UDC 004.412:519.237.5

A MULTIPLE NON-LINEAR REGRESSION MODEL TO ESTIMATE THE AGILE TESTING EFFORTS FOR SMALL WEB PROJECTS

Prykhodko N. V. – PhD, Associate Professor, Associate Professor of the Finance Department, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.
Prykhodko S. B. – Dr. Sc., Professor, Head of the Department of Software of Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

ABSTRACT

Context. Software testing effort estimation is one of the important problems in software development and software testing life cycle. The object of the study is the process of estimating the agile testing efforts for small Web projects. The subject of the study is the multiple regression models for estimating the agile testing efforts for small Web projects.

Objective. The goal of the work is the creation of the multiple non-linear regression model for estimating the agile testing efforts for small Web projects on the basis of the Johnson multivariate normalizing transformation.

Method. The model, confidence and prediction intervals of multiple non-linear regression for estimating the agile testing efforts for small Web projects are constructed on the basis of the Johnson multivariate normalizing transformation for non-Gaussian data with the help of appropriate techniques. The techniques based on the multiple non-linear regression analysis using the multivariate normalizing transformations to build the models, equations, confidence and prediction intervals of multiple non-linear regressions are used. The techniques allow to take into account the correlation between random variables in the case of normalization of multivariate non-Gaussian data. In general, this leads to a reduction of the mean magnitude of relative error, the widths of the confidence and prediction intervals in comparison with the linear models and nonlinear models constructed using univariate normalizing transformations.

Results. Comparison of the constructed model with the linear model and non-linear regression models based on the decimal logarithm and the Johnson univariate transformation has been performed.

Conclusions. The multiple non-linear regression model to estimate the agile testing efforts for small Web projects is firstly constructed on the basis of the Johnson multivariate transformation for S_B family. This model, in comparison with other regression models (both linear and non-linear), has a smaller value of the mean magnitude of relative error, smaller widths of the confidence and prediction intervals. The prospects for further research may include the application of other multivariate normalizing transformations and data sets to construct the multiple non-linear regression model for estimating the agile testing efforts for small Web projects.

KEYWORDS: agile testing, estimation, testing effort, Web project, multiple non-linear regression model, multivariate normalizing transformation, non-Gaussian data.

ABBREVIATIONS

LB is lower bound;
MD is Mahalanobis distance;
MRE is a magnitude of relative error;
MMRE is a mean magnitude of relative error;
PRED is percentage of prediction;
UB is upper bound.

NOMENCLATURE

$\hat{\mathbf{b}}$ is estimator for vector of linear regression equation parameters, $\mathbf{b} = \{b_1, b_2, \ldots, b_k\}^T$;

$\hat{b}_i$ is estimator for the *i*-th parameter of linear regression equation;

k is a number of independent variables (regressors);

N is a number of data points;

$N(0,1)$ is a Gaussian distribution with zero mathematical expectation and unit variance;

P is a non-Gaussian random vector, $\mathbf{P} = \{Y, X_1, X_2, \ldots, X_k\}^T$;

R^2 is a multiple coefficient of determination;

$\mathbf{S}_N$ is a sample covariance matrix, $\mathbf{S}_N = [S_{ij}]$;

T is a Gaussian random vector, $\mathbf{T} = \{Z_Y, Z_1, Z_2, \ldots, Z_k\}^T$;

$t_{\alpha/2,\nu}$ is a quantile of student's *t*-distribution with ν degrees of freedom and $\alpha/2$ significance level;

X_1 is a number of test cases;

X_2 is a number of design document pages;

X_3 is a number of defects;

Y is an actual testing effort in person hours;

$\mathbf{Z}_X^+$ is a matrix of centered regressors that contains the values $Z_{1_i} - \overline{Z}_1$, $Z_{2_i} - \overline{Z}_2$, …, $Z_{k_i} - \overline{Z}_k$;

$\left(\mathbf{Z}_X^+\right)^T$ is a transpose of $\mathbf{Z}_X^+$;

$\mathbf{z}_X^+$ is a vector with components $Z_{1_i} - \overline{Z}_1$, $Z_{2_i} - \overline{Z}_2$, …, $Z_{k_i} - \overline{Z}_k$ for *i*-row;

$\left(\mathbf{z}_X^+\right)^T$ is a transpose of $\mathbf{z}_X^+$;

$\overline{Z}_Y$ is a sample mean of the values of the variable Z_Y;

$\hat{Z}_Y$ is a prediction linear regression equation result;

α is a significance level;

DOI 10.15588/1607-3274-2019-2-17

β_1 is a multivariate skewness;

β_2 is a multivariate kurtosis;

$\boldsymbol{\gamma}$ is a vector of parameters of the Johnson multivariate translation, $\boldsymbol{\gamma} = (\gamma_Y, \gamma_1, \gamma_2, \ldots, \gamma_k)^T$;

ε is a Gaussian random variable which defines residuals, $\varepsilon \sim N(0,1)$;

$\boldsymbol{\eta}$ is a vector of parameters of the Johnson multivariate translation, $\boldsymbol{\eta} = \operatorname{diag}(\eta_Y, \eta_1, \ldots, \eta_k)$;

$\boldsymbol{\lambda}$ is a vector of parameters of the Johnson multivariate translation, $\boldsymbol{\lambda} = \operatorname{diag}(\lambda_Y, \lambda_1, \ldots, \lambda_k)$;

ν is a number of degrees of freedom;

$\boldsymbol{\Sigma}$ is a covariance matrix, $\boldsymbol{\Sigma} = [\Sigma_{ij}]$;

φ is a vector of parameters of the Johnson multivariate translation, $\varphi = (\varphi_Y, \varphi_1, \varphi_2, \ldots, \varphi_k)^T$;

$\boldsymbol{\psi}$ is a vector of multivariate normalizing transformation, $\boldsymbol{\psi} = \{\psi_Y, \psi_1, \psi_2, \ldots, \psi_k\}^T$.

INTRODUCTION

Software testing effort estimation is one of the important problems in software development and software testing life cycle. Agile testing is a software testing process that follows the principles of agile software development [1–5]. In comparison with waterfall testing, it is a new age software testing approach which leads to a reduction of testing efforts. Agile testing is well suited for small software projects including small Web projects.

The agile testing lifecycle consists of the 5 phases [5], the second of which is agile testing planning, that includes testing effort estimation. A testing effort estimation is a difficult problem, for the solution of which various mathematical models are applied.

Today one of the most well-known effort estimation model is the COCOMO II (COnstructive COst MOdel) [6]. The COCOMO II is a non-linear regression equation with parameters that are derived from historical data of software projects. This equation is built on the basis of univariate normalizing transformation in the decimal logarithm form. The paper [7] proposed the multiple linear and non-linear regression equations for estimating the testing efforts of software projects including large ones. However, a prediction regression equation result is a mean value of dependent random variable. There is no random error term in regression equation. A prediction regression model result is a value of dependent random variable, since there is the random error term in regression model. Therefore, to predict agile testing effort as a value of a dependent random variable there is the need to develop the appropriate non-linear regression models.

The object of study is the process of estimating the agile testing efforts for small Web projects.

The subject of study is the multiple non-linear regression models to estimate the agile testing efforts for small Web projects.

The purpose of the work is to construct the multiple non-linear regression model for estimating the agile testing efforts for small Web projects. The agile testing effort prediction results by constructed model should be better in comparison with other regression models, both linear and nonlinear, primarily on such standard evaluations as mean magnitude of relative error, widths of confidence and prediction intervals.

1 PROBLEM STATEMENT

Suppose given the original sample as the four-dimensional non-Gaussian data set: actual testing effort in person hours Y, the total number of test cases X_1, the total number of design document pages X_2 and the number of defects X_3 from N small Web projects. Suppose that there are 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)

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

It is required to build the multiple non-linear regression model in the form $Y = Y(X_1, X_2, X_3, \varepsilon)$ on the basis of the transformations (1) and (2).

2 REVIEW OF THE LITERATURE

A normalizing transformation is often a good way to build the models, equations, confidence and prediction intervals of multiple non-linear regressions [8–13]. According to [9] transformations are made for essentially four purposes, two of which are: firstly, to obtain approximate normality for the distribution of the error term (residuals) or the dependent random variable, secondly, 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 models, equations, confidence and prediction intervals of multiple non-linear regressions are based on the univariate normalizing transformations (such as, the decimal logarithm, the natural logarithm, the Box-Cox transformation), which do not take into account the correlation between random variables in the case of normalization of multivariate non-Gaussian data. Application of univariate normalizing transformations for building the multiple non-linear regression models does not always lead to good prediction results by such regression models, primarily on such standard evaluations as mean magnitude of relative error, widths of confidence and prediction intervals [13]. This leads to the need to use the multivariate normalizing transformations.

DOI 10.15588/1607-3274-2019-2-17

In [13] the techniques to build the models, confidence and prediction intervals of multiple non-linear regressions for multivariate non-Gaussian data on the basis of the bijective multivariate normalizing transformations were proposed. 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 model, confidence and prediction intervals of linear regression for the normalized data are built. In the third step, the model, confidence and prediction intervals of multiple non-linear regression for multivariate non-Gaussian data are constructed on the basis of the model, confidence and prediction intervals of linear regression for the normalized data and the multivariate normalizing transformation.

Non-linear regression prediction results by models, which constructed in the papers [13, 14] on the basis of the Johnson multivariate normalizing transformation, are better in comparison with other regression models, both linear and nonlinear, primarily on such standard evaluations as mean magnitude of relative error, widths of confidence and prediction intervals.

This leads to the need to develop the multiple non-linear regression model for estimating the agile testing efforts for small Web projects on the basis of the multivariate normalizing transformations.

3 MATERIALS AND METHODS

After normalizing the non-Gaussian data by the transformation (1) the linear regression model is built for normalized data. The linear regression model for normalized data will have the form [13]

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

After that the multiple non-linear regression model is built on the basis of the linear regression model (3) for the normalized data and the transformations (1) and (2). The non-linear regression model will have the form [13]

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

The technique to build a confidence interval of multiple non-linear regression is based on a confidence interval of linear regression for normalized data, and transformations (1) and (2) [13]:

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

where $S_{Z_Y}^2 = \frac{1}{\nu}\sum_{i=1}^{N}\left(Z_{Y_i} - \hat{Z}_{Y_i}\right)^2$, $\nu = N-k-1$; $\left(\mathbf{Z}_X^+\right)^T\mathbf{Z}_X^+$ 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 $S_{Z_qZ_r} = \sum_{i=1}^{N}\left[Z_{q_i} - \bar{Z}_q\right]\left[Z_{r_i} - \bar{Z}_r\right]$, $q,r = 1,2,\dots,k$.

The technique to build a prediction interval of multiple non-linear regression is based on a prediction interval of linear regression for normalized data, and transformations (1) and (2) [13]:

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

For normalizing the multivariate non-Gaussian data, we use the Johnson translation system. In our case the Johnson normalizing translation is given by [14]

$$\mathbf{T} = \boldsymbol{\gamma} + \boldsymbol{\eta}\mathbf{h}\left[\boldsymbol{\lambda}^{-1}\left(\mathbf{P} - \boldsymbol{\varphi}\right)\right] \sim N_m\left(\mathbf{0}_m, \boldsymbol{\Sigma}\right), \quad (5)$$

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

$$h = \begin{cases} \ln(y), & \text{for } S_L \text{ (log normal) family;} \\ \ln\left[y/(1-y)\right], & \text{for } S_B \text{ (bounded) family;} \\ \operatorname{Arsh}(y), & \text{for } S_U \text{ (unbounded) family;} \\ y & \text{for } S_N \text{ (normal) family.} \end{cases} \quad (6)$$

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

The model, equation, confidence and prediction intervals of multiple non-linear regression to estimate agile testing efforts for small Web projects are constructed on the basis of the Johnson multivariate normalizing transformation for the four-dimensional non-Gaussian data set from Table 1 for 40 small Web projects (rows 1–40). Also Table 1 contains the values of squared Mahalanobis distance (MD) for 41 and 40 (after outlier cutoff) data rows. For detecting the outliers in the data from Table 1 we use the technique based on multivariate normalizing transformations and the squared MD [15]. There is one outlier in the data from Table 1 for 0.005 significance level and the Johnson multivariate transformation (5) for S_B family.

Parameters of the multivariate transformation (5) for S_B family were estimated by the maximum likelihood method. Estimators for parameters of the transformation (6) for 41 data rows are: $\hat{\gamma}_Y = 4.09443$, $\hat{\gamma}_1 = 5.47043$, $\hat{\gamma}_2 = 1.09282$, $\hat{\gamma}_3 = 1.37671$, $\hat{\eta}_Y = 1.04794$, $\hat{\eta}_1 = 0.97350$, $\hat{\eta}_2 = 0.70189$, $\hat{\eta}_3 = 0.64464$, $\hat{\varphi}_Y = 0.37266$, $\hat{\varphi}_1 = 1.95622$, $\hat{\varphi}_2 = 0.94564$, $\hat{\varphi}_3 = 2.35215$, $\hat{\lambda}_Y = 327.313$, $\hat{\lambda}_1 = 5438.99$, $\hat{\lambda}_2 = 130.495$ and $\hat{\lambda}_3 = 110.210$. The sample covariance

DOI 10.15588/1607-3274-2019-2-17

matrix $\mathbf{S}_N$ of the $\mathbf{T}$ is used as the approximate moment-matching estimator of $\mathbf{\Sigma}$

$$\mathbf{S}_N = \begin{pmatrix} 1.0000 & 0.9812 & 0.4088 & 0.8497 \\ 0.9812 & 1.0000 & 0.4326 & 0.7519 \\ 0.4088 & 0.4326 & 1.0000 & 0.2029 \\ 0.8497 & 0.7519 & 0.2029 & 1.0000 \end{pmatrix}.$$

The data of system 41 is multivariate outlier, since for this data row the squared MD equals to 20.43 is greater than the value of the quantile of the Chi-Square distribution, which equals to 14.86 for 0.005 significance level. The same result was obtained for the univariate transformation in the decimal logarithm form. In this case the data of system 41 is multivariate outlier too, since for this data row the squared MD equals to 20.26.

The squared MD values for 40 data rows indicate there are no outliers in this data from Table 1.

Table 1 – The data set and squared MDs

No	Y	X_1	X_2	X_3	Squared MD	
					N=41	N=40
1	1.33	4	11	4	5.69	6.14
2	3.51	11	2	7	5.75	5.83
3	3.17	10	4	8	2.67	2.50
4	1.53	5	14	4	3.80	4.38
5	2.35	9	12	3	4.71	5.56
6	3.13	10	8	9	2.56	2.39
7	2.03	6	16	7	3.92	4.07
8	3.22	11	11	6	1.31	1.34
9	2.73	10	16	5	2.83	3.89
10	4.65	16	13	7	0.75	0.68
11	5.52	21	12	4	7.56	7.32
12	2.75	7	23	12	4.28	6.42
13	6.93	25	6	13	2.68	2.83
14	3.73	14	29	4	2.44	2.47
15	5.08	18	22	7	0.62	0.60
16	7.12	25	11	11	1.11	1.06
17	3.24	10	31	9	2.49	2.19
18	4.05	12	35	6	2.63	4.16
19	4.72	16	32	7	0.58	0.51
20	3.49	13	41	4	2.52	2.58
21	6.03	18	24	18	0.64	0.57
22	4.13	12	40	13	2.35	2.20
23	10.18	33	15	23	0.62	0.64
24	9.95	36	29	12	1.21	1.10
25	8.67	24	36	32	1.37	1.92
26	16.53	51	3	45	7.11	7.40
27	12.45	44	37	18	0.97	0.91
28	15.56	56	25	19	2.84	2.64
29	17.47	57	7	40	3.92	3.98
30	11.23	29	48	47	2.78	4.27
31	8.21	29	79	10	2.30	2.17
32	19.95	50	6	90	8.33	9.59
33	16.16	44	53	61	3.12	3.88
34	8.83	33	110	8	6.14	5.95
35	12.97	47	94	16	3.30	3.13
36	21.32	94	48	37	4.73	7.40
37	20.97	83	61	62	8.87	11.09
38	38.22	127	25	78	4.77	4.83
39	26.48	111	71	54	5.50	7.38
40	48.2	173	120	60	11.73	12.03
41	7.07	11	13	51	**20.43**	–

Estimators for parameters of the transformation (5) for S_B family for 40 data rows are: $\hat{\gamma}_Y = 3.8484$, $\hat{\gamma}_1 = 5.4050$, $\hat{\gamma}_2 = 1.0397$, $\hat{\gamma}_3 = 1.3214$, $\hat{\eta}_Y = 0.9990$, $\hat{\eta}_1 = 0.96416$, $\hat{\eta}_2 = 0.68334$, $\hat{\eta}_3 = 0.61537$, $\hat{\varphi}_Y = 0.52944$, $\hat{\varphi}_1 = 2.0172$, $\hat{\varphi}_2 = 1.0107$, $\hat{\varphi}_3 = 2.5590$, $\hat{\lambda}_Y = 298.41$, $\hat{\lambda}_1 = 5439.41$, $\hat{\lambda}_2 = 128.66$ and $\hat{\lambda}_3 = 103.668$. The sample covariance matrix $\mathbf{S}_N$

$$\mathbf{S}_N = \begin{pmatrix} 1.0000 & 0.9898 & 0.4108 & 0.8588 \\ 0.9898 & 1.0000 & 0.4280 & 0.7909 \\ 0.4108 & 0.4280 & 1.0000 & 0.2204 \\ 0.8588 & 0.7909 & 0.2204 & 1.0000 \end{pmatrix}.$$

After normalizing the non-Gaussian data by the multivariate transformation (5) for S_B family the linear regression model is built for normalized data

$$Z_Y = \hat{Z}_Y + \varepsilon = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3 + \varepsilon. \quad (7)$$

Parameters of the linear regression model (7) were estimated by the least square method. Estimators for parameters of the model (7) are such: $\hat{b}_0 = 0$, $\hat{b}_1 = 0.81961$, $\hat{b}_2 = 0.01428$, $\hat{b}_3 = 0.20730$.

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

$$Y = \hat{\varphi}_Y + \hat{\lambda}_Y \left[1 + e^{-\left(\hat{Z}_Y + \varepsilon - \hat{\gamma}_Y\right)/\hat{\eta}_Y} \right]^{-1}, \quad (8)$$

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

The model (8) is the multiple non-linear regression model to estimate the agile testing efforts for small Web projects.

4 EXPERIMENTS

For comparison of the model (8) with other multiple models one linear regression model and two non-linear regression models are built on the basis of 40 data rows from Table 1 and two univariate normalizing transformations: the decimal logarithm transformation and the Johnson transformation.

The multiple linear regression model has the form

$$Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \varepsilon, \quad (9)$$

where the estimators for parameters are: $\hat{b}_0 = 0.26513$, $\hat{b}_1 = 0.23116$, $\hat{b}_2 = -0.00082$, $\hat{b}_3 = 0.08374$.

The multiple non-linear regression model is constructed on the basis of the linear regression model (7)

DOI 10.15588/1607-3274-2019-2-17

for the normalized data and the decimal logarithm transformation

$$Y = 10^{\varepsilon + \hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}, \qquad (10)$$

where the estimators for parameters are: $\hat{b}_0 = -0.4500$, $\hat{b}_1 = 0.78887$, $\hat{b}_2 = 0.00176$, $\hat{b}_3 = 0.18782$.

The multiple non-linear regression model is constructed on the basis of the linear regression model (7) for the normalized data and the Johnson univariate transformation for S_B family (6). In this case the estimators for parameters of the model (8) are: $\hat{\gamma}_Y = 1.72684$, $\hat{\gamma}_1 = 1.78876$, $\hat{\gamma}_2 = 1.2346$, $\hat{\gamma}_3 = 1.1127$, $\hat{\eta}_Y = 0.73211$, $\hat{\eta}_1 = 0{,}73090$, $\hat{\eta}_2 = 0.72269$, $\hat{\eta}_3 = 0.53444$, $\hat{\varphi}_Y = 1.1292$, $\hat{\varphi}_1 = 3{,}3512$, $\hat{\varphi}_2 = 1.1717$, $\hat{\varphi}_3 = 2.90$, $\hat{\lambda}_Y = 64.039$, $\hat{\lambda}_1 = 237.48$, $\hat{\lambda}_2 = 142.95$, $\hat{\lambda}_3 = 92.745$, $\hat{b}_0 = 0$, $\hat{b}_1 = 0.86376$, $\hat{b}_2 = 0.00240$ and $\hat{b}_3 = 0.16633$.

The computer program implementing the constructed models (8), (9) and (10) was developed to conduct experiments. The program was written in the sci-language for the Scilab system. Scilab (http://www.scilab.org) is the free and open source software, the alternative to commercial packages for system modeling and simulation packages such as MATLAB and MATRIXx.

5 RESULTS

If the Gaussian random variable ε equals zero the regression models (8), (9) and (10) are the multiple regression equations for which the prediction results for values of components of vector $\mathbf{X} = \{X_1, X_2, X_3\}$ from Table 1 and values of MRE are shown in the Table 2.

Table 2 – The prediction results and confidence intervals of multiple regressions

No	Multiple linear regression				Multiple non-linear regressions											
					univariate normalizing transformation								the Johnson multivariate normalizing transformation			
					the decimal logarithm				the Johnson transformation							
	$\hat{Y}$	MRE	LB	UB	$\hat{Y}$	MRE	LB	UB	$\hat{Y}$	MRE	LB	UB	$\hat{Y}$	MRE	LB	UB
1	1.516	0,1396	0.934	2.097	1.380	0.0375	1.312	1.451	1.360	0.0225	1.334	1.389	1.298	0.0242	1.252	1.346
2	3.392	0,0335	2.752	4.033	3.394	0.0329	3.192	3.609	3.383	0.0363	3.154	3.636	3.387	0.0352	3.223	3.560
3	3.243	0,0231	2.635	3.851	3.233	0.0197	3.088	3.384	3.188	0.0055	3.042	3.344	3.234	0.0203	3.121	3.352
4	1.744	0,1401	1.179	2.310	1.646	0.0760	1.574	1.722	1.645	0.0751	1.603	1.691	1.593	0.0411	1.541	1.648
5	2.587	0,1008	2.000	3.174	2.479	0.0549	2.371	2.592	2.248	0.0436	2.114	2.399	2.312	0.0160	2.210	2.421
6	3.324	0,0619	2.758	3.890	3.309	0.0571	3.192	3.430	3.239	0.0347	3.121	3.363	3.320	0.0607	3.222	3.421
7	2.225	0,0961	1.685	2.765	2.112	0.0404	2.016	2.213	2.041	0.0056	1.969	2.120	2.079	0.0240	2.001	2.161
8	3.301	0,0252	2.735	3.868	3.308	0.0272	3.204	3.415	3.320	0.0310	3.210	3.435	3.357	0.0425	3.270	3.447
9	2.982	0,0924	2.438	3.527	2.967	0.0867	2.871	3.065	2.983	0.0928	2.888	3.085	3.005	0.1006	2.925	3.087
10	4.539	0,0238	3.982	5.096	4.577	0.0157	4.439	4.720	4.624	0.0055	4.462	4.794	4.673	0.0049	4.551	4.799
11	5.445	0,0137	4.797	6.092	5.105	0.0751	4.822	5.406	5.127	0.0712	4.790	5.493	5.075	0.0806	4.816	5.350
12	2.869	0,0434	2.331	3.408	2.641	0.0397	2.495	2.795	2.461	0.1053	2.347	2.584	2.585	0.0598	2.470	2.708
13	7.128	0,0285	6.509	7.746	7.301	0.0535	7.001	7.614	7.304	0.0540	6.923	7.708	7.315	0.0556	7.047	7.595
14	3.812	0,0221	3.281	4.344	3.714	0.0044	3.560	3.873	3.712	0.0048	3.541	3.895	3.734	0.0010	3.598	3.875
15	4.994	0,0169	4.484	5.504	5.027	0.0104	4.879	5.180	5.101	0.0042	4.929	5.281	5.158	0.0153	5.027	5.292
16	6.956	0,0230	6.372	7.540	7.083	0.0052	6.855	7.319	7.150	0.0042	6.858	7.456	7.167	0.0066	6.952	7.388
17	3.305	0,0200	2.773	3.837	3.317	0.0237	3.189	3.449	3.247	0.0020	3.121	3.380	3.365	0.0387	3.258	3.477
18	3.513	0,1327	2.967	4.058	3.550	0.1235	3.429	3.675	3.572	0.1180	3.445	3.706	3.644	0.1001	3.544	3.748
19	4.524	0,0416	4.016	5.031	4.584	0.0288	4.446	4.726	4.633	0.0184	4.476	4.797	4.718	0.0004	4.601	4.839
20	3.571	0,0234	2.978	4.165	3.505	0.0042	3.355	3.661	3.503	0.0037	3.341	3.676	3.536	0.0132	3.408	3.670
21	5.914	0,0193	5.441	6.386	6.004	0.0043	5.813	6.201	5.804	0.0375	5.598	6.019	6.027	0.0004	5.871	6.189
22	4.095	0,0085	3.497	4.692	4.106	0.0059	3.928	4.292	3.960	0.0411	3.781	4.151	4.153	0.0055	4.006	4.306
23	9.807	0,0366	9.306	10.308	10.133	0.0046	9.851	10.422	9.967	0.0210	9.599	10.348	10.055	0.0123	9.785	10.332
24	9.568	0,0384	9.041	10.095	9.616	0.0336	9.304	9.938	9.847	0.0103	9.471	10.238	9.756	0.0195	9.464	10.058
25	8.463	0,0239	7.795	9.131	8.399	0.0312	8.031	8.784	8.011	0.0760	7.640	8.401	8.369	0.0348	8.085	8.663
26	15.820	0,0430	15.010	16.630	16.158	0.0225	15.120	17.268	15.769	0.0460	14.546	17.068	15.589	0.0569	14.693	16.539
27	11.913	0,0431	11.427	12.399	12.162	0.0232	11.795	12.539	12.313	0.0110	11.858	12.783	12.194	0.0206	11.846	12.552
28	14.780	0,0501	14.067	15.494	14.850	0.0456	14.303	15.418	15.196	0.0234	14.538	15.877	14.720	0.0540	14.191	15.269
29	16.785	0,0392	16.019	17.551	17.280	0.0109	16.440	18.163	17.000	0.0269	16.031	18.012	16.734	0.0421	15.986	17.517
30	10.865	0,0325	9.819	11.912	10.487	0.0662	9.922	11.083	10.015	0.1082	9.433	10.630	10.506	0.0645	10.054	10.979
31	7.741	0,0571	6.788	8.695	7.849	0.0440	7.528	8.183	8.029	0.0221	7.605	8.477	8.127	0.0101	7.825	8.442
32	19.355	0,0298	17.602	21.107	18.142	0.0906	16.983	19.379	19.878	0.0036	18.087	21.774	19.676	0.0137	18.356	21.087
33	15.501	0,0408	14.285	16.716	15.305	0.0529	14.513	16.139	14.950	0.0749	14.080	15.861	15.494	0.0412	14.806	16.213
34	8.473	0,0404	7.049	9.897	8.339	0.0556	7.907	8.794	8.674	0.0177	8.044	9.354	8.741	0.0100	8.262	9.249
35	12.392	0,0445	11.315	13.469	12.551	0.0323	12.012	13.115	12.862	0.0083	12.121	13.641	12.803	0.0129	12.256	13.374
36	25.053	0,1751	24.147	25.958	25.353	0.1892	24.289	26.464	25.685	0.2048	24.603	26.786	24.519	0.1501	23.577	25.496
37	24.593	0,1728	23.696	25.490	25.333	0.2081	24.204	26.514	25.102	0.1971	23.980	26.247	25.158	0.1997	24.163	26.191
38	36.133	0,0546	34.746	37.520	36.936	0.0336	35.110	38.858	37.161	0.0277	35.682	38.622	36.607	0.0422	34.926	38.359
39	30.387	0,1476	29.413	31.361	31.054	0.1727	29.615	32.564	31.097	0.1744	29.813	32.388	30.135	0.1380	28.907	31.410
40	45.181	0,0626	43.360	47.003	44.994	0.0665	42.372	47.778	46.127	0.0430	44.434	47.737	43.296	0.1018	40.695	46.035

DOI 10.15588/1607-3274-2019-2-17

The prediction results by model (8) and values of MRE are shown in the Table 2 for two cases: Johnson's univariate and multivariate normalizing transformations. Table 2 also contains the prediction results by linear regression model (9) for values of components of vector **X** from Table 1 and MRE values. The MRE values for the multiple non-linear regression model (8) based on the Johnson multivariate transformation are smaller than for the linear regression model (9) for 25 rows of data: 1, 3–7, 10, 14–16, 18–24, 27, 31, 32, 34–36, 38, 39. Also the MRE values for the non-linear regression model (8) based on the Johnson multivariate transformation are smaller than for the multiple non-linear regression model (10) following the decimal logarithm univariate transformation for 22 rows of data: 1, 2, 4, 5, 7, 10, 14, 18, 19, 21, 22, 24, 27, 30–37, 39. And ones are smaller than for the non-linear regression model (8) following the Johnson univariate transformation for only 18 rows of data: 2, 4, 5, 10, 12, 14, 18, 19, 21–23, 25, 30, 31, 33, 34, 36 and 39.

MMRE and 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 R^2 is approximately the same as for PRED(0.25). The values of R^2, MMRE and PRED(0.25) equal respectively 0.9847, 0.0565 and 1.0 for linear regression model (9), and equal respectively 0.9810, 0.0503 and 1.0 for the model (10), and equal respectively 0.9828, 0.0478 and 1.0 for the model (8) for the Johnson univariate transformation, and equal respectively 0.9818, 0.0443 and 1.0 for the model (8) for the Johnson multivariate transformation. The value of MMRE is better for the model (8) for the Johnson multivariate transformation in comparison with all previous models.

The confidence and prediction intervals of multiple non-linear regression are defined for the data from Table 1. Table 2 contains the lower (LB) and upper (UB) bounds of the confidence intervals of linear and multiple non-linear regressions on the basis of univariate and multivariate transformations respectively for 0.05 significance level. The widths of the confidence interval of multiple non-linear regression on the basis of the Johnson multivariate transformation are smaller than for linear regression (9) for 34 rows of data: 1–25, 27–35. Also the widths of the confidence interval of multiple non-linear regression on the basis of the Johnson multivariate transformation are less for more data rows than for multiple non-linear regressions following the univariate transformations, both decimal logarithm and the Johnson. The widths of the confidence interval of multiple non-linear regression on the basis of the Johnson multivariate transformation are smaller than following the decimal logarithm univariate transformation for 37 rows of data: 1–31, 33, 36–40. And ones are smaller than following the Johnson univariate transformation for 34 rows of data: 2, 3, 5, 6, 8–11, 13–37 and 39. Approximately the same results are obtained for the prediction intervals of regressions.

Table 3 contains the lower (LB) and upper (UB) bounds of the prediction intervals of multiple 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 (9) are negative for the four rows of data: 1, 4, 5 and 7. All the lower bounds of the prediction interval of multiple non-linear regressions are positive. The widths of the prediction interval of multiple non-linear regression on the basis of the Johnson multivariate transformation are smaller than for linear regression (9) for 35 rows of data: 1–35.

Table 3 – The bounds of the prediction intervals

No	Bounds for linear regression		Bounds for multiple non-linear regression			
			decimal logarithm transformation		Johnson multivariate transformation	
	LB	UB	LB	UB	LB	UB
1	–1.174	4.205	1.186	1.606	1.184	1.431
2	0.689	6.096	2.906	3.966	2.968	3.877
3	0.548	5.939	2.782	3.756	2.850	3.682
4	–0.942	4.431	1.417	1.912	1.439	1.773
5	–0.104	5.278	2.134	2.879	2.050	2.620
6	0.637	6.010	2.855	3.834	2.928	3.775
7	–0.456	4.906	1.817	2.454	1.854	2.341
8	0.615	5.988	2.857	3.829	2.962	3.815
9	0.300	5.664	2.562	3.435	2.658	3.407
10	1.855	7.224	3.955	5.297	4.097	5.340
11	2.740	8.149	4.377	5.955	4.412	5.850
12	0.188	5.550	2.264	3.080	2.284	2.938
13	4.430	9.826	6.291	8.473	6.366	8.416
14	1.133	6.492	3.200	4.310	3.279	4.263
15	2.319	7.669	4.344	5.818	4.517	5.900
16	4.266	9.647	6.117	8.201	6.248	8.229
17	0.625	5.984	2.860	3.847	2.965	3.830
18	0.830	6.195	3.064	4.112	3.209	4.150
19	1.849	7.198	3.961	5.306	4.137	5.390
20	0.879	6.264	3.018	4.070	3.109	4.034
21	3.245	8.582	5.186	6.951	5.268	6.906
22	1.402	6.788	3.535	4.768	3.640	4.748
23	7.133	12.481	8.759	11.722	8.754	11.554
24	6.889	12.246	8.304	11.135	8.489	11.217
25	5.753	11.173	7.231	9.756	7.281	9.627
26	13.072	18.568	13.802	18.916	13.436	18.079
27	9.242	14.584	10.508	14.075	10.607	14.018
28	12.059	17.502	12.810	17.214	12.782	16.948
29	14.049	19.520	14.853	20.103	14.496	19.309
30	8.038	13.692	8.997	12.224	9.107	12.124
31	4.947	10.535	6.763	9.108	7.066	9.356
32	16.197	22.512	15.500	21.234	16.896	22.892
33	12.607	18.394	13.141	17.825	13.420	17.880
34	5.486	11.460	7.160	9.712	7.551	10.127
35	9.554	15.231	10.809	14.575	11.095	14.774
36	22.275	27.831	21.839	29.433	21.321	28.160
37	21.818	27.368	21.804	29.432	21.874	28.896
38	33.163	39.103	31.740	42.984	31.889	41.925
39	27.586	33.188	26.713	36.101	26.235	34.553
40	41.985	48.377	38.533	52.537	37.588	49.717

Also the widths of the prediction interval of multiple non-linear regression on the basis of the Johnson multivariate transformation are smaller for more data rows than for multiple non-linear regressions following the univariate transformations, both decimal logarithm and the Johnson. The widths of the prediction interval of

DOI 10.15588/1607-3274-2019-2-17

multiple non-linear regression on the basis of the Johnson multivariate transformation are smaller than following the decimal logarithm univariate transformation for 38 rows of data: 1–31, 33, 35–40. And ones are smaller than following the Johnson univariate transformation for 26 rows of data: 1, 2, 4–19, 21, 23–26, 28–30.

The null hypothesis that the observed frequency distribution of residuals for linear regression models (7) and (9) is the same as the normal distribution was tested by Pearson's chi-squared test. We can accept the null hypothesis that the distribution of residuals for linear regression model (7) is the same as the normal distribution for normalized data, which normalized by the Johnson multivariate transformation only, since the chi-squared test statistic value equals to 5.59 is smaller than the critical value of the chi-square, which equals to 7.81 for 3 degrees of freedom and 0.05 significance level. The chi-squared test statistic values equal to 60.61, 12.41 and 17.34 respectively for the model (9), the model (7) for normalized data, which normalized by the decimal logarithm univariate transformation and the Johnson univariate transformation for S_B family.

Following [16] multivariate skewness β_1 and kurtosis β_2 are estimated for 40 data rows from Table I and the normalized data on the basis of the decimal logarithm transformation, the Johnson univariate and multivariate transformations for S_B family. The measures β_1 and β_2 allow one to test two hypotheses that are compatible with the assumption of multivariate normality. The estimator of multivariate skewness given by [16]

$$\hat{\beta}_1 = \frac{1}{N^2}\sum_{i-1}^{N}\left\{\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)^T \mathbf{S}_N^{-1}\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)\right\}^3 . \quad (11)$$

The estimator of multivariate kurtosis given by [16]

$$\hat{\beta}_2 = \frac{1}{N}\sum_{i-1}^{N}\left\{\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)^T \mathbf{S}_N^{-1}\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)\right\}^2 . \quad (12)$$

In our case, in the formulas (11) and (12), 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 $\beta_1 = m(m+1)(m+2)/N$ and $\beta_2 = \mathrm{m}(\mathrm{m}+2)$ hold under multivariate normality. The given equalities are necessary conditions for multivariate normality. In our case $\beta_1 = 3$ and $\beta_2 = 24$. The estimators of multivariate skewness and kurtosis equal 19.38, 4.18, 5.30, 4.65, and 47.37, 23.22, 26.32, 24.29 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.

6 DISCUSSION

As it evident from the Table 3, the values of lower bounds of the prediction intervals of linear regression (9) for estimating the agile testing efforts for small Web projects are negative for some data rows. In our opinion, the presence of negative values may be explained by two reasons. Firstly, for the initial data from Table 1, four basic assumptions that justify the use of linear regression model, one of which is normality of the error distribution, are not valid. Moreover, the chi-squared test statistic value for residuals in linear regression model (9) is larger than for residuals in linear regression model (7) for normalized data, which normalized by the Johnson multivariate transformation, more than 10 times. Secondly, there is reason to reject the hypothesis that the sample of normalized data comes from a multivariate normal distribution. Note all the lower bounds of the prediction intervals of multiple non-linear regressions are positive.

Also note that in our case for the data from Table 1, 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 multiple non-linear regression for a larger number of data rows compared to the Johnson multivariate transformation.

The widths of the confidence and prediction intervals of multiple non-linear regression on the basis of the Johnson multivariate transformation are smaller for more data rows than for linear regression and multiple non-linear regressions following the univariate transformations, both decimal logarithm and the Johnson. Also the MMRE value is smaller for the model (8) for the Johnson multivariate transformation in comparison with all other models, both linear and non-linear, based on univariate transformations. This may be explained best multivariate normalization and the fact that there is no reason to reject the null hypothesis that the distribution of residuals for linear regression model (7) is the same as the normal distribution for normalized data, which normalized by the Johnson multivariate transformation only.

CONCLUSIONS

The important problem of increase of confidence of agile testing effort estimation for small Web projects is solved.

The scientific novelty of obtained results is that the multiple non-linear regression model to estimate the agile testing efforts for small Web projects is firstly constructed on the basis of the Johnson multivariate transformation for S_B family. This model, in comparison with other regression models (both linear and non-linear), has a smaller value of the mean magnitude of relative error, smaller widths of the confidence and prediction intervals of multiple non-linear regression.

The practical significance of obtained results is that the software realizing the constructed model is developed in the sci-language for Scilab. The experimental results

DOI 10.15588/1607-3274-2019-2-17

allow to recommend the constructed model for use in practice.

Prospects for further research may include the application of other multivariate normalizing transformations and data sets to construct the multiple non-linear regression model for estimating the agile testing efforts for small Web projects.

ACKNOWLEDGEMENTS

We are thankful to Tetiana Avdieieva for the data provided on the agile testing efforts for small Web projects.

REFERENCES

1. Olausson M., Rossberg J., Ehn J., Sköld M. Introduction to Agile Planning, Development, and Testing, *Pro Team Foundation Service*. Berkeley, CA, Apress, 2013, Chapter 2, pp. 9–19. DOI: 10.1007/978-1-4302-5996-1_2
2. Olausson M., Rossberg J., Ehn J., Sköld M. Agile Testing, *Pro Team Foundation Service*. Berkeley, CA, Apress, 2013, Chapter 19, pp. 19–33. DOI: 10.1007/978-1-4302-5996-1_19
3. Penmetsa J. R., Mohanty H., Mohanty J., Balakrishnan A. (eds). Agile Testing, *Trends in Software Testing*. Singapore, Springer, 2017, Chapter 2, pp. 19–33. DOI: 10.1007/978-981-10-1415-4_2
4. Nader-Rezvani N. Agile Quality Test Strategy, *An Executive's Guide to Software Quality in an Agile Organization.* Berkeley, CA, Apress, 2019, Chapter 7, pp. 121–138. DOI: 10.1007/978-1-4842-3751-9_7
5. Agile Testing – Principles, methods & advantages [Electronic resource]. Access mode: https://reqtest.com/testing-blog/agile-testing-principles-methods-advantages/
6. Boehm B. W., Abts C., Brown A. W. et al. Software Cost Estimation with COCOMO II, Upper Saddle River, NJ, Prentice Hall PTR, 2000, 506 p.
7. Lazić L., Đokić I., Milinković S. Challenges in estimating software testing effort, *Proceedings of INFOTEH-JAHORINA*, 2014, Vol. 13, pp. 637–642.
8. Bates D. M., Watts D. G. Nonlinear Regression Analysis and Its Applications. New York, John Wiley & Sons, 1988, 384 p. DOI:10.1002/9780470316757
9. Seber G.A.F., Wild C. J. Nonlinear Regression. New York, John Wiley & Sons, 1989, 768 p. DOI: 10.1002/0471725315
10. Ryan T.P. Modern regression methods. New York, John Wiley & Sons, 1997, 529 p. DOI: 10.1002/9780470382806
11. Johnson R. A., Wichern D. W. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007, 800 p.
12. Samprit Chatterjee, Jeffrey S. Somonoff. Chatterjee Samprit. Handbook of Regression Analysis. New York, John Wiley & Sons, 2012, 252 p. DOI:10.1002/9781118532843
13. Prykhodko N. V., Prykhodko S. B. Constructing the non-linear regression models on the basis of multivariate normalizing transformations, *Electronic modeling*, 2018, Vol. 40, No. 6, pp. 101–110. DOI: 10.15407/emodel.40.06.101
14. Prykhodko N. V., Prykhodko S. B. The non-linear regression model to estimate the software size of open source Java-based systems, *Radio Electronics, Computer Science, Control,* 2018, No. 3 (46), pp. 158–166. DOI 10.15588/1607-3274-2018-3-17
15. Prykhodko S., Prykhodko N., Makarova L. et al. Detecting Outliers in Multivariate Non-Gaussian Data on the basis of Normalizing Transformations / [S. Prykhodko, *// Electrical and Computer Engineering : the 2017 IEEE First Ukraine Conference (UKRCON) «Celebrating 25 Years of IEEE Ukraine Section»*. Kyiv, Ukraine, May 29–June 2, 2017, proceedings. Kyiv, IEEE, 2017, pp. 846–849. DOI: 10.1109/UKRCON.2017.8100366
16. Mardia K. V. Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 1970, Vol. 57, pp. 519–530. DOI: 10.1093/biomet/57.3.519

Received 11.01.2019.
Accepted 06.03.2019.

УДК 004.412:519.237.5

МНОЖИННА НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ ТРУДОМІСТКОСТІ AGILE ТЕСТУВАННЯ ДЛЯ МАЛИХ ВЕБ-ПРОЕКТІВ

Приходько Н. В. – канд. екон. наук, доцент, доцент кафедри фінансів Національного університету кораблебудування імені адмірала Макарова, Миколаїв, Україна.

Приходько С. Б. – д-р техн. наук, професор, завідувач кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова, Миколаїв, Україна.

АНОТАЦІЯ

Актуальність. Оцінювання трудомісткості тестування програмного забезпечення є однією з важливих проблем у розробці програмного забезпечення та життєвому циклі тестування програмного забезпечення. Об'єктом дослідження є процес оцінювання трудомісткості agile тестування для малих веб-проектів. Предметом дослідження є моделі множинної регресії для оцінювання трудомісткості agile тестування для малих веб-проектів.

Мета. Метою роботи є створення моделі множинної нелінійної регресії для оцінювання трудомісткості agile тестування для малих веб-проектів на основі багатовимірного нормалізуючого перетворення Джонсона.

Метод. Модель, довірчі інтервали та інтервали передбачення багатовимірної нелінійної регресії для оцінювання трудомісткості agile тестування для малих веб-проектів побудовані на основі багатовимірного нормалізуючого перетворення Джонсона для негаусівських даних за допомогою відповідних методів. Методи побудови моделей, рівнянь, довірчих інтервалів і інтервалів передбачення нелінійних регресій засновані на багатовимірному нелінійному регресійному аналізі з використанням багатовимірних нормалізуючих перетворень. Розглянуто відповідні методи. Ці методи дозволяють враховувати кореляцію між випадковими величинами в разі нормалізації багатовимірних негаусівських даних. Загалом, це призводить до зменшення середньої величини відносної похибки, ширини довірчих інтервалів і інтервалів передбачення в порівнянні з лінійними моделями та нелінійними моделями, побудованими з використанням одновимірних нормалізуючих перетворень.

Результати. Здійснено порівняння побудованої моделі з моделями лінійної регресії та нелінійними регресіями на основі десяткового логарифму та одновимірного перетворення Джонсона.

Висновки. Модель нелінійної регресії для оцінювання трудомісткості agile тестування для малих веб-проектів побудована на основі багатовимірного перетворення Джонсона для сімейства S_B. Ця модель в порівнянні з іншими регресійній моделі (як лінійними, так і нелінійними) має менше значення середньої величини відносної похибки, менші ширини довірчих інтервалів і інтервалів передбачення. Перспективи подальших досліджень можуть включати застосування інших багатовимірних нормалізують перетворень і наборів даних для побудови моделі нелінійної регресії для оцінювання трудомісткості agile тестування для малих веб-проектів.

DOI 10.15588/1607-3274-2019-2-17

КЛЮЧОВІ СЛОВА: agile тестування, оцінювання, трудомісткість тестування, Веб проект, модель множинної нелінійної регресії, багатовимірне нормалізуюче перетворення, негаусівські дані.

УДК 004.412:519.237.5

МНОЖЕСТВЕННАЯ НЕЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ ДЛЯ ОЦЕНКИ ТРУДОЕМКОСТИ AGILE ТЕСТИРОВАНИЯ ДЛЯ МАЛЫХ ВЕБ-ПРОЕКТОВ

Приходько Н. В. – канд. экон. наук, доцент, доцент кафедры финансов Национального университету кораблестроения имени адмирала Макарова, Николаев, Украина.

Приходько С. Б. – д-р техн. наук, профессор, заведующий кафедрой программного обеспечения автоматизированных систем Национального университету кораблестроения имени адмирала Макарова, Николаев, Украина.

АННОТАЦИЯ

Актуальность. Оценивания трудоемкости тестирования программного обеспечения является одной из важных проблем в разработке программного обеспечения и жизненном цикле тестирования программного обеспечения. Объектом исследования является процесс оценки трудоемкости agile тестирования для малых веб-проектов. Предметом исследования являются модели множественной регрессии для оценки трудоемкости agile тестирования для малых веб-проектов.

Цель. Цель работы – создание модели нелинейной регрессии для оценки трудоемкости agile тестирования для малых веб-проектов на основе многомерного нормализирующего преобразования Джонсона.

Метод. Модель, доверительные интервалы и интервалы прогнозирования многомерной нелинейной регрессии для оценки трудоемкости agile тестирования для малых веб-проектов построены на основе многомерного нормализирующего преобразования Джонсона для негауссовских данных с помощью соответствующих методов. Методы построения моделей, уравнений, доверительных интервалов и интервалов предсказания нелинейных регрессий основаны на многократном нелинейном регрессионном анализе с использованием многомерных нормализующих преобразований. Рассмотрены соответствующие методы. Методы позволяют учитывать корреляцию между случайными величинами в случае нормализации многомерных негауссовских данных. В общем, это приводит к уменьшению средней величины относительной погрешности, ширины доверительных интервалов и интервалов предсказания по сравнению с линейными моделями и нелинейными моделями, построенными с использованием одномерных нормализующих преобразований.

Результаты. Проведено сравнение построенной модели с линейной моделью и нелинейными регрессионными моделями на основе десятичного логарифма и одномерного преобразования Джонсона.

Выводы. Модель нелинейной регрессии для оценки трудоемкости agile тестирования для малых веб-проектов построена на основе многомерного преобразования Джонсона для семейства S_B. Эта модель по сравнению с другими регрессионными моделями (как линейными, так и нелинейными) имеет меньшее значение средней величины относительной погрешности, меньшие ширины доверительных интервалов и интервалов предсказания. Перспективы дальнейших исследований могут включать применение других многомерных нормализующих преобразований и наборов данных для построения модели нелинейной регрессии для оценки трудоемкости agile тестирования для малых веб-проектов.

КЛЮЧЕВЫЕ СЛОВА: agile тестирование, оценка, трудоемкость тестирования, Веб проект, модель множественной нелинейной регрессии, многомерное нормализующее преобразования, негаусивськи данные.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Introduction to Agile Planning, Development, and Testing / [M. Olausson, J. Rossberg, J. Ehn, M. Sköld] // Pro Team Foundation Service. – Berkeley, CA: Apress, 2013. – Chapter 2. – P. 9–19. DOI: 10.1007/978-1-4302-5996-1_2
2. Agile Testing / [M. Olausson, J. Rossberg, J. Ehn, M. Sköld] // Pro Team Foundation Service. – Berkeley, CA: Apress, 2013. – Chapter 19. – P. 19–33. DOI: 10.1007/978-1-4302-5996-1_19
3. Agile Testing / [J. R. Penmetsa, H. Mohanty, J. Mohanty, A. Balakrishnan (eds).] // Trends in Software Testing. – Singapore : Springer, 2017. – Chapter 2. – P. 19–33. DOI: 10.1007/978-981-10-1415-4_2
4. Nader-Rezvani N. Agile Quality Test Strategy / N. Nader-Rezvani // An Executive's Guide to Software Quality in an Agile Organization. – Berkeley, CA: Apress, 2019. – Chapter 7.– P. 121–138. DOI: 10.1007/978-1-4842-3751-9_7
5. Agile Testing – Principles, methods & advantages [Electronic resource]. – Access mode: https://reqtest.com/testing-blog/agile-testing-principles-methods-advantages/
6. Boehm B. W. Software Cost Estimation with COCOMO II / [B. W. Boehm, C. Abts, A. W. Brown et al.]. – Upper Saddle River, NJ: Prentice Hall PTR, 2000. – 506 p.
7. Lazić L. Challenges in estimating software testing effort / L. Lazić, I. Đokić, S. Milinković // Proceedings of INFOTEH-JAHORINA. – 2014. – Vol. 13. – P. 637–642.
8. Bates D. M. Nonlinear Regression Analysis and Its Applications / D. M. Bates, D. G. Watts. – New York : John Wiley & Sons, 1988. – 384 p. DOI:10.1002/9780470316757
9. Seber G.A.F. Nonlinear Regression / G.A.F. Seber, C.J. Wild. – New York : John Wiley & Sons, 1989. – 768 p. DOI: 10.1002/0471725315
10. Ryan T. P. Modern regression methods / T. P. Ryan. – New York : John Wiley & Sons, 1997. – 529 p. DOI: 10.1002/9780470382806
11. Johnson R. A. Applied Multivariate Statistical Analysis / R. A. Johnson, D. W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.
12. Chatterjee Samprit. Handbook of Regression Analysis / Samprit Chatterjee, Jeffrey S. Somonoff. – New York : John Wiley & Sons, 2012. – 252 p. DOI:10.1002/9781118532843
13. Prykhodko N. V. Constructing the non-linear regression models on the basis of multivariate normalizing transformations / N. V. Prykhodko, S. B. Prykhodko // Электронное моделирование. – 2018. – Т. 40, № 6. – С. 101–110. DOI: 10.15407/emodel.40.06.101
14. Prykhodko N. V. The non-linear regression model to estimate the software size of open source Java-based systems / N. V. Prykhodko, S. B. Prykhodko // Радіоелектроніка, інформатика, управління. – 2018. – № 3 (46). – С. 158–166. DOI 10.15588/1607-3274-2018-3-17
15. Prykhodko S. Detecting Outliers in Multivariate Non-Gaussian Data on the basis of Normalizing Transformations / [S. Prykhodko, N. Prykhodko, L. Makarova et al.] // Electrical and Computer Engineering : the 2017 IEEE First Ukraine Conference (UKRCON) «Celebrating 25 Years of IEEE Ukraine Section», Kyiv, Ukraine, May 29 – June 2, 2017 : proceedings. – Kyiv : IEEE, 2017. – P. 846–849. DOI: 10.1109/UKRCON.2017.8100366
16. Mardia K.V. Measures of multivariate skewness and kurtosis with applications / K. V. Mardia // Biometrika. – 1970. – Vol. 57. – P. 519–530. DOI: 10.1093/biomet/57.3.519

DOI 10.15588/1607-3274-2019-2-17