. – Dniprodzerzhins'k: DDTU, 2011. – №1 (24). – S.9-13.
12. Ventcel', E.S. Teorija verojatnostej: Ucheb. dlja vuzov [Tekst] / E.S. Ventcel'. – M.: Vyssh. shk., 1999. – 576 s.
13. Magnus, Ja.R. Ehkonometrika. Nachal'nyj kurs: Ucheb. – 6-e izd., pererab. i dop. [Tekst] / Ja.R. Magnus, P.K. Katyshev, A.A. Pereseckij. – M.: Delo, 2004. – 576 s.
14. Prikhod'ko, S.B. Vybore analiticheskij modeli zakona raspredelenija vremeni narabotki mezhdju otkazami ustrojstv terminal'noj seti [Tekst] / S.B. Prikhod'ko, L.N. Makarova // Naukovi praci: naukovo-metodichnij zhurnal. – Vip. 179. T. 191. Komp'juterni tekhnologii. – Mikolaiv: Vid-vo ChDU im. Petra Mogili, 2012 – S.42-45.
15. Prikhod'ko, S.B. Opredelenie doveritel'nykh intervalov statisticheskikh momentov vremeni narabotki mezhdju otkazami ustrojstv terminal'noj seti [Tekst] / S.B. Prikhod'ko, L.N. Makarova // Naukovi praci: naukovo-metodichnij zhurnal. – Vip. 201. T. 213. Komp'juterni tekhnologii. – Mikolaiv: Vid-vo ChDU im. Petra Mogili, 2013. – S.82-86.