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 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

$b$ is estimator for vector of linear regression equation parameters, $b = [b_1, b_2, \ldots, b_k]^T$;
$\hat{b}_i$ 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,
$P = [Y, X_1, X_2, \ldots, X_k]^T$;
$R^2$ is a multiple coefficient of determination;
$S_N$ is a sample covariance matrix, $S_N = [S_y]$;
$S$ is a sample covariance matrix; $S = [S_{x}]$;
$S_{11}$, $S_{22}$, $S_{12}$ are sample covariance and correlations;
$\mu$ is a sample mean of the values of the variable $Z$;
$\alpha$ is a significance level;
$\sigma$ is a standard deviation of the variable $Z$.

$T$ is a Gaussian random vector,
$T = [Z_Y, Z_1, Z_2, \ldots, Z_k]^T$;
$t_{\nu/2, \nu}$ is a quantile of student’s $t$-distribution with $\nu$ 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;
$Z_X^*$ is a matrix of centered regressors that contains the values $Z_{i1} - \bar{Z}_1$, $Z_{i2} - \bar{Z}_2$, $\ldots$, $Z_{ik} - \bar{Z}_k$;
$\left[Z_X^*\right]^T$ is a transpose of $Z_X^*$;
$z_X^*$ is a vector with components $Z_{i1} - \bar{Z}_1$, $Z_{i2} - \bar{Z}_2$, $\ldots$, $Z_{ik} - \bar{Z}_k$ for $i$-row;
$\left[z_X^*\right]^T$ is a transpose of $z_X^*$;
$\bar{Z}_Y$ is a prediction linear regression equation result;
$\alpha$ is a significance level;
\[ \beta_1 \text{ is a multivariate skewness;} \]
\[ \beta_2 \text{ is a multivariate kurtosis;} \]
\[ \gamma \text{ is a vector of parameters of the Johnson multivariate translation, } \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_k)^T; \]
\[ \varepsilon \text{ is a Gaussian random variable which defines residuals, } \varepsilon \sim \mathcal{N}(0,1); \]
\[ \eta \text{ is a vector of parameters of the Johnson multivariate translation, } \eta = \text{diag}(\eta_1, \eta_1, \ldots, \eta_k); \]
\[ \lambda \text{ is a vector of parameters of the Johnson multivariate translation, } \lambda = \text{diag}(\lambda_1, \lambda_1, \ldots, \lambda_k); \]
\[ \nu \text{ is a number of degrees of freedom;} \]
\[ \Sigma \text{ is a covariance matrix, } \Sigma = \left[\Sigma_{ij}\right]; \]
\[ \phi \text{ is a vector of parameters of the Johnson multivariate translation, } \phi = (\phi_1, \phi_1, \phi_2, \ldots, \phi_k)^T; \]
\[ \psi \text{ is a vector of multivariate normalizing transformation, } \psi = (\psi_1, \psi_2, \ldots, \psi_k)^T. \]

**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 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 \( \mathbf{P} = \begin{bmatrix} Y, X_1, X_2, \ldots, X_k \end{bmatrix}^T \) to Gaussian random vector \( \mathbf{T} = \begin{bmatrix} Z_Y, Z_1, Z_2, \ldots, Z_k \end{bmatrix}^T \) is given by

\[
\mathbf{T} = \psi(\mathbf{P}) \tag{1}
\]

and the inverse transformation for (1)

\[
\mathbf{P} = \psi^{-1}(\mathbf{T}) \tag{2}
\]

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.
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 transformants are more accurate than other methods in the same case. For normalizing the multivariate non-Gaussian data, we use the Johnson translation. In our case the Johnson normalizing translation is given by [14]

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

where \( h(y, y_1, \ldots, y_k) = \{h_1(y_1), \ldots, h_k(y_k)\}^T \); \( h_1, \ldots, h_k \) is one of the translation functions.

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]

\[
Z_f = \bar{Z}_f + \varepsilon = \bar{Z}_f + \left[ Z_X^\gamma \right] \mathbf{b} + \varepsilon.
\]  

(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]

\[
Y = \psi_Y^{-1}\left[ \bar{Z}_f + \left[ Z_X^\gamma \right] \mathbf{b} + \varepsilon \right].
\]  

(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]:

\[
\psi_Y^{-1}\left[ \bar{Z}_f \pm t_{\nu/2, N} S_{\bar{Z}_f} \frac{1}{N} \left[ Z_X^\gamma \right] \left[ Z_X^\gamma \right]^T \left[ Z_X^\gamma \right] \right]^{1/2}.
\]

(5)

where \( S_{\bar{Z}_f} = \frac{1}{N} \sum_{i=1}^{N} \left( Z_{f_i} - \bar{Z}_f \right)^2 \), \( \nu = N - k - 1 \); \( \left[ Z_X^\gamma \right] \left[ Z_X^\gamma \right]^T \) is the \( k \times k \) matrix

\[
\left[ Z_X^\gamma \right] \left[ Z_X^\gamma \right]^T = \begin{pmatrix}
S_{Z_1} & S_{Z_1} & \ldots & S_{Z_1} \\
S_{Z_2} & S_{Z_2} & \ldots & S_{Z_2} \\
\vdots & \vdots & \ddots & \vdots \\
S_{Z_k} & S_{Z_k} & \ldots & S_{Z_k}
\end{pmatrix},
\]

and \( S_{Z_q} = \sum_{i=1}^{N} \left( Z_{q_i} - \bar{Z}_q \right) \), \( q, r = 1, 2, \ldots, k \).

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]:

\[
\psi_Y^{-1}\left[ \bar{Z}_f \pm t_{\alpha/2, N} S_{\bar{Z}_f} \frac{1}{N} \left[ Z_X^\gamma \right] \left[ Z_X^\gamma \right]^T \left( \mathbf{x}_k \right) \right]^{1/2}.
\]

(6)

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

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

where \( h(y, y_1, \ldots, y_k) = \{h_1(y), \ldots, h_k(y)\}^T \); \( h_1, \ldots, h_k \) is one of the translation functions.

Here \( y = (X - \varphi)/\lambda \); \( Arsh(y) = \ln(y + \sqrt{y^2 + 1}) \). 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 data rows are: 4.09443, 5.47043, 1.09282, 3.13761, 1.04794, 0.97350, 0.70189, 0.64464, 0.37266, 1.95622, 0.94564, 2.35215, 327.313, 5438.99, 130.495 and 110.210. The sample covariance

© Prykhodko N. V., Prykhodko S. B., 2019
DOI 10.15588/1607-3274-2019-2-17

160
The squared MD values for 40 data rows indicate that the data of system 41 is multivariate outlier too, 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

| №  | Y   | X_1 | X_2 | X_3 | Squared MD |
|----|-----|-----|-----|-----|------------|
| 1  | 1.33| 4   | 11  | 14  | 5.69       |
| 2  | 3.51| 1    | 2   | 7   | 5.75       |
| 3  | 3.17| 10   | 4   | 8   | 2.67       |
| 4  | 1.53| 5    | 4   | 8   | 4.30       |
| 5  | 2.32| 9    | 12  | 3   | 4.71       |
| 6  | 3.13| 10   | 8   | 9   | 2.56       |
| 7  | 2.03| 6    | 16  | 7   | 3.92       |
| 8  | 3.22| 11   | 11  | 6   | 1.31       |
| 9  | 2.73| 10   | 16  | 5   | 2.83       |
| 10 | 4.65| 16   | 13  | 7   | 0.75       |
| 11 | 5.52| 21   | 12  | 4   | 7.56       |
| 12 | 2.75| 7    | 23  | 12  | 4.28       |
| 13 | 6.93| 25   | 6   | 13  | 2.68       |
| 14 | 3.73| 14   | 29  | 4   | 2.44       |
| 15 | 5.68| 18   | 22  | 7   | 0.62       |
| 16 | 7.12| 25   | 11  | 11  | 1.11       |
| 17 | 3.24| 10   | 31  | 9   | 2.49       |
| 18 | 4.05| 12   | 35  | 6   | 2.63       |
| 19 | 4.72| 16   | 52  | 7   | 0.58       |
| 20 | 3.49| 13   | 41  | 4   | 2.52       |
| 21 | 6.03| 18   | 24  | 18  | 0.64       |
| 22 | 4.13| 12   | 40  | 13  | 2.35       |
| 23 | 10.18|33  | 15  | 23  | 0.62       |
| 24 | 9.95| 36   | 29  | 12  | 1.21       |
| 25 | 8.87| 24   | 36  | 32  | 1.37       |
| 26 | 16.53|51  | 3   | 45  | 7.11       |
| 27 | 12.45|44  | 37  | 18  | 0.97       |
| 28 | 15.56|46  | 25  | 19  | 2.84       |
| 29 | 19.47|57  | 7   | 40  | 3.92       |
| 30 | 11.23|29  | 48  | 47  | 2.78       |
| 31 | 8.21| 29   | 79  | 10  | 2.30       |
| 32 | 19.95|50  | 6   | 90  | 8.33       |
| 33 | 16.16|44  | 53  | 61  | 3.12       |
| 34 | 8.83| 33   | 110 | 8   | 6.14       |
| 35 | 12.97|47  | 94  | 16  | 3.30       |
| 36 | 21.32|94  | 48  | 37  | 4.73       |
| 37 | 20.97|83  | 61  | 62  | 8.87       |
| 38 | 38.22|127  |25  | 78  | 4.77       |
| 39 | 26.48|111  |71  | 54  | 5.50       |
| 40 | 48.22|173  |120 | 60  | 11.73      |
| 41 | 7.07| 11   | 13  | 51  | 20.43      |
for the normalized data and the decimal logarithm transformation

\[ Y = 10^{a + b_1 X_1 + b_2 X_2 + b_3 X_3}, \]

(10)

where the estimators for parameters are: \( b_0 = -0.4500, \)

\( b_1 = 0.78887, \)

\( b_2 = 0.00176, \)

\( b_3 = 0.18782. \)

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:

\[ \hat{\gamma}_Y = 1.72684, \quad \hat{\gamma}_1 = 1.78876, \quad \hat{\gamma}_2 = 1.2346, \quad \hat{\gamma}_3 = 1.1127, \]

\[ \hat{\eta}_1 = 0.73211, \quad \hat{n}_1 = 0.72296, \]

\[ \hat{n}_2 = 0.53444, \quad \hat{\phi}_1 = 1.1292, \quad \hat{\phi}_1 = 3.3512, \quad \hat{\phi}_2 = 1.1717, \]

\[ \hat{\phi}_3 = 2.90, \quad \hat{\lambda}_Y = 64.039, \quad \hat{\lambda}_1 = 237.48, \quad \hat{\lambda}_2 = 142.95, \]

\[ \hat{\lambda}_3 = 92.745, \quad \hat{b}_0 = 0, \quad \hat{b}_1 = 0.86376, \quad \hat{b}_2 = 0.00240 \]

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 \( \varepsilon \) 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 transformation |
|    | \( \hat{\gamma}_1 \) | \( \hat{\gamma}_1 \) | \( \hat{\gamma}_1 \) | \( \hat{\gamma}_1 \) | \( \hat{\gamma}_1 \) |
|    | \( \hat{\gamma}_2 \) | \( \hat{\gamma}_2 \) | \( \hat{\gamma}_2 \) | \( \hat{\gamma}_2 \) | \( \hat{\gamma}_2 \) |
|    | \( \hat{\gamma}_3 \) | \( \hat{\gamma}_3 \) | \( \hat{\gamma}_3 \) | \( \hat{\gamma}_3 \) | \( \hat{\gamma}_3 \) |
|    | \( \hat{\eta}_1 \) | \( \hat{\eta}_1 \) | \( \hat{\eta}_1 \) | \( \hat{\eta}_1 \) | \( \hat{\eta}_1 \) |
|    | \( \hat{\eta}_2 \) | \( \hat{\eta}_2 \) | \( \hat{\eta}_2 \) | \( \hat{\eta}_2 \) | \( \hat{\eta}_2 \) |
|    | \( \hat{\eta}_3 \) | \( \hat{\eta}_3 \) | \( \hat{\eta}_3 \) | \( \hat{\eta}_3 \) | \( \hat{\eta}_3 \) |
|    | \( \hat{\phi}_1 \) | \( \hat{\phi}_1 \) | \( \hat{\phi}_1 \) | \( \hat{\phi}_1 \) | \( \hat{\phi}_1 \) |
|    | \( \hat{\phi}_2 \) | \( \hat{\phi}_2 \) | \( \hat{\phi}_2 \) | \( \hat{\phi}_2 \) | \( \hat{\phi}_2 \) |
|    | \( \hat{\phi}_3 \) | \( \hat{\phi}_3 \) | \( \hat{\phi}_3 \) | \( \hat{\phi}_3 \) | \( \hat{\phi}_3 \) |
|    | \( \hat{\lambda}_Y \) | \( \hat{\lambda}_Y \) | \( \hat{\lambda}_Y \) | \( \hat{\lambda}_Y \) | \( \hat{\lambda}_Y \) |
|    | \( \hat{\lambda}_1 \) | \( \hat{\lambda}_1 \) | \( \hat{\lambda}_1 \) | \( \hat{\lambda}_1 \) | \( \hat{\lambda}_1 \) |
|    | \( \hat{\lambda}_2 \) | \( \hat{\lambda}_2 \) | \( \hat{\lambda}_2 \) | \( \hat{\lambda}_2 \) | \( \hat{\lambda}_2 \) |
|    | \( \hat{\lambda}_3 \) | \( \hat{\lambda}_3 \) | \( \hat{\lambda}_3 \) | \( \hat{\lambda}_3 \) | \( \hat{\lambda}_3 \) |

© Prykhodko N. V., Prykhodko S. B., 2019
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 $\mathbf{X}$ 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.

| No | LB | UB | LB | UB |
|----|----|----|----|----|
| Table 3 – The bounds of the prediction intervals | Bounds for linear regression | Bounds for multiple non-linear regression |
|    | decimal logarithm transformation | Johnson multivariate transformation |
| 1  | -1.174 | 4.205 | 1.186 | 1.606 |
| 2  | 0.689 | 6.096 | 2.906 | 3.966 |
| 3  | 0.548 | 5.939 | 2.782 | 3.756 |
| 4  | -0.942 | 4.431 | 1.417 | 1.912 |
| 5  | -0.104 | 5.278 | 2.134 | 2.879 |
| 6  | 0.637 | 6.010 | 2.852 | 3.833 |
| 7  | -0.456 | 4.906 | 1.817 | 2.454 |
| 8  | 0.615 | 5.988 | 2.857 | 3.829 |
| 9  | 0.300 | 5.664 | 2.562 | 3.435 |
| 10 | 1.855 | 7.224 | 3.955 | 5.297 |
| 11 | 2.740 | 8.149 | 4.377 | 5.955 |
| 12 | 0.188 | 5.550 | 2.264 | 3.080 |
| 13 | 4.430 | 9.826 | 6.288 | 8.473 |
| 14 | 1.133 | 6.492 | 3.200 | 4.310 |
| 15 | 2.319 | 7.669 | 4.344 | 5.818 |
| 16 | 4.266 | 9.647 | 6.117 | 8.201 |
| 17 | 0.625 | 5.984 | 2.860 | 3.847 |
| 18 | 0.830 | 6.195 | 3.064 | 4.112 |
| 19 | 1.849 | 7.198 | 3.961 | 5.306 |
| 20 | 0.879 | 6.264 | 3.018 | 4.070 |
| 21 | 3.245 | 8.582 | 5.186 | 6.951 |
| 22 | 1.402 | 6.788 | 3.553 | 4.768 |
| 23 | 17.333 | 12.841 | 8.759 | 11.722 |
| 24 | 5.753 | 11.713 | 7.231 | 9.756 |
| 25 | 13.072 | 18.568 | 13.802 | 18.916 |
| 26 | 9.242 | 14.968 | 10.298 | 14.875 |
| 27 | 12.059 | 17.502 | 12.810 | 17.214 |
| 28 | 14.049 | 19.520 | 14.853 | 20.103 |
| 29 | 8.038 | 13.692 | 8.997 | 12.224 |
| 30 | 4.947 | 10.535 | 6.763 | 9.108 |
| 31 | 16.197 | 22.512 | 15.500 | 21.234 |
| 32 | 12.607 | 18.394 | 13.141 | 17.825 |
| 33 | 5.486 | 11.460 | 7.160 | 9.712 |
| 34 | 9.554 | 15.231 | 10.809 | 14.575 |
| 35 | 22.275 | 27.831 | 21.839 | 29.433 |
| 36 | 21.818 | 27.368 | 21.804 | 29.432 |
| 37 | 33.163 | 39.103 | 31.740 | 42.984 |
| 38 | 27.386 | 33.188 | 26.713 | 36.101 |
| 39 | 41.985 | 48.377 | 38.533 | 52.537 |
| 40 | 37.588 | 49.717 | 34.553 | 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

© Prykhodko N. V., Prykhodko S. B., 2019
DOI 10.15588/1607-3274-2019-2-17

163
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 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 $\beta_1$ and kurtosis $\beta_2$ are estimated for 40 data rows 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. The measures $\beta_1$ and $\beta_2$ allow one to test two hypotheses that are compatible with the assumption of multivariate normality. The estimator of multivariate skewness given by [16]

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

(11)

The estimator of multivariate kurtosis given by [16]

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

(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 $\beta_2 = m(m+2)$ 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.

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 effort 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

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.
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 Avidieieva for the data provided on the agile testing efforts for small Web projects.

REFERENCES

1. Olausson M., Rossberg J., Ehj I., 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., Ehj I., 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. Pemmetsa I. 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://request.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.

© Prykhodko N. V., Prykhodko S. B., 2019
DOI 10.15588/1607-3274-2019-2-17

165
МНОЖЕСТВЕННАЯ НЕЛИНЕЙНАЯ РЕГРЕSSIONНСЯ МОДЕЛЬ ДЛЯ ОЦЕНКИ ТРУДОЕМКОСТИ AGILE ТЕСТИРОВАНИЯ ДЛЯ МАЛЫХ ВЕБ-ПРОЕКТОВ

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

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

АННОТАЦИЯ

Актуальность. Оценивание 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 нормализующего преобразования Джонсона.

Метод. Модель, доверительные интервалы и интервалы прогнозирования многомерной нелинейной регрессии для оценки трудоемкости agile тестирования для малых веб-проектов построена на основе многомерного нормализующего преобразования Джонсона для негауссовских данных с помощью соответствующих методов. Методы построения моделей, уравнений, доверительных интервалов и интервалов прогнозирования 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 прогнозирования 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 семейства многомерного нормализующего преобразования Джонсона. Выводы. Модель нелинейной регрессии для оценки трудоемкости agile тестирования для малых веб-проектов построена на основе многомерного нормализующего преобразования Джонсона для семейства $S_{P}$. Эта модель по сравнению с другими регрессионными моделями (как линейными, так и нелинейными) имеет меньше значение средней величины относительной погрешности, меньше ширины доверительных интервалов и интервалов прогнозирования. 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 множинної нелінійної регресії, багатовимірне нормалююче перетворення, негаусівські дані.

ЛІТЕРАТУРА

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. Pennetta, H. Moshanty, H. 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://request.com/testing/blog/agile-testing-principles-methods-advantages/
6. Boehm B. Software Cost Estimation with COCOMO II / [B. W. Boehm, C. Abis, 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. – New York : Prentice Hall, 2007. – 700 p. DOI: 10.1002/978111832843
12. Chattor Jeere Samprit. Handbook of Regression Analysis / S. B. Prakhodko // / Electronic modelisation. – 2018. – T. 40, № 6. – C. 101–110. DOI: 10.15407/emodel.40.06.101
13. Prakhodko N. V. Constructing the non-linear regression models on the basis of multivariate normalizing transformations / N. V. Prakhodko, S. B. Prakhodko // / Radio engineering, computer science, control. 2019. № 57. – P. 519–530. DOI: 10.1093/biomet/57.3.519

Приходько Н. В., Приходько С. Б., 2019
DOI 10.15588/1607-3274-2019-2-17

166