Estimating the Software Size of Open-Source PHP-Based Systems Using Non-Linear Regression Analysis

Sergiy Prykhodko[1], Natalia Prykhodko[2], Lidiia Makarova[1]

1. Department of Software of Automated Systems, Admiral Makarov National University of Shipbuilding, UKRAINE, Mykolaiv, Heroes of Ukraine ave. 9, email: sergiy.prykhodko@nuos.edu.ua

2. Finance Department, Admiral Makarov National University of Shipbuilding, UKRAINE, Mykolaiv, Heroes of Ukraine ave. 9, email: natalia.prykhodko@nuos.edu.ua

Abstract: **The equation, confidence and prediction intervals of multivariate non-linear regression for estimating the software size of open-source PHP-based systems are constructed on the basis of the Johnson multivariate normalizing transformation. Comparison of the constructed equation with the linear and non-linear regression equation based on the Johnson univariate transformation is performed.**

Keywords: **software size estimation, PHP-based system, multivariate non-linear regression analysis, normalizing transformation, non-Gaussian data.**

I. Introduction

Software size is one of the most important internal metrics of software. The information obtained from estimating the software size are useful for predicting the software development effort by such model as COCOMO II. The papers [1, 2] proposed the linear regression equations for estimating the software size of some programming languages, such as VBA, PHP, Java and C++. 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 models, 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 open-source PHP-based systems.

A normalizing transformation is often a good way to build the equations, confidence and prediction intervals of multiply non-linear regressions [3-5]. According [4] transformations are used for essentially four purposes, two of which are: first, to obtain approximate normality for the distribution of the error term (residuals), 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 better than the linear relationship between dependent and independent random variables. Well-known techniques for building the 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.

In this paper, we build the equation, confidence and prediction intervals of multivariate non-linear regression for estimating the software size of open-source PHP-based systems on the basis of the Johnson multivariate normalizing transformation (the Johnson normalizing translation) with the help of appropriate techniques proposed in [5].

II. The Techniques

The techniques to build the equations, confidence and prediction intervals of non-linear regressions are based on the multiple non-linear regression analysis using the multivariate normalizing transformations. A 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

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

and the inverse transformation for (1)

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

The linear regression equation for normalized data according to (1) will have the form [4]

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

where $\hat{Z}_Y$ is prediction linear regression equation result for values of components of vector $\mathbf{z}_X=\{Z_1, Z_2, \ldots, Z_k\}$; $\mathbf{Z}_X^+$ is the matrix of centered regressors that contains the values $Z_{1_i}-\bar{Z}_1$, $Z_{2_i}-\bar{Z}_2$, …, $Z_{k_i}-\bar{Z}_k$; $\hat{\mathbf{b}}$ 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

$$\hat{Y}=\psi_1^{-1}\left[\bar{Z}_Y+\left(\mathbf{Z}_X^+\right)\hat{\mathbf{b}}\right], \quad (4)$$

where $\hat{Y}$ is prediction non-linear regression equation result.

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

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

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

CEUR-WS.org/Vol-2300/Paper48.pdf

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 is built on the basis of the interval (5) and inverse transformation (2)

$$\psi_1^{-1}\left(\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_Y}\left\{\frac{1}{N} + \left(\mathbf{z}_X^+\right)^T\left[\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+\right]^{-1}\left(\mathbf{z}_X^+\right)\right\}^{1/2}\right). \quad (6)$$

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 is built on the basis of the interval (7) and inverse transformation (2)

$$\psi_1^{-1}\left(\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_Y}\left\{1 + \frac{1}{N} + \left(\mathbf{z}_X^+\right)^T\left[\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+\right]^{-1}\left(\mathbf{z}_X^+\right)\right\}^{1/2}\right). \quad (8)$$

III. THE JOHNSON NORMALIZING TRANSLATION

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

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

where $\boldsymbol{\Sigma}$ is the covariance matrix; $m = k + 1$; $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$, $\boldsymbol{\varphi}$ and $\boldsymbol{\lambda}$ are parameters of translation (9); $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_m)^T$; $\boldsymbol{\eta} = diag(\eta_1, \eta_2, \dots, \eta_m)$; $\boldsymbol{\lambda} = diag(\lambda_1, \lambda_2, \dots, \lambda_m)$; $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_m)^T$; $\mathbf{h}[(y_1, \dots, y_m)] = \{h_1(y_1), \dots, h_m(y_m)\}^T$; $h_i(\cdot)$ 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 $y = (x - \varphi)/\lambda$; $\text{Arsh}(y) = \ln\left(y + \sqrt{y^2 + 1}\right)$.

IV. THE EQUATION, CONFIDENCE AND PREDICTION INTERVALS OF NON-LINEAR REGRESSION TO ESTIMATE THE SOFTWARE SIZE

The equation, confidence and prediction intervals of non-linear regression to estimate the software size of open-source PHP-based 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 average number of attributes per class X_3, the total number of classes X_1 and the total number of relationships X_2 in conceptual data model from 32 information systems developed using the PHP programming language with HTML and SQL. Table I contains the data from [1] on four metrics of software for 32 open-source PHP-based systems.

TABLE I. THE DATA ON SOFTWARE METRICS

i	Y	X_1	X_2	X_3
1	3.038	5	2	10.6
2	22.599	17	7	7
3	32.243	21	13	4.524
4	16.164	13	11	7.077
5	83.862	35	24	6.571
6	24.22	13	9	8.077
7	63.929	35	19	8.029
8	2.543	5	3	9.4
9	6.697	5	5	7
10	55.537	25	14	8.64
11	55.752	39	10	9.077
12	62.602	30	17	7
13	67.111	23	22	14.957
14	2.552	3	1	8.333
15	12.17	10	5	3.7
16	12.757	13	9	5
17	5.695	7	3	8.429
18	7.744	9	6	9.222
19	7.514	4	1	8
20	11.054	9	9	3.667
21	29.77	17	15	3.412
22	11.653	9	8	8.778
23	6.847	5	4	3.6
24	13.389	7	5	11.714
25	14.45	12	6	16.583
26	4.414	6	3	3.667
27	2.102	3	1	3.333
28	42.819	20	18	3.5
29	4.077	4	2	9
30	57.408	33	14	9.242
31	7.428	7	3	7
32	8.947	15	5	4

For detecting the outliers in the data from Table 1 we use the technique based on multivariate normalizing transformations and the squared Mahalanobis distance [6]. 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. The same result was obtained in [6] for the transformation (9) for S_U family. In [1] it was also assumed that the data contains no outliers.

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: $\hat{\gamma}_Y = 9.63091$, $\hat{\gamma}_1 = 15.5355$, $\hat{\gamma}_2 = 25.4294$, $\hat{\gamma}_3 = 0.72801$, $\hat{\eta}_Y = 1.05243$, $\hat{\eta}_1 = 1.58306$, $\hat{\eta}_2 = 2.54714$, $\hat{\eta}_3 = 0.54312$,

$\hat{\varphi}_Y = -1.4568$, $\hat{\varphi}_1 = -1{,}8884$, $\hat{\varphi}_2 = -6{,}9746$, $\hat{\varphi}_3 = 3.2925$, $\hat{\lambda}_Y = 153102.605$, $\hat{\lambda}_1 = 243051.0$, $\hat{\lambda}_2 = 311229.5$ and $\hat{\lambda}_3 = 13.900$. The sample covariance matrix S_N of the $\mathbf{T}$ is used as the approximate moment-matching estimator of $\mathbf{\Sigma}$

$$S_N = \begin{pmatrix} 1.0000 & 0.9514 & 0.9333 & 0.1574 \\ 0.9514 & 1.0000 & 0.9006 & 0.1345 \\ 0.9333 & 0.9006 & 1.0000 & 0.0554 \\ 0.1574 & 0.1345 & 0.0554 & 1.0000 \end{pmatrix}.$$

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

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

Estimators for parameters of the Eq. (11) are such: $\hat{b}_0 = 1.02 \cdot 10^{-5}$, $\hat{b}_1 = 0.56085$, $\hat{b}_2 = 0.42491$, $\hat{b}_3 = 0.05846$.

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 Eq. (11), $Z_j = \gamma_j + \eta_j \ln \dfrac{X_j - \varphi_j}{\varphi_j + \lambda_j - X_j}$, $\varphi_j < X_j < \varphi_j + \lambda_j$, $j = 1,2,3$.

The prediction results by Eq. (12) for values of components of vector $\mathbf{X} = \{X_1, X_2, X_3\}$ from Table I are shown in the Table II for two cases: univariate and multivariate normalizing transformations.

For univariate normalizing transformations (10) of S_B family the estimators for parameters are such: $\hat{\gamma}_Y = 0.77502$, $\hat{\gamma}_1 = 0.59473$, $\hat{\gamma}_2 = 0.57140$, $\hat{\gamma}_3 = 0.68734$, $\hat{\eta}_Y = 0.44395$, $\hat{\eta}_1 = 0.48171$, $\hat{\eta}_2 = 0.49553$, $\hat{\eta}_3 = 0.51970$, $\hat{\varphi}_Y = 2.063$, $\hat{\varphi}_1 = 2.900$, $\hat{\varphi}_2 = 0.900$, $\hat{\varphi}_3 = 3.304$, $\hat{\lambda}_Y = 83.059$, $\hat{\lambda}_1 = 36.695$, $\hat{\lambda}_2 = 23.525$ and $\hat{\lambda}_3 = 13.660$. In the case of univariate normalizing transformations the estimators for parameters of the Eq. (11) are such: $\hat{b}_0 = 3.11 \cdot 10^{-7}$, $\hat{b}_1 = 0.43519$, $\hat{b}_2 = 0.52239$ and $\hat{b}_3 = 0.08546$.

Table II also contains the prediction results by linear regression equation from [1] for values of components of vector $\mathbf{X} = \{X_1, X_2, X_3\}$ from Table I. Note the prediction results by linear regression equation from [1] are negative for the three rows of data: 14, 19 and 27. All prediction results by non-linear regression equation (12) are positive.

Magnitude of relative error (MRE), mean magnitude of relative error (MMRE) and percentage of prediction (PRED(0.25)) are accepted as standard evaluations of prediction results by regression equations. The values of MRE for linear regression equation from [1], non-linear regression equation (12) for two cases (univariate and multivariate normalizing transformations) are shown in the Table II. The acceptable values of MMRE and PRED(0.25) are not more than 0.25 and not less than 0.75 respectively. The values of MMRE in the Table III indicate that only the value for Eq. (12) on the basis of multivariate normalizing transformation is less than 0.25. Although all values of PRED(0.25) in the Table III are less than 0.75 nevertheless the values are greater for Eq. (12). All values of multiple coefficient of determination R^2 in the Table III are greater than 0.75 but the value of R^2 is greater for Eq. (12) on the basis of multivariate transformation.

TABLE II. PREDICTION RESULTS AND MRE OF REGRESSION EQUATIONS

i	**Linear regression equation**		**Non-linear regression equation**			
			univariate transformation		**multivariate transformation**	
	$\hat{Y}$	***MRE***	$\hat{Y}$	***MRE***	$\hat{Y}$	***MRE***
1	3.237	0.0656	4.675	0.5388	4.550	0.4976
2	24.142	0.0683	19.965	0.1166	19.990	0.1154
3	37.524	0.1638	32.098	0.0045	33.535	0.0401
4	25.916	0.6033	23.171	0.4335	21.292	0.3173
5	74.624	0.1102	80.265	0.0429	83.618	0.0029
6	23.224	0.0411	20.524	0.1526	18.901	0.2196
7	67.215	0.0514	65.913	0.0310	70.647	0.1051
8	4.127	0.6228	5.789	1.2764	5.169	1.0328
9	5.906	0.1181	7.353	0.0980	6.356	0.0509
10	46.843	0.1565	42.098	0.2420	43.126	0.2235
11	57.814	0.0370	67.070	0.2030	49.823	0.1064
12	56.995	0.0896	53.497	0.1454	56.651	0.0951
13	61.856	0.0783	65.500	0.0240	60.617	0.0968
14	-2.395	1.9384	2.202	0.1370	2.447	0.0412
15	9.959	0.1816	9.693	0.2035	10.029	0.1759
16	21.218	0.6632	18.682	0.4644	18.105	0.4192
17	5.976	0.0493	7.083	0.2438	6.687	0.1743
18	13.991	0.8067	12.911	0.6673	11.301	0.4593
19	-1.371	1.1825	2.496	0.6678	3.096	0.5880
20	15.385	0.3918	13.301	0.2032	12.850	0.1625
21	35.179	0.1817	27.321	0.0823	29.061	0.0238
22	17.045	0.4627	15.461	0.3268	13.268	0.1386
23	2.017	0.7054	5.435	0.2062	5.112	0.2534
24	11.462	0.1440	10.367	0.2257	8.661	0.3531
25	22.513	0.5580	20.191	0.3973	15.888	0.0995
26	1.630	0.6307	5.318	0.2048	5.260	0.1916
27	-5.655	3.6902	2.142	0.0192	1.873	0.1090
28	43.975	0.0270	37.967	0.1133	38.631	0.0978
29	0.953	0.7662	3.892	0.0454	3.732	0.0846
30	57.164	0.0043	53.121	0.0747	54.381	0.0527
31	5.044	0.3209	6.861	0.0764	6.571	0.1154
32	16.360	0.8285	12.934	0.4456	14.258	0.5936

The confidence and prediction intervals of non-linear regression are defined by (6) and (8) respectively for the data from Table I.

TABLE III. VALUES OF R^2, MMRE AND PRED(0.25)

Coefficients	**Linear regression equation**	**Non-linear regression equation**	
		univariate transformation	**multivariate transformation**
R^2	0.9491	0.9591	0.9692
MMRE	0.4919	0.2535	0.2199
PRED(0.25)	0.5313	0.7188	0.7188

Table IV contains the lower (LB) and upper (UB) bounds of the prediction intervals of linear and non-linear regressions on the basis of univariate and multivariate transformations respectively for 0.05 significance level.

Note the lower bounds of the prediction interval of linear regression from [1] are negative for the thirteen rows of data: 1, 8, 9, 14, 15, 17, 19, 23, 24, 26, 27, 29 and 31. All the lower bounds of the prediction interval of non-linear regressions are positive. The widths of the prediction interval of non-linear regression on the basis of the Johnson multivariate transformation are less than for linear regression from [1] for the twenty rows of data: 1, 6, 8, 9, 14-20, 22-27, 29, 31 and 32. Also the widths of the prediction interval of non-linear regression on the basis of the Johnson multivariate transformation are less than following the Johnson univariate transformation for the twenty-three rows of data: 1-4, 6, 8-10, 15-18, 20-26, 28, 29, 31 and 32. Approximately the same results are obtained for the confidence interval of non-linear regression.

TABLE IV. BOUNDS OF THE PREDICTION INTERVALS

i	**Bounds for linear regression**		**Bounds for non-linear regression**			
			univariate transformation		**multivariate transformation**	
	LB	***UB***	***LB***	***UB***	***LB***	***UB***
1	-8.886	15.361	2.507	15.664	2.053	8.822
2	12.260	36.024	5.800	53.204	11.088	35.207
3	25.530	49.517	9.341	65.987	19.149	57.962
4	14.031	37.802	6.642	57.342	11.955	37.129
5	61.845	87.403	59.920	84.392	47.603	146.045
6	11.451	34.998	5.956	53.906	10.617	32.866
7	54.797	79.633	31.210	81.247	40.528	122.355
8	-7.849	16.103	2.713	20.215	2.431	9.838
9	-5.998	17.810	2.996	26.099	3.097	11.949
10	34.901	58.785	13.397	72.304	24.761	74.346
11	43.606	72.022	26.251	82.571	26.759	91.726
12	44.844	69.146	19.861	77.358	32.563	97.782
13	47.957	75.755	28.542	81.562	33.153	109.857
14	-14.415	9.625	2.084	2.994	0.811	5.262
15	-2.080	21.999	3.441	33.425	5.255	18.197
16	9.355	33.081	5.492	51.258	10.150	31.513
17	-5.925	17.877	2.964	24.822	3.336	12.381
18	2.136	25.846	4.145	40.894	6.095	20.095
19	-13.374	10.632	2.127	4.916	1.198	6.351
20	3.243	27.527	4.154	42.480	6.867	23.133
21	22.801	47.556	7.324	63.400	15.978	51.960
22	5.148	28.943	4.693	46.152	7.200	23.590
23	-10.093	14.128	2.635	19.103	2.367	9.829
24	-0.715	23.638	3.576	35.238	4.477	15.796
25	9.337	35.689	5.323	56.560	8.396	29.076
26	-10.481	13.741	2.621	18.450	2.464	10.048
27	-17.916	6.606	2.073	2.648	0.410	4.484
28	31.335	56.615	10.895	70.978	21.432	68.748
29	-11.043	12.949	2.371	12.014	1.575	7.423
30	44.632	69.696	19.170	77.441	30.902	94.883
31	-6.838	16.926	2.926	23.959	3.273	12.168
32	4.173	28.547	4.090	41.560	7.530	26.021

Following [7] multivariate kurtosis β_2 is estimated for the data on metrics of software from Table I and the normalized data on the basis of 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 28.66, 37.29 and 23.08 for the data from Table I, 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 only and does not hold for other data.

V. CONCLUSION

The non-linear regression equation to estimate the software size of open-source PHP-based 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, one should use multivariate 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 nonlinear.

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] Matinee Kiewkanya, and Suttipong Surak, "Constructing C++ software size estimation model from class diagram", in *13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, July 13-15, 2016, Khon Kaen, Thailand, pp. 1-6.

[3] D. M. Bates, and D. G. Watts, Nonlinear regression analysis and its applications. Wiley, 1988.

[4] T. P. Ryan, Modern regression methods. Wiley, 1997.

[5] S. B. Prykhodko, "Developing the software defect prediction models using regression analysis based on normalizing transformations", in *Abstracts of the Research and Practice Seminar on Modern Problems in Testing of the Applied Software (PTTAS-2016)*, May 25-26, 2016, Poltava, Ukraine, pp. 6-7.

[6] S. Prykhodko, N. Prykhodko, L. Makarova, and A. Pukhalevych, "Application of the squared Mahalanobis distance for detecting outliers in multivariate non-Gaussian data", in *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.

[7] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications", *Biometrika*, 57, 1970, pp. 519–530.