UDC 004.412:519.237

S.B. PRYKHODKO, N.V. PRYKHODKO, A.V. MANDRA, A.Y. YAREMA
Admiral Makarov National University of Shipbuilding

THE NON-LINEAR REGRESSION EQUATION TO ESTIMATE THE SOFTWARE SIZE OF VB-BASED INFORMATION SYSTEMS

The equation, confidence and prediction intervals of multivariate non-linear regression for estimating the software size of VB-based information systems are constructed on the basis of the Johnson multivariate normalizing transformation. Comparison of the constructed equation with the linear and two non-linear regression equations based on the decimal logarithm and the Johnson univariate transformation is performed.

Keywords: non-linear regression equation, confidence interval, prediction interval, multivariate normalizing transformation, multivariate non-Gaussian data.

С.Б. ПРИХОДЬКО, Н.В. ПРИХОДЬКО, А.В. МАНДРА, А.Я. ЯРЕМА
Національний університет кораблебудування імені адмірала Макарова

НЕЛІНІЙНЕ РЕГРЕСІЙНЕ РІВНЯННЯ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНИХ СИСТЕМ НА VB

Рівняння, довірчі інтервали і інтервали прогнозування багатовимірної нелінійної регресії для оцінювання розміру програмного забезпечення інформаційних систем на VB побудовані на основі багатовимірного нормалізуючого перетворення Джонсона. Виконано порівняння побудованого рівняння з лінійним та двома нелінійними регресійними рівняннями, що базуються на десятковому логарифмі і одновимірному перетворенні Джонсона.

Ключові слова: нелінійне рівняння регресії, довірчий інтервал, інтервал передбачення, багатовимірне нормалізуюче перетворення, багатовимірні негаусовські дані.

С.Б. ПРИХОДЬКО, Н.В. ПРИХОДЬКО, А.В. МАНДРА, А.Я. ЯРЕМА
Национальный университет кораблестроения имени адмирала Макарова

НЕЛИНЕЙНОЕ РЕГРЕССИОННОЕ УРАВНЕНИЕ ДЛЯ ОЦЕНКИ РАЗМЕРА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННЫХ СИСТЕМ НА VB

Уравнение, доверительные интервалы и интервалы прогнозирования многомерной нелинейной регрессии для оценки размера программного обеспечения информационных систем на VB построены на основе многомерного нормализующего преобразования Джонсона. Выполнено сравнение построенного уравнения с линейным и двумя нелинейными регрессионными уравнениями, которые базируются на десятичном логарифме и одномерном преобразовании Джонсона.

Ключевые слова: нелинейное уравнение регрессии, доверительный интервал, интервал предсказания, многомерное нормализующее преобразование, многомерные негауссовские данные.

Problem formulation

VB (Visual Basic) is one of the programming languages commonly used in information systems. Software size is one of the most important internal metrics of software including software of information systems. The information obtained from estimating the software size are useful for predicting the software development effort by such well-known model as COCOMO II. This leads to the need to develop appropriate models for estimating the software size [1]. The paper [1] proposed the linear regression equations for estimating the software size of some programming languages, including VB, for information systems. The proposed equations are constructed by multiple linear regression analysis on the basis of the metrics that can be measured from class diagram. However, there are four basic assumptions that justify the use of linear regression equations, one of which is normality of the error distribution. But this assumption is valid only in particular cases. This leads to the need to use the non-linear regression equations including for estimating the software size of VB-based information systems.

Analysis of recent research and publications

A normalizing transformation is often a good way to construct the equations, confidence and prediction intervals of multiply non-linear regressions [2-8]. According [3] transformations are made for essentially four purposes, two of which are: first, to obtain approximate normality for the distribution of the error term (residuals) or the dependent random variable, second, to transform the response and/or the predictor in such a way that the strength of the linear relationship between new variables (normalized variables) is batter than the linear relationship between dependent and independent random variables. Now well-known normalizing transformations are used to

construct the equations, confidence and prediction intervals of non-linear regressions. However, known techniques for constructing equations, confidence and prediction intervals of multivariate non-linear regressions are based on the univariate normalizing 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 to use the multivariate normalizing transformations.

Purpose of the study

The purpose of the study is to construct the non-linear regression equation for estimating the software size of VB-based information systems. The software size prediction results by constructed equation should be better in comparison with other regression equations, both linear and nonlinear, primarily on such standard evaluations as the multiple coefficient of determination and mean magnitude of relative error.

Presentation of the main research material

We build the equation, confidence and prediction intervals of multivariate non-linear regression for estimating the software size of VB-based systems on the basis of the Johnson multivariate normalizing transformation (the Johnson normalizing translation) with the help of appropriate techniques proposed in [8]. As and in [8] the techniques consist of three steps. In the first step, a set of multivariate non-Gaussian data is normalized using a bijective multivariate normalizing transformation. In the second step, the equation, confidence and prediction intervals of linear regression for the normalized data are built. In the third step, the equations, confidence and prediction intervals of non-linear regressions for multivariate non-Gaussian data are constructed on the basis of the equation, confidence and prediction intervals of linear regression for the normalized data and normalizing transformation.

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

$$\mathbf{T} = \boldsymbol{\psi}(\mathbf{P}) \tag{1}$$

and the inverse transformation for (1)

$$\mathbf{P} = \boldsymbol{\psi}^{-1}(\mathbf{T}). \tag{2}$$

The linear regression equation for normalized data according to (1) will have the form [3]

$$\hat{Z}_Y = \overline{Z}_Y + \left(\mathbf{Z}_X^+\right)\hat{\mathbf{b}}, \tag{3}$$

where $\hat{Z}_Y$ is prediction linear regression equation result for values of $\mathbf{z}_X = \{Z_1, Z_2, \ldots, Z_k\}$; $\mathbf{Z}_X^+$ 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, $\mathbf{b} = \{b_1, b_2, \ldots, b_k\}^T$.

The non-linear regression equation will have the form

$$\hat{Y} = \psi_Y^{-1}\left[\overline{Z}_Y + \left(\mathbf{Z}_X^+\right)\hat{\mathbf{b}}\right], \tag{4}$$

where ψ_Y is the first component of vector $\boldsymbol{\psi} = \{\psi_Y, \psi_1, \psi_2, \ldots, \psi_k\}^T$.

The technique to build a confidence interval of non-linear regression is based on transformations (1) and (2), equation (3) and a confidence interval of linear regression for normalized data

$$\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_Y} \left\{\frac{1}{N} + \left(\mathbf{z}_X^+\right)^T \left[\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+\right]^{-1} \left(\mathbf{z}_X^+\right)\right\}^{1/2}. \tag{5}$$

where $t_{\alpha/2,\nu}$ is a quantile of student's *t*-distribution with ν degrees of freedom and $\alpha/2$ significance level; $\left(\mathbf{z}_X^+\right)^T$ is one of the rows of $\mathbf{Z}_X^+$; $S_{Z_Y}^2 = \frac{1}{\nu}\sum_{i=1}^{N}\left(Z_{Y_i} - \hat{Z}_{Y_i}\right)^2$, $\nu = N - k - 1$; $\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+$ is the $k \times k$ matrix

$$\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+ = \begin{pmatrix} S_{Z_1Z_1} & S_{Z_1Z_2} & \dots & S_{Z_1Z_k} \\ S_{Z_1Z_2} & S_{Z_2Z_2} & \dots & S_{Z_2Z_k} \\ \dots & \dots & \dots & \dots \\ S_{Z_1Z_k} & S_{Z_2Z_k} & \dots & S_{Z_kZ_k} \end{pmatrix},$$

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

The confidence interval for non-linear regression will have the form

$$\psi_Y^{-1}\left(\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_Y} \left\{ \frac{1}{N} + \left(\mathbf{z}_X^+\right)^T \left[\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+ \right]^{-1} \left(\mathbf{z}_X^+\right) \right\}^{1/2} \right). \quad (6)$$

The technique to build a prediction interval is based on multivariate transformation (1), the inverse transformation (2), linear regression equation for normalized data (3) and a prediction interval for normalized data

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

The prediction interval for non-linear regression will have the form

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

The equations (4), formulas (6) and (8) are used for constructing the equations, confidence and prediction intervals of non-linear regressions for multivariate non-Gaussian data.

The Johnson normalizing translation. For normalizing the multivariate non-Gaussian data, we use the Johnson translation system. In our case the Johnson normalizing translation is given by

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

where $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$, φ and $\boldsymbol{\lambda}$ are parameters of the Johnson normalizing translation; $\boldsymbol{\gamma} = \left(\gamma_Y, \gamma_1, \gamma_2, \dots, \gamma_k\right)^T$; $\boldsymbol{\eta} = diag\left(\eta_Y, \eta_1, \eta_2, \dots, \eta_k\right)$; $\boldsymbol{\lambda} = diag\left(\lambda_Y, \lambda_1, \lambda_2, \dots, \lambda_k\right)$; $\varphi = \left(\varphi_Y, \varphi_1, \varphi_2, \dots, \varphi_k\right)^T$; $\mathbf{h}\left[\left(y_Y, y_1, \dots, y_k\right)\right] = \left\{h_Y\left(y_Y\right), h_1\left(y_1\right), \dots, h_k\left(y_k\right)\right\}^T$; $h_i(\cdot)$ is one of the translation functions

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

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

The equation, confidence and prediction intervals of non-linear regression to estimate the software size of VB-based information systems. The equation, confidence and prediction intervals of non-linear regression to estimate the software size of VB-based information systems are constructed on the basis of the Johnson multivariate

normalizing transformation for the four-dimensional non-Gaussian data set: actual software size in the thousand lines of code (KLOC) *Y*, the total number of classes X_1, the total number of relationships X_2 and the average number of attributes per class X_3, in conceptual data model from 32 information systems developed using the VB programming language with SQL. Table I contains the data from [1] on four metrics of software for 32 VB-based systems in the industry.

Table 1

The data and prediction result by regression equations for 32 VB-based systems

No	Y	X_1	X_2	X_3	Linear regression		Non-linear regression			
							univariate transformation		multivariate transformation	
					$\hat{Y}$	RME	$\hat{Y}$	RME	$\hat{Y}$	RME
1	37.54	27	8	8	37.033	0.0135	22.907	0.3898	31.405	0.1634
2	14.723	8	6	25.375	14.083	0.0435	14.891	0.0114	14.810	0.0059
3	24.667	12	10	16.917	21.452	0.1304	20.952	0.1506	22.070	0.1053
4	42.1	19	25	16.526	47.831	0.1361	47.723	0.1336	44.179	0.0494
5	87.23	35	38	8.343	81.900	0.0611	87.925	0.0080	86.198	0.0118
6	31.445	14	21	6.214	33.256	0.0576	33.644	0.0699	31.760	0.0100
7	67.04	35	27	16.829	72.155	0.0763	73.476	0.0960	75.467	0.1257
8	30.79	17	20	6.176	36.239	0.1770	35.208	0.1435	34.930	0.1345
9	22.402	13	14	5.769	23.864	0.0652	24.419	0.0900	25.200	0.1249
10	69.713	28	31	9.571	64.787	0.0707	67.247	0.0354	64.601	0.0733
11	16.17	6	9	27.333	15.321	0.0525	15.682	0.0302	16.168	0.0001
12	90.854	37	39	21.27	89.791	0.0117	90.836	0.0002	90.824	0.0003
13	64.35	27	33	5.481	64.393	0.0007	67.280	0.0455	63.865	0.0075
14	27.076	13	16	8	26.808	0.0099	27.530	0.0167	27.254	0.0066
15	20.933	10	10	6.5	15.467	0.2611	19.159	0.0847	19.812	0.0535
16	40.341	22	20	5.818	42.996	0.0658	39.299	0.0258	41.355	0.0251
17	37.54	27	8	8	37.033	0.0135	22.907	0.3898	31.405	0.1634
18	14.723	8	6	25.375	14.083	0.0435	14.891	0.0114	14.810	0.0059
19	24.667	12	10	16.917	21.452	0.1304	20.952	0.1506	22.070	0.1053
20	42.1	19	25	16.526	47.831	0.1361	47.723	0.1336	44.179	0.0494
21	87.23	35	38	8.343	81.900	0.0611	87.925	0.0080	86.198	0.0118
22	31.445	14	21	6.214	33.256	0.0576	33.644	0.0699	31.760	0.0100
23	67.04	35	27	16.829	72.155	0.0763	73.476	0.0960	75.467	0.1257
24	30.79	17	20	6.176	36.239	0.1770	35.208	0.1435	34.930	0.1345
25	22.402	13	14	5.769	23.864	0.0652	24.419	0.0900	25.200	0.1249
26	69.713	28	31	9.571	64.787	0.0707	67.247	0.0354	64.601	0.0733
27	16.17	6	9	27.333	15.321	0.0525	15.682	0.0302	16.168	0.0001
28	90.854	37	39	21.27	89.791	0.0117	90.836	0.0002	90.824	0.0003
29	64.35	27	33	5.481	64.393	0.0007	67.280	0.0455	63.865	0.0075
30	27.076	13	16	8	26.808	0.0099	27.530	0.0167	27.254	0.0066
31	20.933	10	10	6.5	15.467	0.2611	19.159	0.0847	19.812	0.0535
32	40.341	22	20	5.818	42.996	0.0658	39.299	0.0258	41.355	0.0251

For detecting the outliers in the data from Table 1 we use the technique based on multivariate normalizing transformations and the squared Mahalanobis distance [9]. There are no outliers in the data from Table I for 0.005 significance level and the Johnson multivariate transformation (9) for S_B family.

Parameters of the multivariate transformation (9) for S_B family were estimated by the maximum likelihood method. Estimators for parameters of the transformation (9) are: $\hat{\gamma}_Y = 0.3176$, $\hat{\gamma}_1 = 0.112778$, $\hat{\gamma}_2 = 0.11510$, $\hat{\gamma}_3 = 0.498667$, $\hat{\eta}_Y = 0.46211$, $\hat{\eta}_1 = 0.581163$, $\hat{\eta}_2 = 0.386574$, $\hat{\eta}_3 = 0.37802$, $\hat{\varphi}_Y = 13.9525$, $\hat{\varphi}_1 = 4.9519$, $\hat{\varphi}_2 = 5.90$, $\hat{\varphi}_3 = 5.381$, $\hat{\lambda}_Y = 77.7976$, $\hat{\lambda}_1 = 33.3568$, $\hat{\lambda}_2 = 33.20$ and $\hat{\lambda}_3 = 22.052$. The sample covariance matrix S_N of the $\mathbf{T}$ is used as the approximate moment-matching estimator of $\mathbf{\Sigma}$

$$S_N = \begin{pmatrix} 1.0000 & 0.9389 & 0.9354 & -0.2069 \\ 0.9389 & 1.0000 & 0.7858 & -0.2495 \\ 0.9354 & 0.7858 & 1.0000 & -0.2036 \\ -0.2069 & -0.2495 & -0.2036 & 1.0000 \end{pmatrix}.$$

After normalizing the non-Gaussian data by the multivariate transformation (9) for S_B family the linear regression equation (3) is built for normalized data

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

Parameters of the linear regression equation (11) were estimated by the least square method. Estimators for parameters of the equation (11) are: $\hat{b}_0 = -1.07 \cdot 10^{-4}$, $\hat{b}_1 = 0.52447$, $\hat{b}_2 = 0.54906$, $\hat{b}_3 = 0.03614$.

After that the non-linear regression equation (4) is built

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

where $\hat{Z}_Y$ is prediction result by the equation (11), $Z_j = \gamma_j + \eta_j \ln\left[\left(X_j - \varphi_j\right)/\left(\varphi_j + \lambda_j - X_j\right)\right]$, $\varphi_j < X_j < \varphi_j + \lambda_j$, $j = 1,2,3$.

The prediction results by equation (12) for values of components of vector $\mathbf{X} = \{X_1, X_2, X_3\}$ from Table 1 and values of magnitude of relative error MRE are shown in the Table I for two cases: the Johnson univariate and multivariate normalizing transformations. Table 1 also contains the prediction results by linear regression equation for values of components of vector **X** from Table 1 and MRE values.

For univariate normalizing transformations (10) of S_B family the estimators for parameters are such: $\hat{\gamma}_Y = 0.25082$, $\hat{\gamma}_1 = 0.067046$, $\hat{\gamma}_2 = 0.11745$, $\hat{\gamma}_3 = 0.505658$, $\hat{\eta}_Y = 0.354337$, $\hat{\eta}_1 = 0.403719$, $\hat{\eta}_2 = 0.393218$, $\hat{\eta}_3 = 0.38007$, $\hat{\varphi}_Y = 14.623$, $\hat{\varphi}_1 = 5.90$, $\hat{\varphi}_2 = 5.90$, $\hat{\varphi}_3 = 5.381$, $\hat{\lambda}_Y = 76.331$, $\hat{\lambda}_1 = 31.20$, $\hat{\lambda}_2 = 33.20$ and $\hat{\lambda}_3 = 22.052$. In the case of univariate normalizing transformations the estimators for parameters of the equation (11) are such: $\hat{b}_0 = 1.4 \cdot 10^{-5}$, $\hat{b}_1 = 0.38440$, $\hat{b}_2 = 0.66171$ and $\hat{b}_3 = 0.046502$.

Also the non-linear regression equation (4) is built by the decimal logarithm transformation

$$\hat{Y} = 10^{b_0} X_1^{b_1} X_2^{b_2} X_3^{b_3} . \quad (13)$$

where the estimators for parameters of the equation (13) are: $\hat{b}_0 = 0.09412$, $\hat{b}_1 = 0.715453$, $\hat{b}_2 = 0.374691$ and $\hat{b}_3 = 0.110754$.

The values of multiple coefficient of determination R^2, mean magnitude of relative error MMRE and percentage of prediction PRED(0.25) equal respectively 0.9789, 0.0771 and 0.9375 for linear regression equation, equal respectively 0.9698, 0.0736 and 1 for the equation (13), and equal respectively 0.9626, 0.0832 and 0.9375 for the equation (12) for the Johnson univariate transformation. The values of R^2, MMRE and PRED(0.25) are better for the equation (12) for the Johnson multivariate transformation, in comparison with linear regression equation, equation (13) and equation (12) with parameters for the Johnson univariate transformation, and are 0.9813, 0.0561 and 1. Only the values of PRED(0.25) are the same for both the equation (12) for the Johnson multivariate transformation and equation (13).

The confidence and prediction intervals of non-linear regression are defined by (6) and (8) respectively for the data from Table 1. The widths of the confidence interval of non-linear regression on the basis of the Johnson multivariate transformation are less than for linear regression for the twenty-two rows of data: 1-3, 5-7, 9, 11, 12, 14, 15, 17-19, 21-23, 25, 27, 28, 30 and 31. Also the widths of the confidence interval of non-linear regression on the basis of the Johnson multivariate transformation are less for more data rows than for non-linear regressions following the univariate transformations, both decimal logarithm and the Johnson. The widths of the confidence interval of non-linear regression on the basis of the Johnson multivariate transformation are less than following the

decimal logarithm univariate transformation for the twenty-two rows of data: 1-5, 7, 10-13, 15, 17-21, 23, 26-29 and 31. And ones are less than following the Johnson univariate transformation for the twenty-five rows of data: 3-11, 13-16, 19-27, 29-32. Approximately the same results are obtained for the prediction intervals of regressions.

Following [10] multivariate kurtosis β_2 is estimated for the data on metrics of software from Table 1 and the normalized data on the basis of the decimal logarithm transformation, the Johnson univariate and multivariate transformations for S_B family. 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 21.27, 22.03, 33.88 and 22.92 for the data from Table 1, the normalized data on the basis of the decimal logarithm transformation, 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 decimal logarithm transformation and the Johnson multivariate transformation, it does not hold for other data. Note that in our case, the poor normalization of multivariate non-Gaussian data using the Johnson univariate transformation leads to an increase in the widths of the confidence and prediction intervals of non-linear regression for a larger number of data rows compared to both the Johnson multivariate transformation and the decimal logarithm transformation.

Conclusions

The non-linear regression equation to estimate the software size of VB-based information systems is improved on the basis of the Johnson multivariate transformation for S_B family. This equation, in comparison with other regression equations (both linear and nonlinear), has a larger multiple coefficient of determination and a smaller value of MMRE.

When building the equations, confidence and prediction intervals of non-linear regressions for multivariate non-Gaussian data to estimate the software size of VB-based systems, one should use multivariate normalizing transformations.

Usually poor normalization of multivariate non-Gaussian data or application of univariate transformations instead of multivariate ones to normalize such data may lead to increase of width of the confidence and prediction intervals of regressions, both linear and non-linear, to estimate the software size of VB-based systems.

In the future, we intend to try other multivariate normalizing transformations and non-Gaussian data sets.

References

1. Hee Beng Kuan Tan, Yuan Zhao and Hongyu Zhang, "Estimating LOC for information systems from their conceptual data models", in Proceedings of the 28th International Conference on Software Engineering (ICSE '06), May 20-28, 2006, Shanghai, China, PP. 321-330.
2. 1. D.M. Bates, and D.G. Watts. Nonlinear Regression Analysis and Its Applications. Wiley, 1988, 384 p.
3. T.P. Ryan. Modern regression methods. Wiley, 1997, 529 p.
4. G.A.F. Seber, and C.J. Wild. Nonlinear Regression. John Wiley & Sons, Inc., 2003, 792 p.
5. R.A. Johnson, and D.W. Wichern. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007, 800 p.
6. Iain Pardoe. Applied regression modelling. Wiley, 2012, 325 p.
7. S Chatterjee, and J.S. Simonoff. Handbook of Regression Analysis. John Wiley & Sons, Inc., 2013, 236 p.
8. S.B. Prykhodko, "Developing the software defect prediction models using regression analysis based on normalizing transformations" in "Modern problems in testing of the applied software" (PTTAS-2016), Abstracts of the Research and Practice Seminar, Poltava, Ukraine, May 25-26, 2016, PP. 6-7.
9. S. Prykhodko, N. Prykhodko, L. Makarova, and K. Pugachenko, "Detecting Outliers in Multivariate Non-Gaussian Data on the basis of Normalizing Transformations", in 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, 2017, PP. 846-849.
10. K.V. Mardia, "Measures of multivariate skewness and kurtosis with applications", Biometrika, 57, PP. 519–530, 1970.