doi:https://doi.org/10.15407/emodel.40.06.101

UDC 004.412:519.237.5

N.V. Prykhodko, PhD, Economics, **S.B. Prykhodko**, Dr.Sci.Tech.
Admiral Makarov National University of Shipbuilding
(9, Heroes of Ukraine Ave., 54025, Mykolaiv, Ukraine
Tel.: (0512) 424280, e-mail: university@nuos.edu.ua)

Constructing the Nonlinear Regression Models on the Basis of Multivariate Normalizing Transformations

The techniques to build the models, equations, confidence and prediction intervals of nonlinear regressions on the basis of multivariate normalizing transformations for non-Gaussian data are considered. The examples of application of the techniques for the four-dimensional non-Gaussian data set for two cases such as: univariate and multivariate normalizing transformations are given. The values of the multiple coefficient of determination such as: the mean magnitude of relative error and the percentage of prediction which are given are better for the nonlinear regression model for the Johnson multivariate transformation compared to the univariate one. The widths of the prediction interval of non-linear regression on the basis of the Johnson multivariate transformation are less than following Johnson univariate transformation for 26 of 30 rows of data. Approximately the same results are obtained for confidence intervals of nonlinear regression. In general, when constructing the models, equations, confidence and prediction intervals of non-linear regressions for multivariate non-Gaussian data, one should use multivariate normali- zing transformations. Normalizing data with univariate transformations instead of multivariate one may lead to increasing of width of the confidence and prediction intervals of non-linear regression.

K e y w o r d s: non-linear regression model, prediction interval, normalizing transformation, multivariate non-Gaussian data.

Introduction. A normalizing transformation is often a good way to build models, equations, confidence and prediction intervals of nonlinear regressions [1—7]. However, well-known techniques for building models, equations, confidence and prediction intervals of non-linear regressions are based on the univariate normalizing transformations (such as log and Box-Cox transformations) which do not take into account the correlation between random variables in the case of normalization of multivariate non-Gaussian data. This leads to the need of using the multivariate normalizing transformations, which take into account the correlation to build models, equations, confidence and prediction intervals of nonlinear regression.

In this paper we demonstrate that there may be data sets for which the results of building the models such as confidence and prediction intervals of non-linear regressions depend on which normalizing transformation is applied, univariate or multivariate. We consider the techniques to build the models, confidence and prediction intervals of non-linear regression for multivariate non-Gaussian data. As and in [5] the techniques consist of three steps. In the first step a set of multivariate non-Gaussian data is normalized using a multivariate normalizing transformation. In the second step, the models, confidence and prediction intervals of linear regression for the normalized data are built. In the third step, the models, confidence and prediction intervals of non-linear regressions for multivariate non-Gaussian data on the basis of the models, confidence and prediction intervals of linear regression for the normalized data and the normalizing transformation are constructed.

Nonlinear regression model. In reference [2] authors define a nonlinear regression model as «a model for the relationship between a response and predictor(s) in which at least one parameter does not enter linearly into the model». According to [1—3, 6—8] the general nonlinear regression model may be represented as

$$Y = f(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon, \tag{1}$$

where f is a nonlinear function; $\mathbf{x}$ is a vector of regressors (independent variables); $\boldsymbol{\theta}$ is a vector of parameters; ε is the error term that has the same properties as in linear regression, i.e. the Gaussian random variable which defines residuals, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

We have additive error term in the model (1). According to [7, 8] the nonlinear regression model with multiplicative error term may be represented as

$$Y = f(\mathbf{x}, \boldsymbol{\theta})\, \varepsilon, \tag{2}$$

where $\varepsilon \sim N(1, \sigma_\varepsilon^2)$.

Duncan [8] considered two models for the error structure: the additive form (1) and the multiplicative form (2) and simulated three distributions for ε — the normal, the scale-contaminated normal, and the double exponential. In practice we will generally not know the form of the error term, but an additive error term undoubtedly is more common than multiplicative one. Bates and Watts [1] pointed out that the assumption of the additivity of the error is closely tied to the assumption of constant variance of the disturbances (residuals). It may be the case that the residuals can be considered as having constant variance, but as entering the model multiplicatively. In either case, one of the corrective actions is to take a transformation of the response (dependent variable). If the error is multiplicative, we can treat the nonlinear regression model as intrinsically linear and use the normalizing transformation [4]. That is, we define the nonlinear regression model as $Y = f(\mathbf{x}, \boldsymbol{\theta}, \varepsilon)$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

The techniques. Consider bijective multivariate normalizing transformation of non-Gaussian random vector $\mathbf{P} = \{Y, X_1, X_2, ..., X_k\}^T$ to Gaussian random vector $\mathbf{T} = \{Z_Y, Z_1, Z_2, ..., Z_k\}^T$ is given by

$$\mathbf{T} = \boldsymbol{\psi}(\mathbf{P}) \tag{3}$$

and the inverse transformation for (3)

$$\mathbf{P} = \boldsymbol{\psi}^{-1}(\mathbf{T}). \tag{4}$$

The linear regression model for normalized data according to (3) will have the form [6]

$$Z_Y = \hat{Z}_Y + \varepsilon = \overline{Z}_Y + (\mathbf{Z}_X^+)\,\hat{\mathbf{b}} + \varepsilon, \tag{5}$$

where $\hat{Z}_Y$ is prediction linear regression equation result for values of components of vector $\mathbf{z}_X = \{Z_1, Z_2, ..., Z_k\}$; $\mathbf{Z}_X^+$ is the matrix of centered regressors that contains the values $Z_{1_i} - \overline{Z}_1$, $Z_{2_i} - \overline{Z}_2$, ..., $Z_{k_i} - \overline{Z}_k$; $\hat{\mathbf{b}}$ is estimator for vector of linear regression equation parameters, $\mathbf{b} = \{b_1, b_2, ..., b_k\}^T$; ε is the Gaussian random variable which defines residuals, $\varepsilon \sim N(0,1)$.

The nonlinear regression model will have the form

$$Y = \psi_Y^{-1}[\overline{Z}_Y + (\mathbf{Z}_X^+)\,\hat{\mathbf{b}} + \varepsilon], \tag{6}$$

where ψ_Y is the first component of vector $\boldsymbol{\psi}$, $\boldsymbol{\psi} = \{\psi_Y, \psi_1, \psi_2, ..., \psi_k\}^T$. The technique to build a prediction interval of non-linear regression is based on a prediction interval of linear regression for normalized data, transformations (3) and (4):

$$\psi_Y^{-1}\left(\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_Y} \left\{1 + \frac{1}{N} + (\mathbf{z}_X^+)^T [(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+]^{-1} (\mathbf{z}_X^+)\right\}^{1/2}\right), \tag{7}$$

where $S_{Z_Y}^2 = \frac{1}{\nu}\sum_{i=1}^{N}(Z_{Y_i} - \hat{Z}_{Y_i})^2$, $\nu = N - k - 1$; $(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+$ is the $k \times k$ matrix

$$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ = \begin{pmatrix} S_{Z_1 Z_1} & S_{Z_1 Z_2} & \cdots & S_{Z_1 Z_k} \\ S_{Z_1 Z_2} & S_{Z_2 Z_2} & \cdots & S_{Z_2 Z_k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{Z_1 Z_k} & S_{Z_2 Z_k} & \cdots & S_{Z_k Z_k} \end{pmatrix},$$

where $S_{Z_q Z_r} = \sum_{i=1}^{N} [Z_{q_i} - \overline{Z}_q][Z_{r_i} - \overline{Z}_r]$, $q, r = 1, 2, ..., k$.

A confidence interval of nonlinear regression is defined like (7) with the only difference that in the sum in curly brackets (7) there will not be 1:

$$\psi_Y^{-1}\left(\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_r} \left\{\frac{1}{N} + (\mathbf{z}_X^+)^T [(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+]^{-1} (\mathbf{z}_X^+)\right\}^{1/2}\right), \quad (8)$$

Examples. We consider the examples of building the models, equations, confidence and prediction intervals of nonlinear regressions for multivariate non-Gaussian data for two cases: univariate and multivariate normalizing transformations.

Table 1 contains the data on metrics of software for open-source Java-based system [9, 10]. Recall that the first metric Y involves actual software size in the thousand lines of code, the second X_1, third X_2 and fourth X_3 metrics determine respectively the total number of classes, the total number of relationships and the average number of attributes per class in conceptual data model. Table 1 also contains the lower bounds (LB) and upper bounds (UB) of the prediction intervals of nonlinear regressions, which calculation is considered below.

For normalizing the multivariate non-Gaussian data from Table 1, we use the Johnson univariate and multivariate transformations (the Johnson translation system) for S_B family. In our case the Johnson normalizing translation is given by [11]:

$$\mathbf{T} = \gamma + \boldsymbol{\eta}\mathbf{h}\,[\boldsymbol{\lambda}^{-1}(\mathbf{P} - \varphi)] \sim N_m(\mathbf{0}_m, \boldsymbol{\Sigma}), \quad (9)$$

where $\mathbf{h}\,[(y_Y, y_1, ..., y_k)] = \{h_Y(y_Y), h_1(y_1), ..., h_k(y_k)\}^T$; $h_i(\cdot)$ is one of the translation functions

$$h = \begin{cases} \ln(y), & \text{for } S_L \text{ (log normal) family;} \\ \ln[y/(1-y)], & \text{for } S_B \text{ (bounded) family;} \\ \text{Arsh}(y), & \text{for } S_U \text{ (unbounded) family;} \\ y, & \text{for } S_N \text{ (normal) family.} \end{cases} \quad (10)$$

Here $y = (X - \varphi)/\lambda$; $\text{Arsh}(y) = \ln(y + \sqrt{y^2 + 1})$. In our case X equals Y, X_1, X_2 or X_3 respectively.

Parameters of Johnson univariate and multivariate transformations for S_B family were estimated by the maximum likelihood method. Estimators for parameters of the univariate transformation (10) for metric Y are: $\hat{\gamma}_Y = 0{,}46387$; $\hat{\eta}_Y = 0{,}50326$; $\hat{\varphi}_Y = 2{,}817$; $\hat{\lambda}_Y = 89{,}930$. Estimators for parameters of the multivariate transformation (9) for metric Y are: $\hat{\gamma}_Y = 9{,}6309$; $\hat{\eta}_Y = 1{,}05243$; $\hat{\varphi}_Y = -1{,}4568$; $\hat{\lambda}_Y = 153102{,}6$. Estimators for other parameters of the Johnson univariate and

Table 1. **Metrics of software and the prediction intervals of software size regressions**

No	Y	X_1	X_2	X_3	The bounds of the prediction intervals			
					Univariate		Multivariate	
					LB	UB	LB	UB
1	11,717	8	6	4,25	5,679	22,412	7,536	19,277
2	47,52	23	19	9,565	32,573	75,467	36,600	68,504
3	84,01	26	40	11,462	77,736	91,114	73,863	96,264
4	26,999	15	14	8,933	17,436	58,472	21,814	48,847
5	41,72	20	15	5,9	20,726	63,360	25,763	54,772
6	13,015	5	6	12,4	5,559	24,105	7,061	19,398
7	30,402	18	7	6,611	11,701	46,283	14,924	36,873
8	29,159	23	10	6,957	18,875	61,455	22,764	51,145
9	53,443	28	25	4,179	40,607	80,655	45,398	77,523
10	18,694	13	9	6,615	10,693	42,556	14,102	34,502
11	26,384	16	6	5,125	9,052	37,791	11,900	30,337
12	38,721	19	16	6,579	21,431	64,236	26,458	55,673
13	75,643	26	30	6,154	49,028	83,854	52,398	82,637
14	46,72	21	24	6,048	32,094	74,958	37,508	69,345
15	6,413	7	5	4,143	4,920	18,108	6,290	15,899
16	79,534	20	37	4,85	51,026	85,596	52,840	84,458
17	36,343	18	17	5,333	20,068	62,454	25,536	54,455
18	59,684	22	31	6,182	43,924	81,998	48,094	79,560
19	50,454	15	20	11,6	25,246	70,052	29,796	61,411
20	3,055	4	1	7	3,009	4,748	2,516	4,443
21	63,257	34	17	3,971	45,506	84,384	44,137	78,697
22	91,28	35	28	13,571	78,586	91,334	74,127	97,180
23	32,707	11	17	7,545	15,088	54,334	19,782	45,949
24	11	5	5	3,6	4,077	13,054	4,988	12,419
25	5,543	6	4	3,833	4,203	13,643	5,062	12,421
26	22,686	12	11	6,667	11,471	44,858	15,245	36,852
27	3,911	3	2	6,667	3,048	5,105	2,834	5,520
28	20,841	14	7	3	6,446	29,640	9,577	26,119
29	9,269	6	5	3,5	4,360	14,963	5,483	13,826
30	7,732	7	2	11,143	4,206	14,773	4,601	11,617

Table 2. **Estimators for parameters**

j	Univariate transformation				Multivariate transformation			
	$\hat{\gamma}_j$	$\hat{\eta}_j$	$\hat{\varphi}_j$	$\hat{\lambda}_j$	$\hat{\gamma}_j$	$\hat{\eta}_j$	$\hat{\varphi}_j$	$\hat{\lambda}_j$
1	0,38093	0,62689	2,634	33,711	15,5355	1,58306	–1,8884	243051,0
2	0,60545	0,62215	0,700	41,428	25,4294	2,54714	–6,9746	311229,5
3	0,65592	0,72789	2,839	11,780	0,72801	0,54312	3,2925	13,90

multivariate transformations are in Table 2. The sample covariance matrix $\mathbf{S}_N$ of the $\mathbf{T}$ is used as the approximate moment-matching estimator of $\mathbf{\Sigma}$

$$\mathbf{S}_N = \begin{pmatrix} 1,0000 & 0,9514 & 0,9333 & 0,1574 \\ 0,9514 & 1,0000 & 0,9006 & 0,1345 \\ 0,9333 & 0,9006 & 1,0000 & 0,0554 \\ 0,1574 & 0,1345 & 0,0554 & 1,0000 \end{pmatrix}.$$

For detecting the outliers in the data from Table 1 we use the technique based on multivariate normalizing transformations and the squared Mahalanobis distance (MD) [12, 13]. There are no outliers in the data from Table 1 for 0,005 significance level and the Johnson multivariate transformation (9) for S_B family. In [9, 10] it was also assumed that the data contains no outliers. The values of squared MD for normalized data by the Johnson univariate transformation (10) for S_B family from Table 1 indicate the data of system 22 is multivariate outlier, since for this data row the squared MD equals to 17,73 is greater than the value of the quantile of the Chi-Square distribution, which equals to 14,86 for 0,005 significance level. Although note that without using normalization the data of system 11 is multivariate outlier since for this data row the squared MD equals to 15,44.

After normalizing the non-Gaussian data, the linear regression model (5) is built

$$Z_Y = \hat{Z}_Y + \varepsilon = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3 + \varepsilon. \quad (11)$$

Parameters of the linear regression model (11) were estimated by the least square method. Estimators for parameters of the model (11) for the Johnson univariate and multivariate transformation are: $\hat{b}_0 = 0$; $\hat{b}_1 = 0{,}46976$; $\hat{b}_2 = 0{,}53539$; $\hat{b}_3 = 0{,}11397$ and $\hat{b}_0 = 0$; $\hat{b}_1 = 0{,}56085$; $\hat{b}_2 = 0{,}42491$; $\hat{b}_3 = 0{,}05846$ respectively.

Next, the non-linear regression model (6) is built

$$Y = \hat{\varphi}_Y + \hat{\lambda}_Y [1 + e^{-(\hat{Z}_Y + \varepsilon - \hat{\gamma}_Y)/\hat{\eta}_Y}]^{-1}, \quad (12)$$

where

$$Z_j = \gamma_j + \eta_j \ln \frac{X_j - \varphi_j}{\varphi_j + \lambda_j - X_j}, \ \varphi_j < X_j < \varphi_j + \lambda_j, \ j = 1,2,3.$$

The model (12) becomes a nonlinear regression equation when there is no error term ε. A mean magnitude of relative error (MMRE) and percentage of prediction (PRED(0,25)) are accepted as standard evaluations of prediction results by regression models and equations. The acceptable values of MMRE and PRED(0,25) are not more than 0,25 and not less than 0,75 respectively. The acceptable value of multiple coefficient of determination R^2 is approximately the same as for PRED(0,25). The values of R^2, MMRE and PRED(0,25), which equal 0,9672, 0,1389 and 0,8667 respectively, are better for the model (12) for the Johnson multivariate transformation in comparison with univariate one, for which these values are 0,9574, 0,1579 and 0,8000 respectively.

Table 1 contains the lower LB and upper UB bounds of the prediction intervals of nonlinear regressions, which calculated by (7) on the basis of the Johnson univariate and multivariate transformations respectively for 0,05 significance level. Note, the widths of the prediction interval of non-linear regression on the basis of Johnson multivariate transformation are less than following the Johnson univariate transformation for 26 rows of data: 1, 2, 4-19, 21, 23-26, 28-30. Approximately the same results are obtained for confidence intervals, which calculated by (8).

Following [14] multivariate kurtosis β_2 is estimated for the data from Table 1 and the normalized data on the basis of Johnson univariate and multivariate transformations for S_B family. The estimator of multivariate kurtosis given by [14]:

$$\hat{\beta}_2 = \frac{1}{N} \sum_{i-1}^{N} \{(\mathbf{Z}_i - \overline{\mathbf{Z}})^T \mathbf{S}_N^{-1} (\mathbf{Z}_i - \overline{\mathbf{Z}})\}^2. \tag{13}$$

In our case, in the formula (13), the vectors $\mathbf{Z}$ and $\overline{\mathbf{Z}}$ should be replaced by the vectors $\mathbf{P}$ and $\overline{\mathbf{P}}$ or $\mathbf{T}$ and $\overline{\mathbf{T}}$, respectively, for the initial (non-Gaussian) or normalized data. It is known that $\beta_2 = m(m+2)$ holds under multivariate normality. The given equality is a necessary condition for multivariate normality. In our case $\beta_2 = 24$. The estimators of multivariate kurtosis equal 27,17, 32,05 and 24,02 for the data from Table 1, the normalized data on the basis of the Johnson univariate and multivariate transformations respectively. The values of these estimators indicate that the necessary condition for multivariate normality is practically performed for the normalized data on the basis of the Johnson multivariate transformation for family S_B, it does not hold for other data.

Conclusions

In general, when constructing the models, equations, confidence and prediction intervals of non-linear regressions for multivariate non-Gaussian data, one should use multivariate normalizing transformations. Normalizing data with univariate transformations instead of multivariate one may lead to increasing of width of the confidence and prediction intervals of non-linear regression.

REFERENCES

1. Bates, D.M. and Watts, D.G. (1988), Nonlinear Regression Analysis and Its Applications, 2nd edition, John Wiley & Sons, New York, USA, DOI: 10.1002/9780470316757.
2. Chatterjee, S. and Simonoff, J.S. (2013), Handbook of Regression Analysis, John Wiley & Sons, New York, USA.
3. Drapper, N.R. and Smith, H. (1998), Applied Regression Analysis, John Wiley & Sons, New York, USA.
4. Freund, Rudolf J., Wilson, William J. and Sa, Ping (2006), Regression analysis: statistical modeling of a response variable, 2nd edition, Elsevier Academic Press, Burlington, MA, London, GB.
5. Prykhodko, S.B. (2016), Developing the software defect prediction models using regression analysis based on normalizing transformations, *Modern Problems in Testing of the Applied Software: The Research and Practice Seminar* (PTTAS-2016), Poltava, Ukraine, May 25-26, 2016, pp. 6-7.
6. Ryan, T.P. (1997), Modern regression methods, John Wiley & Sons, New York, USA, DOI: 10.1002/9780470382806.
7. Seber, G.A.F. and Wild, C.J. (1989), Nonlinear Regression, John Wiley & Sons, New York, USA, DOI: 10.1002/0471725315.
8. Duncan, G.T. (1978), An empirical study of jackknife constructed confidence regions in nonlinear regression, Technometrics, Vol. 20, no. 2, pp. 123-129, DOI: 10.2307/1268703.
9. Tan, H.B.K., Zhao, Y. and Zhang, H. (2006), Estimating LOC for information systems from their conceptual data models, *Proceedings of the 28th International Conference on Software Engineering* (ICSE ’06), Shanghai, China, May 20-28, 2006, pp. 321-330, DOI: 10.1145/1134285.1134331.
10. Tan, H.B.K., Zhao, Y. and Zhang, H. (2009), Conceptual data model-based software size estimation for information systems, *Transactions on Software Engineering and Methodology*, Vol. 19, Issue 2, October 2009, Article no. 4, DOI: 10.1145/1571629.1571630.
11. Stanfield, P.M., Wilson, J.R., Mirka, G.A., Glasscock, N.F., Psihogios, J.P. and Davis, J.R. (1996), Multivariate input modeling with Johnson distributions, *Proceedings of the 28th Winter simulation conference* WSC’96, Coronado, CA, USA, December 8-11, 1996, ed. S. Andradyttir, K.J. Healy, D.H. Withers and B.L. Nelson, IEEE Computer Society Washington, DC, USA, pp. 1457-1464.
12. Prykhodko, S., Prykhodko, N., Makarova, L. and Pukhalevych, A. (2018), Application of the Squared Mahalanobis Distance for Detecting Outliers in Multivariate Non-Gaussian Data, *Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering* (TCSET), Lviv-Slavske, Ukraine, February 20-24, 2018, pp. 962-965, DOI: 10.1109/TCSET.2018.8336353.

13. Prykhodko, S., Prykhodko, N., Makarova, L. and Pugachenko, K. (2017), Detecting Outliers in Multivariate Non-Gaussian Data on the basis of Normalizing Transformations, *Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering* (UKRCON) «Celebrating 25 Years of IEEE Ukraine Section», May 29 -June 2, 2017, Kyiv, Ukraine, pp. 846-849, DOI: 10.1109/UKRCON.2017.8100366.
14. Mardia, K.V. (1970), Measures of multivariate skewness and kurtosis with applications, *Biometrika*, Vol. 57, Issue 3, pp. 519-530, DOI: 10.1093/biomet/57.3.519.

Received 17.09.18

Н.В. Приходько, С.Б. Приходько

ПОСТРОЕНИЕ НЕЛИНЕЙНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ
НА ОСНОВЕ МНОГОМЕРНЫХ НОРМАЛИЗИРУЮЩИХ ПРЕОБРАЗОВАНИЙ

Рассмотрены методы построения моделей, уравнений, доверительных интервалов и интервалов прогнозирования нелинейных регрессий на основе многомерных нормализирующих преобразований для негауссовых данных. Приведены примеры применения методов для набора четырехмерных негауссовых данных в двух случаях: одномерного и многомерного нормализирующих преобразований Джонсона. Значения множественного коэффициента детерминации, средней величины относительной ошибки и процента прогнозирования для нелинейной регрессионной модели при многомерном преобразовании Джонсона лучше по сравнению с одномерным. Ширина интервала предсказания нелинейной регрессии на основе многомерного преобразования Джонсона меньше, чем после одномерного преобразования Джонсона для 26 из 30 строк данных. Приблизительно такие же результаты получены для доверительных интервалов нелинейной регрессии. В общем случае при построении моделей, уравнений, доверительных интервалов и интервалов прогнозирования нелинейных регрессий для многомерных негауссовых данных следует использовать многомерные нормализирующие преобразования. Применение одномерных преобразований вместо многомерных для нормализации таких данных может приводить к увеличению ширины доверительных интервалов и интервалов предсказания нелинейной регрессии.

К л ю ч е в ы е с л о в а: нелинейная регрессионная модель, интервал прогнозирования, нормализующее преобразование, многомерные негауссовые данные.

Н.В. Приходько, С.Б. Приходько

ПОБУДОВА НЕЛІНІЙНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ
НА ОСНОВІ БАГАТОВИМІРНИХ НОРМАЛІЗУЮЧИХ ПЕРЕТВОРЕНЬ

Розглянуто методи побудови моделей, рівнянь, довірчих інтервалів і інтервалів прогнозування нелінійних регресій на основі багатовимірних нормалізуючих перетворень для негаусових даних. Наведено приклади застосування методів для набору чотиривимірних негаусових даних у двох випадках: одновимірного і багатовимірного нормалізуючих перетворень Джонсона. Значення множинного коефіцієнта детермінації, середньої величини відносної похибки і відсотка прогнозування для нелінійної регресійної моделі при багатовимірному перетворенні Джонсона краще в порівнянні з одномірними. Ширина інтервалу передбачення нелінійної регресії на основі багатовимірного перетворення Джонсона менше, ніж після одновимірного перетворення Джонсона для 26 з 30 рядків даних. Приблизно такі ж результати отримано для довірчих інтервалів нелінійної регресії.

У загальному випадку при побудові моделей, рівнянь, довірчих інтервалів і інтервалів прогнозування нелінійних регресій для багатовимірних негаусових даних слід використовувати багатовимірні нормалізуючі перетворення. Застосування одновимірних перетворень замість багатовимірних для нормалізації таких даних може призводити до збільшення ширини довірчих інтервалів і інтервалів передбачення нелінійної регресії.

К л ю ч о в і с л о в а: нелінійна регресійна модель, інтервал прогнозування, нормалізуюче перетворення, багатовимірні негаусові дані.

PRYKHODKO Natalia Vasilivna, PhD (economics), Assistant Professor at the Finance Department of Admiral Makarov National University of Shipbuilding, Ministry of Education and Science of Ukraine, graduated from the Nikolayev Shipbuilding Institute named after Admiral S.O. Makarov in 1986, received the PhD degree from the St. Petersburg State Academy of Economics & Engineering in 1993, nostrified her PhD Diploma in Kharkiv State University of Economics in 1999. Sphere of scientific research: modeling of stochastic systems with the use of statistical approach, data processing, nonlinear regression analysis.

PRYKHODKO Sergiy Borisovich, Doctor of Sciences (D.Sc.) in Mathematical Modeling and Computational Methods, Professor and Head of the Department of Software of Automated Systems of Admiral Makarov National University of Shipbuilding, Ministry of Education and Science of Ukraine, graduated from the Nikolayev Shipbuilding Institute named after Admiral S.O. Makarov in 1981, received the D.Sc. degree from the V.M. Glushkov Institute of Cybernetics of NAS of Ukraine in 2013. Sphere of scientific research: modeling, identification and simulation of stochastic systems, data and signal processing, nonlinear regression analysis, signal recognition, empirical software engineering.