WEIGHTED APPROACH IN ESTIMATING EVALUATION INDEXES FOR MEASUREMENT SYSTEMS WITH CORRELATED CHARACTERISTICS

Rogério Santana Peruchi [1,2,3,*], Anderson Paulo de Paiva [2], Pedro Paulo Balestrassi [2], Rapinder Sawhney [3], Gagan Rajpal [3], Lavanya Marella [3]

[1] CAPES Foundation, Ministry of Education of Brazil, Brasília, DF 70040-020, Brazil.

[2] Institute of Industrial Engineering and Management, Federal University of Itajubá, Itajubá, MG 37500-903, Brazil

[3] Department of Industrial and Systems Engineering, University of Tennessee, Knoxville, TN 37996, USA

E-mail: rsantana@utk.edu, andersonppaiva@unifei.edu.br, pedro@unifei.edu.br, sawhney@utk.edu, grajpal@utk.edu, mvklavanya@gmail.com

ABSTRACT

This article explores how measurement systems having correlated characteristics are analyzed through studies of gage repeatability and reproducibility (GR&R). The main contribution of this research is the proposal of four new indexes for multivariate analysis of a measurement system. To prove their efficiency, the study generates simulated data with different correlation structures for measurement systems classified as acceptable, marginal, and unacceptable. The proposed indexes are compared with univariate and multivariate indexes in the literature. It was observed that, compared to the other indexes, the most efficient weighted approach in assessing a multivariate measurement system was by the explanation percentage of the eigenvalues extracted from measurement system matrix.

Keywords: *measurement system analysis, repeatability and reproducibility, multivariate analysis of variance, simulation*

1. Introduction

To draw inferences regarding products and process quality, manufacturers use quantitative methods. Such methods (e.g., process capability indexes and control charts) incorporate data into the decision-making process. Prior to obtaining data, a manufacturer should help ensure its validity by evaluating the measuring device (Majeske, 2008). According to Wu et al. (2009), the inevitable variations in process measurements arise from two sources: the manufacturing process and the gage. In manufacturing, a measurement system is not used to produce an exact dimension of a part. Such a system provides measurements that, due to errors (random and systematic), vary from the true value (AIAG, 2010). In any activity involving measurements, whatever observed variability not due to the product/process itself, σ_P^2, is due to measurement error or variability in the measurement system, σ_{SM}^2 (Costa et al., 2005; Li and Al-Refaie, 2008; Senol, 2004; Woodall and Borror, 2008).

Quality improvement projects are often characterized by their objective to reduce variability and achieve zero-defect production. If a product fails to conform to these standards, analysts generally blame the process and then act to improve process capability. In some instances, however, the process capability may be fine. Yet the measurement error, when compared to the variability of the process, remains unacceptable (Al-Refaie and Bata, 2010). Hence, before a team of analysts tries to improve a process, they should investigate both the variability of the measurement process as well as the variability of the manufacturing process. To identify the components of variations in the precision and accuracy assessments of measuring instruments, researchers often rely on measurement system analysis (MSA). The purposes of MSA are to: (1) determine the extent of the observed variability caused by a test instrument, (2) identify the sources of variability in a testing system, and (3) assess the capability of a test instrument (Burdick et al., 2003). According to He et al. (2011), MSA is an important element of Six Sigma as well as of the ISO/TS 16949 standards. Gage Repeatability and Reproducibility (GR&R) is the most common study in MSA to evaluate statistical variations in the measurement process. Repeatability is the variation in measurements obtained by one measuring instrument when used several times by one appraiser while measuring an identical characteristic on the same part. Reproducibility is the variation in the average of measurements made by different appraisers using the same gage when measuring a characteristic on one part (Awad et al., 2009; Burdick et al., 2003; Erdmann et al., 2010; Polini and Turchetta, 2004; Van Den Heuvel and Trip, 2002; Wu et al., 2009). GR&R aims to determine that a measurement system's variability is less than that of the monitoring process (Al-Refaie and Bata, 2010; Wang and Chien, 2010).

As emphasized above, a team of analysts, before analyzing the process capability of a quality improvement project, should evaluate the capability of the measurement system. Two methods commonly used in the analysis of a GR&R study are: (1) an analysis of variance (ANOVA) approach followed by estimation of the appropriate variance components; and (2) an X-bar and Range chart that estimates the standard deviations of the components of gage variability (Wang and Chien, 2010). Analysts prefer the ANOVA method because it measures the operator-to-part interaction gage error; this variation is not included in the X-bar and Range method (AIAG, 2010). Burdick et al. (2003) provided a good review of methods for conducting and analyzing measurement system capability studies, which are based on the analysis of variance approach. Dejaegher et al. (2006) used Six Sigma to measure, analyze, and improve the capability of a procedure required in the testing of the quality of an active pharmaceutical ingredient. This was done using multiple GR&R studies to analyze the capability of the measurement procedure. A design of experiments was next designed to improve this procedure. Kaija et al. (2010) used the Six Sigma DMAIC (define, measure, analyze, improve, control) approach to evaluate a process of printing a dielectric layer with an inkjet printer. Initially, a GR&R study was conducted to evaluate the proportion of variation caused by the measurement system and process variation. Experiments were then planned and analyzed to identify the parameters having the most significant effects on the output variables of the dielectric layer's insulating layer and surface roughness. Li and Al-Refaie (2008) used the Six Sigma DMAIC procedure to improve quality through enhancing the measuring system capability of the wood industry. The measurement system assessed through GR&R had been considered unacceptable. To improve it, analysts implemented corrective actions, including operator training, proper selection of measuring instruments, and improved measuring procedures. In a second GR&R study, the authors concluded that the corrective actions had reduced the *%R&R* index (percentage of repeatability and reproducibility) by 39.38% and had improved the *ndc* index (number of distinct categories) by 168.84%.

Other studies have employed GR&R to evaluate measurement systems. In proposing a procedure to evaluate measurement systems and process capabilities, Al-Refaie and Bata (2010) used GR&R along with four quality measures. The quality measures were: precision-to-tolerance ratio (*P/T*), signal-to-noise ratio (*SNR*), discrimination ratio (*DR*), and process capability index (*Cp* or *Cpk*). Costa et al. (2005) addressed the design and implementation of a measurement system that permitted the evaluation—quantitatively, objectively, and systematically—of the superficial paper waviness in industrial practice. The process of designing the measurement system was presented considering all its stages, from selection and evaluation of the measuring device (using GR&R) to the generation and validation of the statistical model of measurement. Senol (2004) used an experimental design including laboratory factors as a measurement variability factor in MSA studies. This study concluded that environmental and atmospheric conditions, often overlooked in GR&R studies, might represent a significant contribution to the variability in measurements. Majeske (2012a) presented a methodology for MSA under two conditions: the measuring device is robust to minor differences in how a part is oriented in the measuring fixture and the quality characteristic has within-part variability. Using a standard MSA approach, the gauge did not satisfy the approval criteria, suggesting that the gauge was not capable of precisely measuring the component. As a result, the author concluded that a gauge that has very good precision can fail the standard techniques if it does not satisfy their assumptions. Weaver et al. (2012) used a Bayesian approach to estimate variance components in GR&R studies. In their article, worked examples of gauge R&R data analysis for types of studies common in industrial applications were provided. The results of this study indicate that a Bayesian approach to analyzing these data is much simpler and requires very few changes to an estimation procedure when adapted to a new situation. Majeske (2012b) developed two-sample hypothesis tests for five different MSA criteria to compare the ability of the two systems to provide precise measurements. The techniques were demonstrated using data from an automotive body manufacturing facility that compares a coordinate measuring machine to a noncontact vision-based measurement system.

The bulk of the studies associated with analyzing the quality and efficiency of measurement systems are so far limited to a discussion of one single critical-to-quality characteristic (CTQ). Currently, the ANOVA method for GR&R studies can be applied only to univariate data (Wang and Yang, 2007). In assessing measurement systems that measure multiple characteristics, the analyst must consider the correlation structure of the CTQs, a task more suited to multivariate methods. Flynn et al. (2009) used regression analysis to analyze the comparative performance capability between two functionally equivalent but technologically different automatic measurement systems. The systems were used for acceptance testing of a unit under test. For such accurate measurements as repeatability and reproducibility, the "pass/fail" criteria for a test unit were inappropriate. Hence, the authors proposed a methodology that used principal component analysis (PCA) and multivariate analysis of variance (MANOVA) to examine whether there was a statistically significant difference between the system's measurements. He et al. (2011) proposed an online multivariate MSA approach to detecting faulty test instruments in a multisite testing system. The multivariate data were transformed using PCA. The values of the principal components of each test instrument were then compared with the control limits obtained by analyzing the principal components of all test instruments. Majeske (2008) used the MANOVA method to estimate the variance-covariance matrix for GR&R studies with one, two, and three significant factors. This work evaluated a measurement system using data from a GR&R study of a sheet-metal body panel. In so doing it demonstrated how to adjust a MANOVA model and estimate multivariate criteria (P/T_m, $\%R\&R_m$ and SNR_m). Wang and Chien (2010) used the process-oriented basis representation method (POBREP) to evaluate a measurement process with multivariate data. The results showed that POBREP outperformed other methods such as PCA and ANOVA. The POBREP was able to identify specific causes of production problems and map those into a basis matrix. Wang and Yang (2007) presented a GR&R study with multiple characteristics using the PCA method. To assess the adequacy of the measurement system, the study employed two composite indexes: precision-to-tolerance ratio and measurement-variation-to-total-variation-of-measurement-system ratio. The case study showed that, for estimating the indexes, PCA outperformed the ANOVA method.

This article deals with a multivariate analysis of a measurement system through studies of repeatability and reproducibility of the measurement process. Its main objective is two-fold: to propose new indexes for multivariate analysis of a measurement system and to assess the performance between the proposed indexes and those found in the literature. The four new indexes are calculated using a weighted approach for multivariate analysis of variance. To prove their efficiency, simulated data are generated with different correlation structures and measurement systems that are unacceptable, marginal (may be acceptable depending on application), and acceptable. The results obtained by the proposed indexes are then compared to those obtained through indexes found in the literature. The simulation study concludes that the most efficient weighted approach in estimating multivariate indexes was by the explanation percentage of the eigenvalues extracted from measurement system matrix.

The remainder of this paper is structured as follows. Section 2 shows how to evaluate a measurement system using univariate method by ANOVA. Section 3 details the MANOVA method and the four multivariate indexes proposed in this article. In Section 4, a simulation study is conducted to evaluate the indexes' performance, especially the multivariate, for different correlation structures as well as for measurement systems that are unacceptable, marginal, and acceptable. Finally, Section 5 presents the main findings involving the analysis using univariate and multivariate methods for assessing multivariate measurement systems.

2. MSA by univariate GR&R study

In many processes involving measurements of manufactured products for a single CTQ, the variability may be due to a measurement error, to variability in the measuring device, or to variability in the product/process itself. A complete model for a GR&R study with *p* parts, *o* operators, and *r* replicates is made up of a two-factor crossed design with interaction as such (Al-Refaie and Bata, 2010; Burdick et al., 2003; Deldossi and Zappa, 2011; Erdmann et al., 2010):

$$\mathbf{ctq} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,o \\ k = 1,2,\ldots,r \end{cases} \quad (1)$$

where $\mathbf{ctq}$ is the response variable measured; μ is the mean of the measured values; $\alpha_i \sim N(0,\sigma_\alpha)$, $\beta_j \sim N(0,\sigma_\beta)$, $\alpha\beta_{ij} \sim N(0,\sigma_{\alpha\beta})$ and $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon)$ are random variables statistically independent part, operator, interaction and the error term, respectively. The above components of variance can be translated into notation GR&R to (Kaija et al., 2010; Li and Al-Refaie, 2008; Senol, 2004; White and Borror, 2011):

$$\begin{aligned} &\sigma_P^2 = \sigma_\alpha^2, \\ &\sigma_{repeatability}^2 = \sigma_\varepsilon^2, \sigma_{reproducibility}^2 = \sigma_\beta^2 + \sigma_{\alpha\beta}^2 \\ &\sigma_{MS}^2 = \sigma_{repeatability}^2 + \sigma_{reproducibility}^2, \\ &\sigma_T^2 = \sigma_P^2 + \sigma_{MS}^2 \end{aligned} \quad (2)$$

The variance components of Model (1) in Eq (2) can be estimated using the method Analysis of Variance (ANOVA). More details on how to calculate the components of variation using ANOVA can be found in Majeske (2008) and Wang and Chien (2010).

To determine the acceptability of a measurement system, the AIAG (2010) recommended evaluating a measurement system by scaling the standard deviation of measurement error to the total standard deviation of the observed process. This statistic, called the percentage of R&R, is defined as:

$$\%R\,\&\,R = \left(\frac{\sigma_{MS}}{\sigma_T}\right)100\% \quad (3)$$

If the measurement system is, according to the index, less than 10%, it is considered acceptable. If between 10% and 30%, it is considered marginal—acceptable depending on the application, the cost of the measurement device, the cost of repair and other factors. If it exceeds 30%, it is considered unacceptable and should be improved (AIAG, 2010; Al-Refaie and Bata, 2010; Woodall and Borror, 2008).

3. MSA by multivariate GR&R study

When a GR&R study considers a two-factors cross design with interaction for multiple CTQs (*q* characteristics), the model is given by (He et al., 2011; Majeske, 2008; Wang and Chien, 2010):

$$\begin{aligned} \mathbf{CTQ} &= \begin{bmatrix} CTQ_{11} & CTQ_{12} & \cdots & CTQ_{1q} \\ CTQ_{21} & CTQ_{22} & \cdots & CTQ_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ CTQ_{n1} & CTQ_{n2} & \cdots & CTQ_{nq} \end{bmatrix} \\ &= \boldsymbol{\mu} + \boldsymbol{\alpha}_\mathbf{i} + \boldsymbol{\beta}_\mathbf{j} + (\boldsymbol{\alpha\beta})_\mathbf{ij} + \boldsymbol{\varepsilon}_\mathbf{ijk} \end{aligned} \quad (4)$$

where $\mathbf{CTQ} = (\mathbf{CTQ_1}, \mathbf{CTQ_2}, \ldots, \mathbf{CTQ_q},)$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_q)$ are constant vectors; $\boldsymbol{\alpha}_\mathbf{i} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$, $\boldsymbol{\beta}_\mathbf{j} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$, $\boldsymbol{\alpha\beta}_\mathbf{ij} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\alpha\beta})$, and $\boldsymbol{\varepsilon}_\mathbf{ijk} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ are random vectors statistically independent of each other. It is possible to translate the above components of variance into notation GR&R to:

$$\begin{aligned} &\hat{\boldsymbol{\Sigma}}_\mathbf{P} = \hat{\boldsymbol{\Sigma}}_\boldsymbol{\alpha}, \\ &\hat{\boldsymbol{\Sigma}}_\mathbf{repeatability} = \hat{\boldsymbol{\Sigma}}_\boldsymbol{\varepsilon}, \hat{\boldsymbol{\Sigma}}_\mathbf{reproducibility} = \hat{\boldsymbol{\Sigma}}_\boldsymbol{\beta} + \hat{\boldsymbol{\Sigma}}_\boldsymbol{\alpha\beta} \\ &\hat{\boldsymbol{\Sigma}}_\mathbf{MS} = \hat{\boldsymbol{\Sigma}}_\mathbf{repeatability} + \hat{\boldsymbol{\Sigma}}_\mathbf{reproducibility}, \\ &\hat{\boldsymbol{\Sigma}}_\mathbf{T} = \hat{\boldsymbol{\Sigma}}_\mathbf{P} + \hat{\boldsymbol{\Sigma}}_\mathbf{MS} \end{aligned} \quad (5)$$

The variance components of the model (4) in Eq. (5) can be estimated using the method of Multivariate Analysis of Variance (MANOVA). Before estimating the variance-covariance matrices, $\boldsymbol{\Sigma}_\mathbf{P}$, $\boldsymbol{\Sigma}_\mathbf{MS}$ and $\boldsymbol{\Sigma}_\mathbf{T}$ are calculated mean squares matrices for part, operator, part*operator interaction and the error term. More details on how to calculate these components of variation using MANOVA for multivariate GR&R studies can be found in Majeske (2008).

The multivariate version of the *%R&R* index proposed by Majeske (2008) is called here *G* index and is calculated by Eq. (6). λ_{MS_i} and λ_{T_i} $\forall i = 1,2,\ldots,q$ are eigenvalues extracted from variance-covariance matrices, $\boldsymbol{\Sigma}_\mathbf{MS}$ and $\boldsymbol{\Sigma}_\mathbf{T}$.

$$G = \left(\prod_{i=1}^{q} \sqrt{\frac{\boldsymbol{\lambda}_{\mathbf{MS_i}}}{\boldsymbol{\lambda}_{\mathbf{T_i}}}}\right)^{1/q} 100\% \quad (6)$$

To obtain the evaluation index to the measurement system, Majeske (2008) applied geometric mean on the ratio $\sqrt{\lambda_{MS}/\lambda_T}$. This strategy does not determine greater importance to the most significant pair of eigenvalues, extracted from variance-covariance matrices. As a result, this article adopts a weighted approach upon $\sqrt{\lambda_{MS}/\lambda_T}$ ratio to propose four new evaluation indexes for multivariate measurement systems. These new indexes, WA_T, WA_{MS}, WG_T and WG_{MS}, can be obtained based on Eqs. (7) and (8).

$$WA = \sum_{i=1}^{q}\left(W_i\sqrt{\frac{\lambda_{SM_i}}{\lambda_{T_i}}}\right)100\% \tag{7}$$

$$WG = \prod_{i=1}^{q}\left(\sqrt{\frac{\lambda_{SM_i}}{\lambda_{T_i}}}\right)^{W_i}100\% \tag{8}$$

where λ_{SM} and λ_T are eigenvalues extracted from variance-covariance matrices $\hat{\mathbf{\Sigma}}_{\mathbf{MS}}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{T}}$, respectively; $W_i\ \forall i = 1,...,q$ are the explanation percentage of the eigenvalues extracted from either $\hat{\mathbf{\Sigma}}_{\mathbf{T}} : W_i = \left(\lambda_{T_i}\Big/\sum_{j=1}^{q}\lambda_{T_j}\right)$ or $\hat{\mathbf{\Sigma}}_{\mathbf{MS}} : W_i = \left(\lambda_{SM_i}\Big/\sum_{j=1}^{q}\lambda_{SM_j}\right)$. The WA_T and WA_{MS} indexes are obtained by calculating the weighted arithmetic mean according to Eq. (7). The first index, WA_T, weights the $\sqrt{\lambda_{MS}/\lambda_T}$ ratio using the explanation percentage of the eigenvalues extracted from total variation matrix. The second index, WA_{MS}, weights the $\sqrt{\lambda_{MS}/\lambda_T}$ ratio through the explanation percentage of the eigenvalues extracted from measurement system matrix. On the other hand, the WG_T and WG_{MS} indexes are calculated using weighted geometric mean in the Eq. (8). The first index, WG_T, weights the $\sqrt{\lambda_{MS}/\lambda_T}$ ratio using the explanation percentage of the eigenvalues extracted from total variation matrix. The second index, WG_{MS}, weights the $\sqrt{\lambda_{MS}/\lambda_T}$ ratio through the explanation percentage of the eigenvalues extracted from measurement system matrix. The acceptance criterion of the measurement system is the same as described in section 2 (Majeske, 2008).

4. Simulation

5.1 Detailing the simulation study

The purpose of this simulation is to evaluate several possible situations in multivariate analysis of a measurement system and to compare the results achieved, primarily, through multivariate indexes. Simulated data will be generated for measurement systems that are unacceptable ($\%R\&R_m>30\%$), marginal ($10\%<\%R\&R_m<30\%$) and acceptable ($\%R\&R_m<10\%$), as well as correlations that are very low ($W_1\leq 65\%$), low ($65\%<W_1\leq 75\%$), medium ($75\%<W_1\leq 85\%$), high ($85\%<W_1\leq 95\%$), and very high ($W_1>95\%$), a total of 15 scenarios and 1800 simulated measurements. W_1 is the result obtained from $\lambda_{T_1}\Big/\sum_{j=1}^{q}\lambda_{T_j}$. Simulated data were generated from the information in Table 1, according to the same amount of CTQs, parts, operators and replicates in Majeske (2008), $q=4$, $p=5$, $o=2$, and $r=3$. The data for the 15 simulated scenarios can be found in http://www.pedro.unifei.edu.br/download/tables.rar.

Table 1
Mean vectors and variance-covariance matrices used to generate simulated data with different correlations and measurement systems.

Scenarios	Mean vectors										Variance-covariance matrix
	P_1O_1	P_2O_1	P_3O_1	P_4O_1	P_5O_1	P_1O_2	P_2O_2	P_3O_2	P_4O_2	P_5O_2	
1 Very Low corr. Unacceptable MS	4.00	8.00	6.00	10.00	5.00	4.10	8.10	5.90	9.90	4.90	1.10 1.27 1.39 1.50
	8.00	6.00	13.00	9.00	11.00	7.90	6.10	12.90	9.10	10.90	1.27 1.50 1.63 1.76
	9.00	10.00	13.00	16.00	7.00	9.10	10.10	12.90	15.90	7.10	1.39 1.63 1.80 1.92
	7.00	11.00	5.00	10.00	15.00	7.10	10.90	5.10	10.10	15.10	1.50 1.76 1.92 2.10
2 Low corr. Unacceptable MS	4.00	8.00	6.00	10.00	5.00	4.10	8.10	5.90	9.90	4.90	1.10 1.27 1.39 1.50
	8.00	7.00	9.00	12.00	11.00	7.90	6.90	9.10	12.10	10.90	1.27 1.50 1.63 1.76
	9.00	10.00	7.00	13.00	15.00	9.10	10.10	6.90	13.10	14.90	1.39 1.63 1.80 1.92
	7.00	13.00	11.00	14.00	17.00	7.10	13.10	11.10	13.90	16.90	1.50 1.76 1.92 2.10
3 Medium corr. Unacceptable MS	9.00	7.00	5.00	12.00	10.00	9.01	6.99	5.01	12.01	9.99	1.10 1.27 1.39 1.50
	8.00	7.00	9.00	12.00	11.00	7.99	6.99	9.01	12.01	10.99	1.27 1.50 1.63 1.76
	9.00	10.00	7.00	13.00	15.00	9.01	10.01	6.99	13.01	14.99	1.39 1.63 1.80 1.92
	7.00	13.00	9.00	17.00	14.00	7.01	13.01	8.99	16.99	14.01	1.50 1.76 1.92 2.10
4 High corr. Unaccpetable MS	6,00	4,00	8,00	10,00	12,00	6,01	4,01	7,99	9,99	12,01	1.50 1.58 1.63 1.67
	3,00	6,00	9,00	11,00	15,00	3,01	6,01	9,01	10,99	14,99	1.58 1.70 1.73 1.78
	6,00	8,00	11,00	15,00	13,00	6,01	8,01	11,10	15,10	13,10	1.63 1.73 1.80 1.83
	8,00	10,00	12,00	16,00	14,00	7,99	10,01	12,01	16,01	14,01	1.67 1.78 1.83 1.90
5 Very high corr. Unacceptable MS	4.00	6.00	8.00	10.00	12.00	4.01	6.01	7.99	9.99	12.01	1.10 1.27 1.39 1.50
	5.00	7.00	9.00	11.00	13.00	5.01	7.01	9.01	10.99	12.99	1.27 1.50 1.63 1.76
	6.00	8.00	10.00	12.00	14.00	6.01	8.01	9.99	11.99	13.99	1.39 1.63 1.80 1.92
	8.00	10.00	12.00	14.00	16.00	7.99	10.01	12.01	14.01	15.99	1.50 1.76 1.92 2.10
6 Very Low corr. Marginal MS	4.00	8.00	6.00	10.00	5.00	4.10	8.10	5.90	9.90	4.90	0.22 0.25 0.28 0.30
	8.00	6.00	13.00	9.00	11.00	7.90	6.10	12.90	9.10	10.90	0.25 0.30 0.33 0.35
	5.00	8.00	9.00	14.00	12.00	5.10	8.10	8.90	13.90	12.10	0.28 0.33 0.36 0.38
	7.00	13.00	5.00	10.00	17.00	7.10	13.10	5.10	10.10	16.90	0.30 0.35 0.38 0.42
7 Low corr. Marginal MS	6.00	8.00	4.00	11.00	10.00	6.10	8.10	3.90	10.90	9.90	0.22 0.25 0.28 0.30
	8.00	7.00	9.00	12.00	11.00	7.90	6.90	9.10	12.10	10.90	0.25 0.30 0.33 0.35
	7.00	13.00	10.00	11.00	15.00	7.10	13.10	9.90	11.10	14.90	0.28 0.33 0.36 0.38
	9.00	11.00	14.00	13.00	17.00	9.10	10.90	14.10	12.90	16.90	0.30 0.35 0.38 0.42

8 Medium corr. Marginal MS	9.00	7.00	5.00	12.00	10.00	9.01	6.99	5.01	12.01	9.99	0.22	0.25	0.28	0.30
	8.00	7.00	9.00	12.00	11.00	7.99	6.99	9.01	12.01	10.99	0.25	0.30	0.33	0.35
	9.00	10.00	7.00	13.00	15.00	9.01	10.01	6.99	13.01	14.99	0.28	0.33	0.36	0.38
	7.00	13.00	9.00	17.00	14.00	7.01	13.01	8.99	16.99	14.01	0.30	0.35	0.38	0.42
9 High corr. Marginal MS	6,00	4,00	8,00	10,00	12,00	6,01	4,01	7,99	9,99	12,01	0.22	0.25	0.28	0.30
	3,00	6,00	9,00	11,00	15,00	3,01	6,01	9,01	10,99	14,99	0.25	0.30	0.33	0.35
	6,00	8,00	11,00	15,00	13,00	6,01	8,01	11,10	15,10	13,10	0.28	0.33	0.36	0.38
	8,00	10,00	12,00	16,00	14,00	7,99	10,01	12,01	16,01	14,01	0.30	0.35	0.38	0.42
10 Very high corr. Marginal MS	4.00	6.00	8.00	10.00	12.00	4.01	6.01	7.99	9.99	12.01	0.22	0.25	0.28	0.30
	5.00	7.00	9.00	11.00	13.00	5.01	7.01	9.01	10.99	12.99	0.25	0.30	0.33	0.35
	6.00	8.00	10.00	12.00	14.00	6.01	8.01	9.99	11.99	13.99	0.28	0.33	0.36	0.38
	8.00	10.00	12.00	14.00	16.00	7.99	10.01	12.01	14.01	15.99	0.30	0.35	0.38	0.42
11 Very Low corr. Acceptable MS	4.00	8.00	6.00	10.00	5.00	4.10	8.10	5.90	9.90	4.90	0.04	0.04	0.05	0.05
	8.00	6.00	13.00	9.00	11.00	7.90	6.10	12.90	9.10	10.90	0.04	0.05	0.05	0.06
	5.00	8.00	9.00	14.00	12.00	5.10	8.10	8.90	13.90	12.10	0.05	0.06	0.06	0.06
	7.00	13.00	5.00	10.00	17.00	7.10	13.10	5.10	10.10	16.90	0.05	0.06	0.06	0.07
12 Low corr. Acceptable MS	6.00	8.00	4.00	11.00	10.00	6.01	8.01	3.99	10.99	9.99	0.04	0.04	0.05	0.05
	7.00	5.00	9.00	13.00	11.00	6.99	4.99	9.01	13.01	10.99	0.04	0.05	0.05	0.06
	7.00	13.00	10.00	11.00	15.00	7.01	13.01	9.99	11.01	14.99	0.05	0.06	0.06	0.06
	6.00	10.00	14.00	12.00	17.00	6.01	9.99	14.01	12.01	16.99	0.05	0.06	0.06	0.07
13 Medium corr. Acceptable MS	9.00	7.00	5.00	12.00	10.00	9.01	6.99	5.01	12.01	9.99	0.04	0.04	0.05	0.05
	8.00	7.00	9.00	12.00	11.00	7.99	6.99	9.01	12.01	10.99	0.04	0.05	0.05	0.06
	9.00	10.00	7.00	13.00	15.00	9.01	10.01	6.99	13.01	14.99	0.05	0.06	0.06	0.06
	7.00	13.00	9.00	17.00	14.00	7.01	13.01	8.99	16.99	14.01	0.05	0.06	0.06	0.07
14 High corr. Acceptable MS	6,00	4,00	8,00	10,00	12,00	6,01	4,01	7,99	9,99	12,01	0.04	0.04	0.05	0.05
	3,00	6,00	9,00	11,00	15,00	3,01	6,01	9,01	10,99	14,99	0.04	0.05	0.05	0.06
	6,00	8,00	11,00	15,00	13,00	6,01	8,01	11,10	15,10	13,10	0.05	0.06	0.06	0.06
	8,00	10,00	12,00	16,00	14,00	7,99	10,01	12,01	16,01	14,01	0.05	0.06	0.06	0.07
15 Very high corr. Acceptable MS	4.00	6.00	8.00	10.00	12.00	4.01	6.01	7.99	9.99	12.01	0.04	0.04	0.05	0.05
	5.00	7.00	9.00	11.00	13.00	5.01	7.01	9.01	10.99	12.99	0.04	0.05	0.05	0.06
	6.00	8.00	10.00	12.00	14.00	6.01	8.01	9.99	11.99	13.99	0.05	0.06	0.06	0.06
	8.00	10.00	12.00	14.00	16.00	7.99	10.01	12.01	14.01	15.99	0.05	0.06	0.06	0.07

5.2 Criterion of methods' assessment

This simulation study will focus only on the comparison of multivariate index obtained by MANOVA method. For each scenario, it was tried to obtain close *%R&R* index values for **CTQ$_1$**, **CTQ$_2$**, **CTQ$_3$**, and **CTQ$_4$**. Thus, multivariate indexes are expected to be estimated close to those obtained by ANOVA method. The criterion used in this work to determine if the estimated multivariate index is correct is based on confidence intervals for mean calculated from results obtained by ANOVA method. The lower (LCL) and upper (UCL) limits of the confidence intervals are calculated using

$$CI = \overline{CTQ} \pm t_{N-1,\,\alpha/2}\,\frac{s}{\sqrt{N}} \qquad (9)$$

where $\overline{CTQ}$ is the mean of *%R&R* between **CTQ$_1$**, **CTQ$_2$**, **CTQ$_3$** and **CTQ$_4$**; s is the standard deviation; N is the sample size and $t_{N-1,\alpha}$ is the $(1-\alpha)100^{th}$ percentile of a t distribution with $(N-1)$ degrees of freedom. Note that it would not make sense to evaluate situations in which CTQs determine distinct classifications to the measurement system. For instance , CTQ$_1$ and CTQ$_2$ classify the measurement system as unacceptable and, on the other hand, CTQ$_3$ and CTQ$_4$ classify the measurement system as acceptable. In such situations, the confidence interval would be wider, thereby, *%R&R$_m$* indexes would be easily estimated inside the limits.

5.3 Result Analysis

Table 2 presents the results of calculations of the *%R&R* index, the mean value and the 95% confidence interval, as well as the multivariate evaluation indexes. Moreover, Fig. 1 graphically presents how accurate the multivariate indexes were estimated compared to the 95% confidence interval. The analysis and comparison will be performed in two ways: intra- and inter-indexes. The intra-index analysis will provide an overview of the indexes' performance and, on the other hand, the inter-index analysis will seek to justify the indexes' deviations from the confidence intervals.

In the intra-index analysis was verified that WA_{MS} and WG_{MS} indexes were more robust than G, WA_T and WG_T. G index was estimated within the confidence interval only in scenarios S9, S11 and S14. WA_T and WG_T indexes failed in one and six scenarios, respectively. As seen in Table 2, WA_{MS} and WG_{MS} indexes were estimated within the confidence interval for all 15 scenarios evaluated.

For the inter-index analysis, Table 3 presents how the G index was estimated for the 15 simulated scenarios. In addition, through Fig. 1 is verified that this G index was calculated within of the confidence interval only in S9, S11 and S14. This index was obtained using geometric mean of the $\sqrt{\lambda_{MS}/\lambda_T}$ ratio according to the amount of quality characteristics. This simulation study dealt with four characteristics. Thus, four eigenvalues of the $\hat{\mathbf{\Sigma}}_{\mathbf{MS}}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{T}}$ matrices were extracted. If the individual ratio $\sqrt{\lambda_{MS}/\lambda_T}$ for each pair of eigenvalues, in $\hat{\mathbf{\Sigma}}_{\mathbf{MS}}$ and $\hat{\mathbf{\Sigma}}_{\mathbf{T}}$, provide different interpretations, G index may not represent well the performance of the measurement system. Indeed, geometric mean provides the same degree of importance in the analysis of each pair of eigenvalues. Nevertheless, it is known that the first eigenvalues have a greater percentage of explaining the measured phenomenon greater than the last eigenvalues. Therefore, the need is confirmed that some form of weighting for the calculation of this index should be used. In the inter-index analysis by WA_T and WG_T, the weighting approach through the explanation percentage of the eigenvalues extracted from $\hat{\mathbf{\Sigma}}_{\mathbf{T}} : W_i = \left(\lambda_{T_i} \big/ \sum_{j=1}^{q} \lambda_{T_j}\right)$ was not satisfactory. These indexes (mainly WG_T) failed in scenarios with correlations deemed lower due to higher weights assigned to less significant $\sqrt{\lambda_{MS}/\lambda_T}$ ratios.

In the inter-index analysis by WA_{MS} and WG_{MS}, the explanation percentage of the eigenvalues extracted from $\hat{\mathbf{\Sigma}}_{\mathbf{MS}} : W_i = \left(\lambda_{SM_i} \big/ \sum_{j=1}^{q} \lambda_{SM_j}\right)$ showed to be the most efficient weighting approach for assessing a multivariate measurement system. In the most scenarios, the first pair of eigenvalues to calculate $\sqrt{\lambda_{MS}/\lambda_T}$ ratio receives greater degree of importance, including scenarios with lower structure correlations. Conceptually, the weighted approach using the explanation percentages of the eigenvalues extracted from measurement system matrix makes more sense than from total variation matrix, in estimating evaluation index for GR&R studies.

Table 2
Results for calculations of the *%R&R* index, mean and 95% confidence interval, and *%R&R$_m$* index.

SCENARIO			UNIVARIATE (*%R&R*)				MEAN CI			MULTIVARIATE				
S	MS	Corr.	CTQ_1	CTQ_2	CTQ_3	CTQ_4	$\overline{CTQ}$	LCL	UCL	G	WA_T	WG_T	WA_{MS}	WG_{MS}
S1	Unaccept.	VL	49.9	39.3	38.3	34.1	40.42	29.69	51.14	10.78	**31.62**	18.28	**48.84**	**48.51**
S2		L	42.2	55.5	44.3	39.8	45.44	34.42	56.47	13.30	**36.50**	30.28	**45.42**	**44.94**
S3		M	40.8	52.4	42.6	36.9	43.18	32.63	53.72	11.32	**38.27**	31.45	**45.76**	**45.51**
S4		H	45.3	33.2	41.2	47.8	41.86	31.70	52.03	28.15	**42.95**	**42.20**	**44.33**	**44.14**
S5		VH	31.1	34.9	37.8	41.1	36.21	29.45	42.97	64.09	**35.81**	**35.79**	**36.05**	**35.94**
S6	Marginal	VL	15.8	14.1	13.7	10.2	13.48	9.75	17.21	4.97	9.62	6.48	**15.67**	**15.38**
S7		L	18.6	27.2	21.3	24.1	22.82	16.95	28.69	10.04	**19.86**	14.68	**27.15**	**26.87**
S8		M	15.5	23.7	17.0	14.6	17.69	11.16	24.21	5.40	**15.38**	**13.24**	**18.02**	**17.90**
S9		H	13.2	10.3	13.6	16.9	13.50	9.19	17.80	**14.31**	**14.37**	**14.35**	**14.44**	**14.43**
S10		VH	15.2	19.0	19.7	20.9	18.70	14.80	22.59	47.23	**16.95**	**16.94**	**17.33**	**17.12**
S11	Accept.	VL	8.4	6.3	4.9	5.3	6.22	3.67	8.77	**4.08**	**5.00**	**4.41**	**6.75**	**6.49**
S12		L	5.6	4.6	6.7	5.4	5.54	4.15	6.92	2.01	**4.70**	3.53	**6.18**	**6.13**
S13		M	6.2	9.6	6.6	5.9	7.07	4.37	9.76	2.28	**6.07**	**5.32**	**7.05**	**7.00**
S14		H	5.7	4.5	5.9	7.3	5.84	4.00	7.69	**7.22**	**6.58**	**6.56**	**6.65**	**6.63**
S15		VH	6.5	7.6	8.6	9.2	7.95	6.07	9.83	39.35	**7.78**	**7.78**	**8.18**	**7.89**

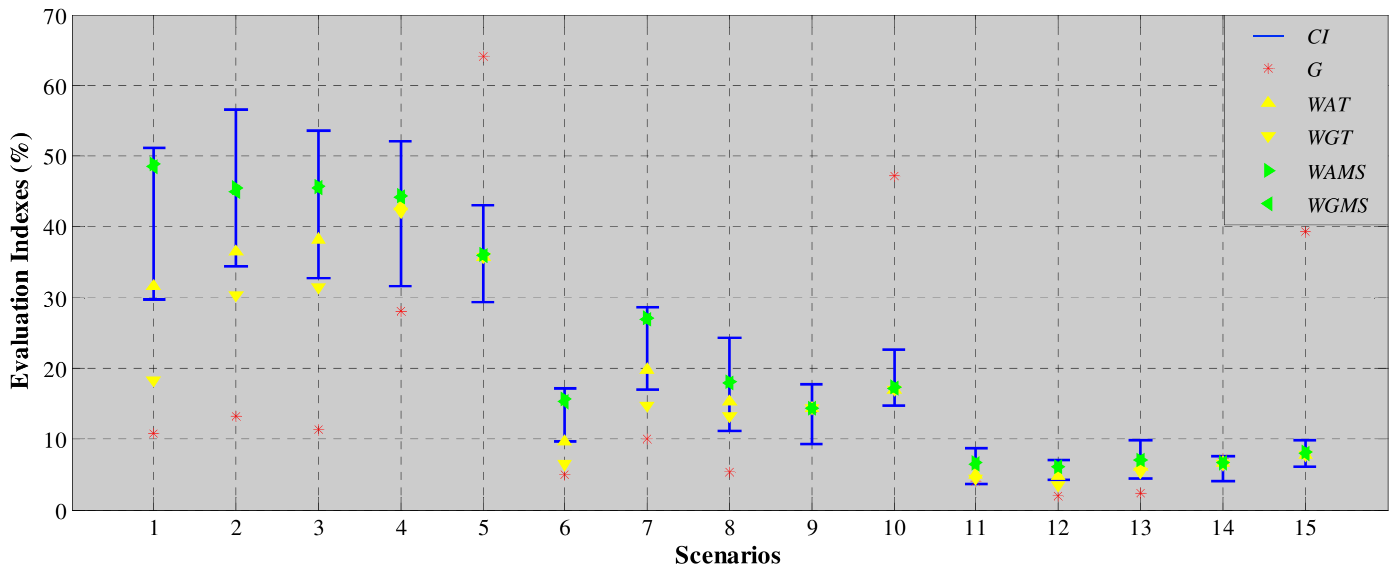

Fig. 1. Multivariate evaluation indexes estimated for 15 distinct scenarios

Table 3
G index for the inter-method analysis.

S	$\overline{\text{CTQ}}$	LCL	UCL	$\sqrt{\lambda_{MS_1}/\lambda_{T_1}}$	$\sqrt{\lambda_{MS_2}/\lambda_{T_2}}$	$\sqrt{\lambda_{MS_3}/\lambda_{T_3}}$	$\sqrt{\lambda_{MS_4}/\lambda_{T_4}}$	G
S1	40.42	29.69	51.14	**49.07(61.2[a])**	3.29 (27.4)	5.13 (10.5)	16.34 (0.9)	10.78
S2	45.44	34.42	56.47	**45.90 (74.6)**	9.35 (13.9)	8.32 (7.7)	8.76 (3.9)	13.30
S3	43.18	32.63	53.72	**45.97 (80.4)**	5.81 (10.8)	7.80 (5.2)	7.87 (3.6)	11.32
S4	41.86	31.70	52.03	**43.30 (87.3)**	52.70 (8.6)	14.66 (3.1)	18.78 (1.1)	28.15
S5	36.21	29.45	42.97	**37.75 (99.8)**	51.77 (0.2)	98.97 (0.0)	92.13 (0.0)	64.09
S6	13.48	9.75	17.21	**15.94 (53.4)**	2.14 (29.0)	2.40 (16.3)	7.43 (1.3)	4.97
S7	22.82	16.95	28.69	**27.39 (67.7)**	4.44 (18.3)	3.37 (13.9)	**24.73 (0.2)**	10.04
S8	17.69	11.16	24.21	**18.15 (81.2)**	3.00 (11.0)	4.45 (4.6)	3.51 (3.2)	5.40
S9	13.50	9.19	17.80	**14.50 (90.4)**	**14.24 (6.9)**	**10.03 (2.6)**	20.27 (0.1)	**14.31**
S10	18.70	14.80	22.59	**16.92 (99.9)**	56.75 (0.0)	70.15 (0.0)	73.86 (0.0)	47.23
S11	6.22	3.67	8.77	**7.27 (50.4)**	2.80 (34.5)	2.21 (34.5)	**6.14 (14.2)**	**4.08**
S12	5.54	4.15	6.92	**6.23 (71.5)**	0.95 (15.5)	0.73 (12.7)	3.80 (0.3)	2.01
S13	7.07	4.37	9.76	**7.11 (81.5)**	1.34 (10.6)	1.69 (4.6)	1.68 (3.3)	2.28
S14	5.84	4.00	7.69	**6.59 (90.8)**	**7.39 (6.4)**	**4.10 (2.8)**	13.63 (0.1)	**7.22**
S15	7.95	6.07	9.83	**7.78 (100.0)**	38.19 (0.0)	98.44 (0.0)	82.04 (0.0)	39.35

[a] $\lambda_{T_i} / \sum_{j=1}^{q} \lambda_{T_j} \quad i = 1,2,\ldots,q$

5. Conclusions

This article addressed the multivariate analysis of measurement systems through studies of repeatability and reproducibility of the measurement process. The main contribution of this paper is its proposal for new indexes for multivariate analysis of the measurement system by multivariate analysis of variance. To prove the efficiency of the indexes, simulated data were generated with different correlation structures for measurement systems considered acceptable, marginal, and unacceptable. The results obtained by the proposed indexes were compared to those obtained by the multivariate index in the literature. Statistical analysis provided the following conclusions:

1. *G* index uses geometric mean for evaluating the measurement system. This approach may be incorrect when the ratio $\sqrt{\lambda_{MS}/\lambda_T}$ for each q pair of eigenvalues provide significant difference for their calculations. Some form of weighting for the calculation of this index must be used;

2. WA_T and WG_T indexes use the weighting approach through the explanation percentage of the eigenvalues extracted from the total variation matrix. These indexes (mainly WG_T) failed in scenarios with lower correlation structure due to higher weights assigned to less significant $\sqrt{\lambda_{MS}/\lambda_T}$ ratios.

3. Taking the simulation study into account, WA_{MS} and WG_{MS} were more robust indexes for assessing a multivariate measurement system. These indexes were able to overcome shortcomings such as: to provide an single assessment for all CTQs in multivariate GR&R study; to estimate indexes inside the confidence interval even when the correlation structure of CTQs is considered very low; and to provide a strategy of weighting that guarantee greater importance for the most significant $\sqrt{\lambda_{MS}/\lambda_T}$ ratio.

Acknowledgments

The authors would like to thank FAPEMIG, CNPq and, mainly, CAPES (through doctorate ‘sandwich’ program, Fellow – Process number 12944/12-2) for their support in this research.

References

AIAG, 2010. Measurement Systems Analysis: Reference Manual, fourth ed. Automotive Industry Action Group, Detroit, MI.

Al-Refaie, A., Bata, N., 2010. Evaluating measurement and process capabilities by GR&R with four quality measures. Measurement 43, 842-851.

Awad, M., Erdmann, T.P., Shanshal, Y., Barth, B., 2009. A measurement system analysis approach for hard-to-repeat events. Quality Engineering 21, 300-305.

Burdick, R.K., Borror, C.M., Montgomery, D.C., 2003. A review of methods for measurement systems capability analysis. Journal of Quality Technology 35, 342-354.

Costa, R., Angélico, D., Reis, M.S., Ataíde, J.M., Saraiva, P.M., 2005. Paper superficial waviness: conception and implementation of an industrial statistical measurement system. Analytica Chimica Acta 544, 135–142.

Dejaegher, B., Jimidar, M., De Smet, M., Cockaerts, P., Smeyers-Verbeke, J., Vander Heyden, Y., 2006. Improving method capability of a drug substance HPLC assay. Journal of Pharmaceutical and Biomedical Analysis 42, 155–170.

Deldossi, L., Zappa, D., 2011. Measurement uncertainty with nested mixed effect models. Quality and Reliability Engineering International 27, 373-379.

Erdmann, T.P., Does, R.J.M.M., Bisgaard, S., 2010. Quality quandaries: a gage R&R study in a hospital. Quality Engineering 22, 46-53.

Flynn, M.J., Sarkani, S., Mazzuchi, T.A., 2009. Regression analysis of automatic measurement systems. IEEE Transactions on Instrumentation and Measurement 58, 3373-3379.

He, S.G., Wang, G.A., Cook, D.F., 2011. Multivariate measurement system analysis in multisite testing: An online technique using principal component analysis. Expert Systems with Applications 38, 14602-14608.

Kaija, K., Pekkanen, V., Mäntysalo, M., Koskinen, S., Niittynen, J., Halonen, E., Mansikkamäki, P., 2010. Inkjetting dielectric layer for electronic applications. Microelectronic Engineering 87, 1984–1991.

Li, M.H.C., Al-Refaie, A., 2008. Improving wooden parts' quality by adopting DMAIC procedure. Quality and Reliability Engineering International 24, 351-360.

Majeske, K.D., 2008. Approval criteria for multivariate measurement systems. Journal of Quality Technology 40, 140-154.

Majeske, K.D. 2012. Approving vision-based measurement systems in the presence of within-part variation, Quality Engineering 24, 49-59.

Majeske, K.D., 2012. Two-sample tests for comparing measurement systems, Quality Engineering 24, 501-513.

Polini, W., Turchetta, S., 2004. Test protocol for micro-geometric wear of sintered diamond tools. Wear 257, 246-256.

Senol, S., 2004. Measurement system analysis using designed experiments with minimum $\alpha - \beta$ Risks and n. Measurement 36, 131-141.

Van Den Heuvel, E., Trip, A., 2002. Evaluation of measurement systems with a small number of observers. Quality Engineering 15, 323-331.

Wang, F.K., Chien, T.W., 2010. Process-oriented basis representation for a multivariate gauge study. Computers and Industrial Engineering 58, 143-150.

Wang, F.K., Yang, C.W., 2007. Applying principal component analysis to a GR&R study. Journal of the Chinese Institute of Industrial Engineering 24, 182-189.

Weaver, B.P., Hamada, M.S., Vardeman, S.B. Wilson, A.G., 2012. A Bayesian approach to the analysis of gauge R&R data, Quality Engineering 24, 486-500.

White, T.K., Borror, C.M., 2011. Two-dimensional guidelines for measurement system indexes. Quality and Reliability Engineering International 27, 479-487.

Woodall, W.H., Borror, C.M., 2008. Some relationships between gage R&R criteria. Quality and Reliability Engineering International 24, 99-106.

Wu, C.W., Pearn, W.L., Kotz, S., 2009. An overview of theory and practice on process capability indexes for quality assurance. International Journal of Production Economics 117, 338-359.