STATISTICAL INFERENCE IN PRINCIPAL COMPONENT ANALYSIS BASED ON STATISTICAL THEORY

Yan Liu

Xi'an Eurasia University
Xi'an, Shanxi
China
e-mail: liuyanl023@sina.com

Abstract. Principal component analysis is a diversified statistical method, while statistical inference is the major research subject in modern statistics, whose theories and methods have comprised the core content of mathematical statistics. According to the relevant knowledge of statistical theory and based on a large sample size, this study explored the statistical inference problem when population followed normal distribution. Besides, statistical methods were applied to further analyze the statistical inference problems in principle component analysis under the condition of population with non-normal distribution or small sample size. First, principal component analysis was performed on parameter estimation and hypothesis testing on the condition that population followed multivariate normal distribution. Then under the condition of complex distribution of population, simulated sampling statistical inference method, i.e., Bootstrap method, was used to do interval estimation and discuss over other statistical inference problems of the characteristic values of the correlation coefficient matrixes in principle component analysis, and then the defects of Bootstrap method were adjusted using Bayesian theory.
Keywords: Bootstrap method; Bayesian estimation; principle component analysis; hypothesis testing.

1. Introduction

Principle component analysis can process a problem of high-latitude space by transforming it into a problem of low-latitude space [1], [2], thus to make the problem simpler and more visualized. Moreover, in the process of principal component analysis, each principal component will generate its weight automatically [3], which resists the influence of human factors in evaluation process and makes evaluation results more objective. Principal component analysis can be divided into principal component analysis for population and principal component analysis for samples [4]. In the analysis of practical problems, population covariance matrix and correlation coefficient matrix are generally unknown; therefore, sample covariance matrix and correlation coefficient matrix need to be calculated for the evaluation or hypothesis testing of samples [5]. When practical problems are analyzed using principal component analysis, principal component which can cover population information is selected through changing variable values and moreover further analysis is made on this basis. But, the premise of principal component selection is to suppose that these principal components can cover population information [6]. Therefore, corresponding hypothesis testing is needed to determine whether the processing of follow-up question is feasible.

In terms of the current situation of the development of statistical theory, studies of the statistical inference problem of principal component analysis mainly focus on the interval estimation of characteristic value of covariance matrix, the applicability test of principal component analysis, the test of principal component quantity selection, the test of characteristic root of population covariance matrix, etc [7]. With the development of modern computer technology in recent years, some new methods and ideas, for example, Bootstrap method, Logistic regression method and Jackknife estimation method, are added into statistical inference processing new technology [8, 9]. Bootstrap method and Jackknife estimation methods are two major simulation methods of statistical inference, which have been explored by many scholars in China and abroad. On the basis of previous studies, this study attempted to create resampling samples using Bootstrap method and then made interval estimation on the characteristic root of population covariance matrix based on the obtained samples. Finally, the existing loophole and defects of Bootstrap method were fixed and overcome using Bayesian theory.

2. Principal component analysis under multivariate normal distribution

2.1. Parameter estimation

Parameter estimation mainly includes confidence interval of characteristic root of population covariance matrix, confidence region of characteristic vector and combination confidence region of principal component score.

(1) Confidence interval of characteristic root of population covariance matrix

In the following, principal component analysis performed on parameter estimation on the premise that population followed multivariate normal distribution. Firstly, the approximate distribution of covariance matrix characteristic root of samples was deduced and then confidence interval with a confidence level of $1-\alpha$ was constructed.

If population X followed normal distribution, then the original sample $X_{(i)} = (x_{i1}, x_{i2}, ..., x_{in})$ and $X \sim N_n(\mu, \Sigma)$. The covariance matrix S of samples was the maximum likelihood estimation of population covariance matrix Σ, and, moreover, we have [10]:

$$S = \frac{1}{m-1}\sum_{i=1}^{m}(X_{(i)} - \overline{X}_{(*)})(X_{(i)} - \overline{X}_{(*)})' = (S_{ij})_{n\times n}.$$

As the maximum likelihood estimation result of Σ characteristic root was $\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_n$, $(\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_n)$ was determined as being in almost normal distribution. As likelihood function of λ_1 was $L(\lambda_i) = (\lambda_i)^{-\frac{n}{2}}\, e^{\frac{a_i' v a_i}{2\lambda_i}}$ and, moreover, $a_i'\Sigma a_i = \lambda_i$, $v \sim W_n(m-1, \Sigma)$, therefore $a_i' \sim \lambda_i \chi^2(m-1)$.

Because of the mutual independence among $\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_n$ and the almost normal distribution of $(\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_n)$, we have:

$$\sqrt{m-2}(\hat{\lambda}_i - \lambda_i) \to N_n\left(0.2\ \frac{\lambda_i^2}{m}\right).$$

Therefore, the confidence interval with a confidence level of $1-\alpha$ was:

$$\frac{\hat{\lambda}_i}{1+\sqrt{\frac{2}{m-2}\,Z_{\frac{1-\alpha}{2}}}} \leq \lambda_i \leq \frac{\hat{\lambda}_i}{1-\sqrt{\frac{2}{m-2}\,Z_{\frac{1-\alpha}{2}}}}.$$

It can be known from the above equation that, the increase of $\hat{\lambda}_i$ could widen the confidence interval of λ_i under the condition of a large sample size. However, when the characteristic value was relatively large, the confidence interval would be wide as well, even though the sample size was relatively large; at this moment, the estimation of characteristic root lost significance.

(2) Confidence region of characteristic vector

In the following content, parameter estimation of characteristic vector $a_1, a_2, ..., a_n$ was performed through building the confidence region of $a_1, a_2, ..., a_n$. If and moreover the maximum likelihood estimation result of Σ was $\hat{\Sigma} = S$, when the characteristic roots of Σ, i.e., $\lambda_1, \lambda_2, ..., \lambda_n$, satisfied the relationship that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_n$, then the corresponding unit orthotropic characteristic vectors were $\hat{a}_1, \hat{a}_2, ..., \hat{a}_n$ respectively and, moreover, $\hat{a}_i = (\hat{a}_{1i}, \hat{a}_{2i}, ..., \hat{a}_{ni})$. If $\hat{a}_{1i} \geq 0$, then $(\hat{a}_i, \hat{a}_2, ..., \hat{a}_n)$ was determined as being in almost normal distribution based on the maximum likelihood estimation. Suppose

$$A_i = \lambda_i \sum_{k=1}^{n} \frac{\lambda_k}{(\lambda_k - \lambda_i)^2}\ a_k a_k'$$

and, moreover, $\sqrt{n}(\hat{a}_i - a_i) \to N_n(0, A_i)$, then the confidence region with a confidence level of $1-\alpha$ was:

$$\frac{n^2(\hat{a}_i - a_i)(\hat{a}_i - a_i)'}{\lambda_i \sum_{k=1}^{n} \frac{\hat{\lambda}_k a + k a_k'}{(\hat{\lambda}_k - \hat{\lambda}_i)^2}} \leq \left(Z_{\frac{1-\alpha}{2}}\right)^2.$$

(3) Combination confidence region of principal component score

If population X followed normal distribution, then the original sample $X_{(i)} = (x_{i1}, x_{i2}, ..., x_{in})$ and, moreover, $X \sim N_n(\mu, \Sigma)$. If $\Sigma > 0$ and the maximum likelihood estimation result ofΣ was $hat\Sigma = S$, when characteristic roots of Σ, i.e., $\lambda_1, \lambda_2, ..., \lambda_n$, satisfied the relationship that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_n$, then the corresponding unit orthotropic characteristic vectors were $\hat{a}_1, \hat{a}_2, ..., \hat{a}_n$ and, moreover, $\hat{a}_i = (\hat{a}_{1i}, \hat{a}_{2i}, ..., \hat{a}_{ni})$. If $\hat{a}_{1i} \geq 0$, then the i-th principal component could be expressed as follows according to the principal component analysis

$$Y_i = a_i X = a_{i1} X_i + a_{i2} X_2 + \cdots + a_{in} X_n.$$

According to the confidence interval of characteristic root and the confidence region of characteristic vector of population covariance matrix obtained by parameter estimation previously, the combination confidence region of principal component score could be deduced.

According to $\sqrt{n}(\hat{a}_i - a_i) \to N_n(0, A_i)$, the principal component score approximately followed normal distribution [11]. Therefore, the following calculation was needed

$$E[\hat{Y}_i] = E[a_i' X_{(i)}] = E[a_i' V - \hat{a}_i' X_{(i)}] + E[\hat{a}_i' X_{(i)}] = \hat{a}_i' \mu_i$$
$$D[\hat{Y}] = D[a_i' X_{(i)}] = D[a_i' X_{(i)} - \hat{a}_i' X_{(i)}] = A_i \Sigma$$

It could be known that $\hat{Y}_i - \hat{a}_i' \mu_i \to N_n(0, A_i, \Sigma)$. Thus the confidence region of principal component score was:

$$\frac{n(\hat{Y} - \hat{a}_i'\mu_i)(\hat{Y} - \hat{a}_i'\mu_i)'}{\left[\hat{\lambda}_i \sum\limits_{k=1}^{n} \frac{\hat{\lambda}_k \hat{a}_k \hat{a}_k'}{(\hat{\lambda}_k - \hat{\lambda}_i)^2}\right] \hat{\Sigma}} \le (Z_{\frac{1-\alpha}{2}})^2$$

2.2. Hypothesis testing

Hypothesis testing mainly includes [12] hypothesis testing on the adaptability of principal component analysis, hypothesis testing of population covariance matrix characteristic root, and hypothesis testing on the selection of principal component number, etc.

(1) Hypothesis testing on the adaptability of principal component analysis

Hypothesis testing on the adaptability of principal component analysis, a common correlation test [13], aims at testing the hypothesis that the correlation coefficients of variables were all equal or not all equal to each other. If the correlation coefficients of n principal components were all equal, then it indicated that, characteristic roots of covariance matrix were all unequal [14]; at that moment, principal component analysis cannot be performed on n principal components. Therefore, the testing of correlation structure should be carried out. The following was an example of Kaiser-Meyer-Olkin (KMO) test.

The calculation formula of KMO [15] was:

$$\text{KMO} = \frac{\sum\limits_{j=1}^{k} \sum\limits_{i \neq j}^{k} r_{ij}^2}{\sum\limits_{j=1}^{k} \sum\limits_{i \neq j}^{k} r_{ij}^2 + \sum\limits_{j=1}^{k} \sum\limits_{i \neq j}^{k} \rho_{ij}^2}$$

where r_{ij} stands for simple correlation and ρ_{ij} stands for partial correlation coefficient. It can be seen from the above formula that, the value of KMO was $[0, 1]$, which fell out of the range of $[0, 0.5)$. Thus, it was unsuitable to carry out a principal component analysis.

(2) Hypothesis testing of population covariance matrix characteristic root

It can be known from statistical theory that, if matrix $A' = A$, then characteristic values of the matrix, i.e., $\lambda_1, \lambda_2, ..., \lambda_n$, were ranked in order of size. Suppose $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ and, moreover, $\gamma_1, \gamma_2, ..., \gamma_n$ as the standard orthogonal characteristic vectors corresponding to characteristic roots, then there is an A for any vector x

$$\max_{x \neq 0} \frac{x'Ax}{x'x} = \lambda_1, \quad \min_{x \neq 0} \frac{x'Ax}{x'x} = \lambda_n.$$

Suppose the covariance matrix of random vector($X = (X_1, X_2, ..., X_m)'$ was Σ, $\lambda_1, \lambda_2, ..., \lambda_n$ were characteristic roots, and $\gamma_1, \gamma_2, ..., \gamma_n$ were standard orthogonal characteristic vectors corresponding to characteristic roots of matrix A, then the i-th principal component was:

$$Y_i = \gamma_{1i}X_1 + \gamma_{2i}X_2 + \cdots + \gamma_{mi}X_m \ (i = 1, 2..., m).$$

At that moment, $\mathrm{Var}(Y_i) = \gamma_i'\Sigma\gamma_i = \lambda_1$ and $\mathrm{cov}(Y_i, Y_j) = \gamma'\Sigma\gamma_j = 0 \ (i \neq j)$. Standard characteristic vectors $\gamma_1, \gamma_2, ..., \gamma_n$ corresponding to nonzero characteristic values of covariance matrix Σ, i.e., $\lambda_1, \lambda_2, ..., \lambda_m$, were taken as coefficient vectors, that was:

$$Y_1 = \gamma_1'X, Y_2 = \gamma_2'X, ..., Y_m = \gamma_m'X,$$

where $Y_1, Y_2, ..., Y_m$ stand for the first principal component, the second principal component,..., and the m-th principal component of vector X. The necessary and sufficient conditions for that the separation of the principal components is equal to vector X were as follows. Firstly, $Y = u'X$ and $u'u = I$, i.e., u was an m-order orthogonal matrix. Secondly, principal components Y were uncorrelated to each other. Thirdly, m components of principal component were ranked according to the size of variance.

Thus X vector and principal component had the following relationship

$$\begin{aligned} Y &= u'X = [u_1', u_2', ..., u_m']^T \cdot X \\ &= \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mm} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} \gamma_1' \\ \gamma_2' \\ \vdots \\ \gamma_m' \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix} \end{aligned}$$

(3) Hypothesis testing of the selection of principal component number

In instance analysis, whether the selected principal components can satisfy the requirements of application needs to be determined at first; therefore, the selection of principal component number requires hypothesis testing. The selection of principal component number mainly relies on the size of contribution. The common testing methods include Bartlett test, mean method and empirical method [16].

Bartlett test: the method determines the number of principal component by testing whether the characteristic root is 0 [17]. If characteristics roots with low ranks had no statistically significant difference with 0, then they were not considered as principal components. Only principal components whose characteristic roots were not zero were selected.

Mean method: the mean value of characteristic roots, i.e., $\overline{\lambda}$, was calculated at first and then principal components whose characteristic roots were larger than $\overline{\lambda}$ were selected.

Empirical method: empirical method determines the number of principal component according to practical experience. Generally, a component with an accumulated contribution higher than 80% can be selected as a principal component. Suppose λ_i stands for the characteristic value of the i-th principal component, then the contribution rate of the i-th principal component y_i was expressed as $\lambda_i \sum_{i=1}^{m} \lambda_2$. Therefore, the accumulated contribution rate of the first m principal components should be expressed as:

$$\frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \times 100\% \quad (n > m).$$

Generally, to reduce the loss of information and variables as well as simplify problem, the value of m should be able to ensure over 80% of accumulated contribution rate.

It can be known from practical analysis condition that, Bartlett test is easy to be affected by the volume of samples. Many principals can be obtained if the volume of sample is large; otherwise, fewer principals can be obtained. Mean method is usually easy to be affected by extremum, which can make the obtained $\overline{\lambda}$ fail to reflect practical situation. Thus, the empirical method is adopted usually as it can reflect practical situation better and satisfy the requirements of application.

3. Discussion on the inference problem of principal component analysis using the Bootstrap method

The Bootstrap method mainly includes a parametric Bootstrap method and a nonparametric Bootstrap method [18]. The parametric Bootstrap method extracts samples according to population distribution function. The nonparametric Bootstrap method has no requirement on population distribution, i.e., sampling with replacement is performed to obtain samples when population distribution is unknown [19]. For example, when a parameter θ is selected for N experiments, an estimated value $\theta^{(n)}$ $(n = 1, 2, ..., N)$ can be obtained every time; then, some properties of θ including mean value, standard deviation and confidence interval [20] are estimated according to its distribution $F(\theta)$. The procedures of the method are as follows.

Figure 1: Sampling principle of the Bootstrap method

However, the distribution condition of population is unknown generally. Therefore, sampling needs to be carried out for the analysis of statistical inference problems using the nonparametric Bootstrap method. The parameter estimation of principal component analysis has been described before. The following content was about non-parametric Bootstrap method. As there is no need to know the distribution of population in advance when the nonparametric Bootstrap method is used, we suppose the size of samples was small and insufficient for the traditional statistical inference.

The specific procedures were as follows.

(1) Suppose the distribution of population X followed $F(x)$, the number of samples was m, and the original sample $X(i) = (x_{i1}, x_{i2}, ..., x_{ip})$ $(i = 1, 2, ..., n)$.

(2) The covariance matrix $\hat{\Sigma} = S$ of samples was calculated according to $X_{(i)}$, then P characteristic roots, i.e., $\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_p$, which were not equal with each other were obtained. The first principal component λ_1 was processed by parameter estimation.

(3) Sampling with replacement was carried out among m samples; the sampling repeated for m times or for B times to obtain a sample with a volume of B.

(4) Suppose the sample obtained after B times of sampling as $X_{(B)}$. The characteristic roots of the covariance matrix of samples, i.e., $\lambda_1^{(b)}, \lambda_2^{(b)}, ..., \lambda_p^{(b)}$ $(p = 1, 2, ..., B)$, could be obtained after the covariance and characteristic root were calculated.

(5) Frequency distribution of characteristic roots of group B samples was analyzed. The Bootstrap percentiles method was used to estimate the confidence interval under the confidence level of 1α.

4. Adjustment of Bootstrap based on Bayesian theory

4.1. The problems existing in Bootstrap method

Subsamples obtained in random sampling are very likely to be similar to the original samples or be regenerated samples which deviate greatly from the original samples when the traditional Bootstrap method is used to perform repeated random sampling with replacement, because the observed value of original sample has been determined [21], [22].

The defect will be more obvious when there are fewer original measured values. The phenomenon that probability concentrates on a certain observed value may result in the deviation of the finally estimated $R(X, F)$ distribution from real distribution, which severely affects the accuracy of prediction results.

On the other hand, subsamples obtained when repeated random sampling with replacement is performed on sample $X = (X_1, X_2, ..., X_n)$ are all samples points in original observation samples. At the moment, the properties of complete real distribution at non-original observed value points cannot be obtained and moreover the real distribution characteristics of $R(X, F)$ cannot be obtained, which further lowers the accuracy of prediction results.

4.2 The adjustment of Bootstrap method using the Bayesian theory

In reality, it is common that sample size is not large enough for statistical inference using traditional method. When the original sample is not less than 10, estimated result obtained using Bootstrap method is usually highly accurate. The advantage of the Bayesian idea in combination with the Bootstrap method is that it can solve the problem of parameter estimation in the situation of unknown and complex population distribution or small sample size. Compared to risk estimation using traditional method, this method has high preciseness. The characteristic of the Bayesian theory is that, it determines prior distribution of parameters with unknown distribution using the Bootstrap method and taking prior distribution as the starting point [23] and adjusts the problems existing in Bootstrap method using the Bayesian theory in combination with Bootstrap method.

The specific procedures were as follows.

Firstly, suppose the population $X \sim F(x)$ as an unknown distribution and $\theta(F)$ as a parameter to be estimated. Besides, $(X_{(i)} = (x_{i1}, x_{2i}, ..., x_{in})$ was supposed as n original samples from the population $X \sim F(x)$. Let the empirical distribution function of $X_{(i)}$ $(i = 1, 2, ..., m)$ be:

$$F_i(x) \sim \frac{1}{n} \sum_{i=1}^{n} P(X_i < x).$$

Secondly, before the acquisition of $\hat{\theta}(F)$ by evaluating $\theta(F)$, statistic $T_n = \hat{\theta}(F) - \theta(F)$ needed to be created to obtain its statistical characteristics.

Thirdly, $m - 1$ original samples following $U(0, 1)$ were extracted and ranked according to their size, denoted as $u_{(1)}, u_{(2)}, ..., u_{(n-1)}$ $(u_{(0)}, = 0,\ u_{(n)}, = 1)$.

Fourthly, suppose $v_i = u_{(i)} - u_{(i-1)}$ $(i = 1, 2..., n-1)$, let:

$$\hat{\theta}_v = \theta\left(\sum_{i=1}^{n} v_i F_i(x)\right),$$
$$D_n = \hat{\theta}_v - \hat{\theta}(F)$$

$\hat{\theta}_v$ stands for the random weighed statistic. T_n and its statistical characteristics could be obtained through D_n because D_n is the approximate distribution of T_n.

Fifthly, if the number of original samples $X_{(i)} = (x_{i1}, x_{i2}, ..., x_{in})$ was not large enough for the Bayesian estimation, repeated sampling could be adopted to obtain regenerated samples and then $\theta(F)$ was processed by the Bayesian estimation.

If population X followed normal distribution and the number of original samples was small, then we have:

$$X_{(i)} \sim N(\mu, \sigma^2)$$

Suppose $\mu = 0$ and $\theta = \sigma^2$ as parameters to be estimated; moreover, $\theta = \sigma^2 \sim G(\alpha, \beta)$. Then it could be known that $\pi(\theta)\infty\theta^{(\beta-1)}e^{-\frac{\alpha}{\theta}}$ $(\theta, \beta > 0)$.

If the sample extracted from regenerated samples was supposed as $X_{(B)}$ and, moreover, $\hat{\sigma}_B^2 = \hat{\theta}_B$, then the Bayesian estimation result obtained by substituting samples into the above estimation result was as follows.

$$\hat{\theta}_B = E(\theta|x) = \frac{2\alpha + s}{2\beta + n - 2} = \frac{2\beta - 2}{2\beta + n - 2}\ E(\theta) + \frac{n}{2\beta + n - 2}\ \hat{\sigma}_B^2$$

Finally, the Bayesian estimation method was used to perform parameter estimation on characteristic roots of population covariance matrix. The operation was as follows.

If population X followed normal distribution, original sample was $X_{(i)} = (x_{i1}, x_{i2}, ..., x_{in})$, and the volume of samples was small, then $X_{(i)} \sim N(\overline{X}, \hat{\Sigma})$ was thought to follow normal distribution. Then sample $X_{(B)}$ with a volume of B was extracted through regeneration sampling; characteristic roots $\lambda_1^{(b)}, \lambda_2^{(b)}, ..., \lambda_p^{(b)}$ $(b = 1, 2, ..., B)$ of covariance matrix were calculated. The confidence interval of the characteristic root of population covariance matrix, with a confidence level of α, could be obtained according to the upper and lower $\frac{100\alpha}{2}\%$ quantiles of frequency distribution of characteristic roots.

5. Conclusion

In this study, statistical theory was applied to discuss over the statistical inference problem of the population following multivariate normal distribution in principal component analysis under the condition of a large sample size; then non-parametric Bootstrap method was used to explore the statistical inference problem in principal component analysis under the condition of a small sample size; finally, the limitations and bugs of Bootstrap method were corrected using the Bayesian estimation. However, Bootstrap method still has some defects during parameter estimation. Hence, further studies using other methods are needed to improve Bootstrap method.

References

[1] Bro, R., Smilde, A.K., *Analytical Methods*, 6 (9) (2014), 433459.

[2] Ohta, N., *Analytical Chemistry*, 45 (3) (2015), 218-219.

[3] Doukas, H., Papadopoulou, A., Savvakis, N., et al., *Renewable & Sustainable Energy Reviews*, 16 (4) (2012), 1949-1957.

[4] Zheng, X., Levine, D., Shen, J., et al., *Bioinformatics*, 28 (24) (2012), 3326-3328.

[5] Sabokbar, H, F., Roodposhti, M.S., Tazik, E., *Geomorphology*, 226 (2014), 1524.

[6] Moore, B., Moore, B., *Social Science Electronic Publishing*, 61 (100) (2012), 41-64.

[7] Kwak, N., *Cybernetics IEEE Transactions*, 44 (5) (2014), 594-609.

[8] Li, Y., Hsing, T., *Annals of Statistics*, 38 (2010) (2012), 3321-3351.

[9] Garcia-Alvarez, D., Fuente, M.J., Sainz, G.I., *Journal of Process Control*, 22 (3) (2012), 551-563.

[10] Helwig, N.E., Hong, S., Polk, J.D., *Human Movement Science*, 31 (3) (2012), 630-648.

[11] Metsalu, T., Vilo, J., *Nucleic Acids Research*, 43 (1) (2015), 566-70.

[12] Josse, J., Husson, F., *Computational Statistics & Data Analysis*, 56 (6) (2012), 1869-1879.

[13] Abson, D.J., Dougill, A.J., Stringer, L.C., *Applied Geography*, 35 (1-2) (2012), 515-524.

[14] Peng, X., Xin, L., Shi, X., et al., *Food Chemistry*, 151 (20) (2014), 532-8.

[15] Zandi, G., Naysary, B., Sok, Kwan, S., *Journal of Accounting & Management Information Systems*, 12 (3) (2013), 471-488.

[16] Peterson, M.D., Liu, D., Iglayreger, H.B., et al., *Cardiovascular Diabetology*, 11 (1) (2012), 1-10.

[17] Yang, S., Wang, M., Jiao, L., *Information Fusion*, (2012), 13(3):177-184.

[18] Babamoradi, H., Berg, F.V.D., Åsmund, Rinnan, *Chemometrics & Intelligent Laboratory Systems*, 120 (2) (2013), 97-105.

[19] Sharpe, J., Fieller, N., *Journal of Applied Statistics*, (2016), 1-15.

[20] Chen, X., Samson, E., Tocqueville, A., et al., *Journal of Cleaner Production*, 87 (1) (2015), 87-95.

[21] Yaqub, T., Katupitya, J., *Journal of Automation Mobile Robotics & Intelligent Systems*, 2 (6) (2008), 44-53.

[22] She, Y., Li, S., Wu, D., *Journal of the American Statistical Association*, (2014), 41-64.

[23] Jolliffe, I.T., Springer Berlin, 87 (100) (2015), 41-64.

Accepted: 08.10.2016