Acta Scientiarum

http://www.uem.br/acta
ISSN printed: 1806-2563
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v38i4.29294

Comparisons of multivariate GR&R methods using bootstrap confidence interval

Rogério Santana Peruchi[1*], Hélio Maciel Junior[2], Nilson José Fernandes[1], Pedro Paulo Balestrassi[2] and Anderson Paulo de Paiva[2]

*[1]Faculdade de Engenharia, Universidade Federal de Goiás, Avenida Doutor Lamartine Pinto de Avelar, 1120, 75704-020, Catalão, Goiás, Brazil. [2]Instituto de Engenharia e Gestão Industrial, Universidade Federal de Itajubá, Itajubá, Minas Gerais, Brazil. *Author for correspondence. E-mail: rogerioperuchi@ufg.br*

ABSTRACT. This paper aimed to compare the performance of multivariate GR&R (gage repeatability and reproducibility) studies based on PCA (principal component analysis) and Manova (multivariate analysis of variance) methods. To estimate the multivariate gauge index, geometric and arithmetic means have been implemented with and without weighting strategies. Bootstrap confidence interval based on BC_a (bias-corrected and accelerated) method has been adopted to determine multivariate gauge index adequacy. This confidence interval was calculated for the mean of univariate gauge indices estimated from each quality characteristic. The result analyses have shown that weighted approaches provided the best estimates of gauge index in multivariate GR&R studies.

Keywords: measurement system analysis, repeatability and reproducibility, multivariate analysis of variance, principal component analysis.

Comparações de métodos GR&R multivariados usando intervalo de confiança bootstrap

RESUMO. Este artigo teve objetivo de comparar o desempenho de estudos GR&R (*gage repeatability and reproducibility*) multivariados baseados nos métodos PCA (*principal component analysis*) e Manova (*multivariate analysis of variance*). As médias aritmética e geométrica com e sem ponderação foram implementadas para estimar os índices de medição multivariados. Para determinar a adequação dos índices de medição multivariados, foi adotado o intervalo de confiança bootstrap BCa (*bias-corrected and accelerated*). Esse intervalo de confiança foi calculado para a média dos índices de medição univariados estimados de cada característica da qualidade. As análises dos resultados mostraram que as abordagens ponderadas apresentaram melhores estimativas dos índices de avaliação em estudos GR&R com múltiplas variáveis.

Palavras-chave: análise de sistemas de medição, repetitividade e reprodutividade, análise multivariada de variância, análise de componentes principais.

Introduction

In any measurement process, at least part of the variation is due to the measurement system. It is unlikely that repeated measurements of any measurand results in exactly the same value (Senol, 2004; Majeske, 2008; Woodal & Borror, 2008; Automotive Industry Action Group [AIAG], 2010; Al-Refaie & Bata, 2010; Wang & Chien, 2010; Peruchi, Balestrassi, Paiva, Ferreira, & Carmelossi, 2013). Measurement system analysis (MSA) is a set of statistical techniques for ensuring that the measurement system variability is not significant in relation to manufacturing process variation. GR&R (gage repeatability and reproducibility) is the most common study in MSA to assess the precision of measurement systems (Peruchi, Paiva, Balestrassi, Ferreira, & Sawhney, 2014; Pereira, Peruchi, Paiva, Costa, & Ferreira, 2016). Repeatability is the variation of the measuring instrument or equipment assessing the same unit (operator or with the same setup and the same period of time). Reproducibility determines the variability arising from different operators, set-ups or period of time (Burdick, Borror, & Montgomery, 2003; Polini & Turchetta, 2004; Awad, Erdmann, Shanshal, & Barth, 2009; Wu, Pearn, & Kotz, 2009; Erdmann, Does, & Bisgaard 2009; Kaija et al., 2010; Weaver, Hamada, Vardeman, & Wilson, 2012; Peruchi et al., 2013). A measurement system is deemed adequate for monitoring a particular application, if R&R variation is relatively smaller than manufacturing process variation (Majeske, 2012; Pereira et al., 2016).

In GR&R studies two methods are usually utilized: (i) analysis of variance (ANOVA); and (ii) the $\overline{X}$ and R chart (Burdick et al., 2003; Wang & Chien, 2010). ANOVA is preferred due to its capacity of estimating the component of reproducibility from interaction between parts and operators. These methods are commonly applied to univariate cases; however, analysts often use more than one characteristic of the product to discriminate among different units (Burdick, Borror, & Montgomery, 2005). The analyst must consider the correlation structure among the characteristics to properly estimate the evaluation indices in these multivariate GR&R studies.

It is never possible to predict the exact values of variance components due to manufacturing and measurement variation in GR&R studies. Confidence intervals are used to quantify the uncertainty associated with the point estimation for each gauge variance component (Burdick et al., 2005). Wang and Li (2003) used Bootstrap method to obtain the confidence intervals of gauge variability when the control chart is used to find the point estimates. Wang and Chern (2012) evaluated the accuracy of the confidence interval for the circle-diameter with circular tolerances by using the Bootstrap method. In this particular research, the Bootstrap method has been applied upon univariate gauge capability indices in order to build confidence intervals. These confidence intervals were used as comparison criterion for evaluating performance of multivariate GR&R methods.

This article deals with repeatability and reproducibility studies applied to multivariate processes. Principal component analysis (PCA) and multivariate analysis of variance (Manova) are the most common multivariate methods used in such complex systems (Wang, 2013). The aim of this paper is to compare PCA and Manova methods with their variations to determine directions for practitioner conducting multivariate GR&R studies. The comparison criterion adopted in this research was the confidence intervals for the mean by BCa (bias-corrected and accelerated) bootstrap procedure of univariate evaluation indices of the measurement system. The results have shown that weighted approaches were the most effective strategies to calculate the evaluation index in multivariate GR&R studies.

Material and methods

In order to achieve the objective of this research, this section presents an overview of multivariate GR&R methods (Manova and PCA) and the bootstrap procedure to calculate the confidence interval. This was the criterion used to evaluate the performance of the multivariate evaluation indices of the measurement system. In the next section, three illustrative examples were assessed and some concerns about multivariate index estimates were provided. Last section addressed the main findings of this research.

GR&R based on multivariate analysis of variance

For GR&R studies considering two factors with interaction for q multiple quality characteristics, the model is given by Equation 1 (Majeske, 2008; Peruchi et al., 2014):

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

where:
$Y = (Y_1, Y_2, \ldots, Y_q)$ and $\mu = (\mu_1, \mu_2, \ldots, \mu_q)$ are constant vectors;
$\alpha_i \sim N(0, \Sigma_\alpha)$, $\beta_j \sim N(0, \Sigma_\beta)$, $\alpha\beta_{ij} \sim N(0, \Sigma_{\alpha\beta})$, and $\varepsilon_{ijk} \sim N(0, \Sigma_\varepsilon)$ are random vectors statistically independent of each other. Variance components in Equation 1 can be estimated using the Manova method proposed by Majeske (2008). These variance components are estimated for obtaining an index that evaluates acceptance of the measurement system, called $\%R\&R_m$ (variation percentage due to repeatability and reproducibility). The index $\%R\&R_m$, or G index for this Manova method, can be calculated by Equation 2:

$$G = \left(\prod_{i=1}^{q} \sqrt{\frac{\lambda_{ms_i}}{\lambda_{t_i}}} \right)^{1/q} 100\% \quad (2)$$

where:
λ_{ms} and λ_t are eigenvalues extracted from the variance-covariance matrices for measurement system ($\hat{\boldsymbol{\Sigma}}_{ms}$) and total variation ($\hat{\boldsymbol{\Sigma}}_t$), respectively. $\%R\&R_m$ less than 10% requires that the measurement system is deemed acceptable. If the index lies in a marginal region between 10 and 30%, the measurement system may be acceptable depending on the application, the measuring device cost, repair cost, or other factors. Moreover, the measurement system is considered unacceptable if the index exceeds 30% (Li & Al-Refaie, 2008; Woodall & Borror, 2008; AIAG, 2010).

To estimate the evaluation index of the measurement system, Equation 2 applies geometric mean on $\sqrt{\lambda_{ms}/\lambda_t}$ ratio. This strategy does not determine the utmost importance for the most significant pairs of eigenvalues extracted from the variance-covariance matrices. Thus, Peruchi, Paiva, Balestrassi, Ferreira, and Sawhney (2014) adopted a weighted approach on $\sqrt{\lambda_{ms}/\lambda_t}$ to propose four new evaluation indices for multivariate measurement systems (%$R\&R_m$). The new indices WA_t, WA_{ms}, WG_t and WG_{ms} can be estimated using Equations 3 and 4.

$$WA = \sum_{i=1}^{q}\left(W_i\sqrt{\frac{\lambda_{ms_i}}{\lambda_{t_i}}}\right)100\% \quad (3)$$

$$WG = \prod_{i=1}^{q}\left(\sqrt{\frac{\lambda_{ms_i}}{\lambda_{t_i}}}\right)^{W_i} 100\% \quad (4)$$

where:

$W_i\ \forall i = 1,\ldots q$ determines the explanation percentage of the eigenvalues extracted from either $\hat{\mathbf{\Sigma}}_t : W_i = \left(\lambda_{t_i}\Big/\sum_{j=1}^{q}\lambda_{t_j}\right)$ or $\hat{\mathbf{\Sigma}}_{ms} : W_i = \left(\lambda_{ms_i}\Big/\sum_{j=1}^{q}\lambda_{ms_j}\right)$. The WA_t and WA_{ms} indices are calculated by the weighted arithmetic mean in Equation 3. On the other hand, the WG_t and WG_{ms} indices are estimated using weighted geometric mean according to Equation 4.

GR&R based on principal component analysis

According to Wang and Chien (2010) and Peruchi, Balestrassi, Paiva, Ferreira, and Carmelossi (2013), to deal with q multiple quality characteristics in GR&R studies, PCA is an alternative method to Manova. The model that represents a multivariate GR&R study using PCA is given by Equation 5:

$$\mathbf{PC_n} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}\begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,o \\ k = 1,2,\ldots,r \\ n = 1,2,\ldots,q \end{cases} \quad (5)$$

where:

PC_n are scores of principal components PC_1, PC_2, ..., PC_q;

μ is a constant;

α_i, β_j, $\alpha\beta_{ij}$ and ε_{ijk} are independent normal random variables with zero means and variances σ_α^2, σ_β^2, $\sigma_{\alpha\beta}^2$ and σ_ε^2 respectively. The %$R\&R_m$ evaluation index of the measurement system is obtained by Equation 6 through the PCA method. More details on how to obtain the scores of principal components and how to evaluate the measurement system using the PCA method, see Wang (2013) and Wang and Chien (2010). The measurement system acceptance criteria are the same as described in the previous subsection.

$$\%R\ \&\ R_{PC} = \left(\frac{\sigma_{ms}}{\sigma_T}\right)100\,\% \quad (6)$$

Wang and Chien (2010) compared the PCA method with two other methods for analyzing the measurement system. However, these authors performed individual analysis for each principal component. This methodology may not be appropriate since the individual analysis might provide different interpretations. When responses are highly correlated (e.g., $\%PC_1 > 95\%$), the first principal component explains reasonably well measurement system's variability. However, when the correlations between the responses are medium or low, additional principal components must be assessed, since the first principal component is incapable of explaining the entire variation of the original responses. Consequently, Peruchi et al. (2013) proposed a method for multivariate GR&R studies using weighted principal components (WPC). In this case, the model in Equation 5 is modified by weighting the scores of principal components based on their respective eigenvalues. The response vector to be analyzed in Equation 5 should be Equation 7:

$$WPC = \sum_{i=1}^{q}\left[\lambda_i\left(PC_i\right)\right] \quad (7)$$

or using the explanation percentage of each principal component as such, according Equation 8:

$$WPC = \sum_{i=1}^{q}\left[\frac{\lambda_i}{\sum_{j=1}^{q}\lambda_j}\left(PC_i\right)\right] \quad (8)$$

The measurement system evaluation index using WPC method follows Equation 6, however, all computations are based on weighted scores of principal components.

Comparison criterion based on Bootstrap confidence interval

Bootstrap is a computational method for assigning accuracy measures of statistical estimates (Efron & Tibshirani, 1993). Confidence intervals is one of the areas that the bootstrap procedure has achieved greater success (Wehrens, Putter, & Buydens, 2000). According to Wang and Chern (2012), the standard method assumes $\mu_{Y(i)}$ and $S_{Y(i)}$ be the mean and standard deviation of the *Y(i)*, with i = 1, 2, ..., B, where *Y(i)* is the i^{th} Bootstrap sample data set and *B* is the number of Bootstrap samples. The bootstrap mean and standard deviation are calculated as such, according Equations 9 and 10:

$$\overline{\mu}_{Bootstrap} = \frac{\sum_{i=1}^{B} \overline{\mu}_{Y(i)}}{B} \quad (9)$$

$$S_{Bootstrap} = \sqrt{\frac{\sum_{i=1}^{B} \left(\overline{Y}(i) - \overline{\mu}_{Bootstrap}\right)^2}{B}} \quad (10)$$

Hence, a 100 (1-α)% confidence interval for $\mu_{Y(i)}$ is Equation 11:

$$\overline{\mu}_{Bootstrap} - Z_{\alpha/2} S_{Bootstrap} \le \mu_{Y(i)} \le \overline{\mu}_{Bootstrap} + Z_{\alpha/2} S_{Bootstrap} \quad (11)$$

where:

$Z_{\alpha/2}$ is the upper α/2 quartile of the standard normal distribution.

The bootstrap confidence interval (BCI) should not only reproduce results quite similar to statistical theoretical calculation but also provide adequate coverage probability. Bootstrap-t method presents reasonable theoretical coverage, but it tends to be unstable in practical situations. The percentile method is more stable, however it determines poor coverage properties. The BC_a (bias-corrected and accelerated) method is an improved version of the previously mentioned ones. BC_a considers both lack of symmetry in data distribution and adaptative shape when the statistics of interest varies (Efron & Tibshirani, 1993). The first two steps of BC_a method are identical to the bias-corrected percentile method. In the third step, accelerated corrected percentile endpoints of the standard normal distribution are obtained by Equations 12 and 13 (Wang & Chern, 2012):

$$P_{AL} = \Phi\left(Z_{P_0} + \frac{Z_{P_0} + Z_{\alpha/2}}{1 - a\left(Z_{P_0} + Z_{\alpha/2}\right)} \right) \quad (12)$$

$$P_{AU} = \Phi\left(Z_{P_0} + \frac{Z_{P_0} + Z_{1-\alpha/2}}{1 - a\left(Z_{P_0} + Z_{1-\alpha/2}\right)} \right) \quad (13)$$

where:

$$a = \frac{\sum_{i=1}^{n} \left(\overline{\mu}_{Y(\cdot)} - \dot{\mu}_{Y(i)}\right)^3}{6\left[\sum_{i=1}^{n} \left(\overline{\mu}_{Y(\cdot)} - \mu_{Y(i)}\right)^2\right]^{3/2}},$$

$\dot{\mu}_{Y(i)}$ is calculated from the original sample with i^{th} point deleted, and $\overline{\mu}_{Y(\cdot)} = \sum_{i=1}^{n} \dot{\mu}_{Y(i)}/n$. Thus, a 100 (1-α)% confidence interval for μ_Y using the BC_a method is Equation 14:

$$\mu_Y\left(P_{AL}B\right) \le \mu_Y \le \mu_Y\left(P_{AU}B\right) \quad (14)$$

As mentioned previously, BC_a confidence interval of univariate gauge indices was the criterion for multivariate GR&R method adequacy. The algorithm to estimating these confidence intervals is given as follows:

I. Select parts, operators, number of replicates and, then collect the dataset;

II. Perform univariate GR&R study and estimate *%R&R* index for each response;

III. Using BC_a method, generate a Bootstrap resample (*B* = 2000) from the univariate *%R&R* indices;

IV. Compute the confidence interval for the mean of *%R&R* index at the 95% confidence level using BCa method.

A multivariate GR&R method was deemed adequate if its gauge index, using either Equations 2 at 4 and 6, was estimated within the bootstrap confidence interval.

Results and discussion

Case 1: automotive body stamped-panel measurement system

Majeske (2008) presented an application of the automobile industry in which an analyst built a machine to measure four *Y* characteristics in a sheet steel panel. This study consisted of p = 5 parts, the operators *o* = 2, and *r* = 3 replicates. Adjusting the data to an ANOVA model, variance components for manufacturing process (part-to-part), measurement system, total variation, and the univariate index *%R&R* were estimated and stored in Table 1. More details on how to calculate *%R&R* using ANOVA method can be seen in Automotive Industry Action Group (AIAG, 2010).

Applying the proposed procedure for estimating the BC_a confidence interval of *%R&R* indices, the first two steps have already been performed, as can be seen in Table 1. Then, 2000 bootstrap sample was generated from the *%R&R* indices as suggested by

Wang and Chern (2012) and Efron and Tibshirani (1993). After that, the bootstrap confidence interval [10.86 and 20.42%] based on BC_a method was built using Equations 12 at 14. These BC_a confidence intervals have been estimated by using Matlab® software. Eventually, variance-covariance matrices (Manova method) and standard deviation based on scores of principal components (PCA method) for manufacturing process (part-to-part), measurement system, and total variation were estimated and stored in Table 2.

Table 1. Variation components, univariate gauge indices and bootstrap confidence interval for case 1.

Source	Y_1	Y_2	Y_3	Y_4	BLCL	BUCL
Part-to-part	0.018	0.252	0.208	0.986		
Measurement System	0.001	0.006	0.005	0.008		
Total Variation	0.019	0.258	0.213	0.995		
%R&R	22.20	15.66	15.09	9.26	10.86	20.42

%R&R$_m$ indices based on Manova were calculated by extracting eigenvalues from variance-covariance matrices, in Table 2, using Equatons 2 at 4. *%R&R$_m$* indices using PCA method were obtained by standard deviation related to either scores or weighted scores of principal components, according to Equation 6. Additionally, Figure 1 illustrates the multivariate evaluation indices and the BC_a confidence intervals estimated from case 1. The multivariate indices calculated by Manova presented estimates within the bootstrap confidence interval [10.86; 20.42], using both simple geometric mean (*G* index) and weighted approaches for arithmetic and geometric means (WA_t, WG_t, WA_{ms} and WG_{ms} indices). Through the PCA method, the principal components PC_1, PC_2 and PC_3 together account for 99% explanation of the original variables. PC_1 and PC_2 estimated within the BCI, but PC_3 (*%R&R$_m$* = 9.6%) was estimated outside BCI. Wang and Chien (2010) recommended evaluating components representing at least 95% of explanation, so this approach was deemed failed. Through the weighted arithmetic mean of the principal component scores, WPC adequately estimated the multivariate index of the measurement system.

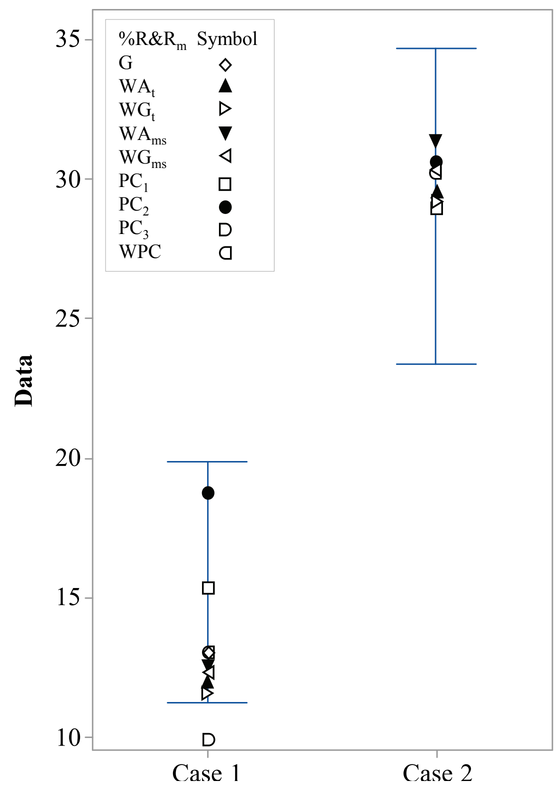

Figure 1. Multivariate gauge indices and bootstrap confidence intervals for cases 1 and 2; Source: the authors.

Table 2. Variation components and multivariate gauge indices for case 1.

Source	Manova: G	WA_t	WG_t	WA_{ms}	WG_{ms}	PCA: PC_1	PC_2	PC_3	WPC
Part-to-part	$\begin{pmatrix} 0.01811 & 0.01600 & -0.02180 & 0.00763 \\ 0.01600 & 0.25163 & -0.15732 & 0.35463 \\ -0.02180 & -0.15732 & 0.20856 & -0.39249 \\ 0.00763 & 0.35463 & -0.39249 & 0.98631 \end{pmatrix}$					3.040	1.194	0.413	21.880
Measurement system	$\begin{pmatrix} 0.00094 & 0.00168 & -0.00141 & 0.00189 \\ 0.00168 & 0.00632 & -0.00475 & 0.00702 \\ -0.00141 & -0.00475 & 0.00486 & -0.00581 \\ 0.00189 & 0.00702 & -0.00581 & 0.00852 \end{pmatrix}$					0.077	0.042	0.004	0.335
Total variation	$\begin{pmatrix} 0.01905 & 0.01768 & -0.02321 & 0.00574 \\ 0.01768 & 0.25795 & -0.16207 & 0.36165 \\ -0.02321 & -0.16207 & 0.21342 & -0.39890 \\ 0.00574 & 0.36165 & -0.39830 & 0.99483 \end{pmatrix}$					3.117	1.236	0.417	22.215
%R&R$_m$	12.28[a]	11.73[a]	11.66[a]	12.00[a]	11.92[a]	15.70[b]	18.36[b]	9.60	12.28[b]

[a]evaluation index within the confidence interval based on Manova; [b]evaluation index within the confidence interval based on PCA.

Case 2: turning process measurement system

A recent study by Peruchi et al. (2014) analyzed roughness measurements of work pieces made up of AISI 12L14 steel from a turning process. Five roughness parameters were evaluated in a multivariate GR&R study with p = 12 parts, *o* = 3 operators and *r* = 4 replicates. Similarly to the case 1, variance components for manufacturing process (part-to-part), measurement system, total variation, and the univariate index *%R&R* were estimated and presented in Table 3.

Table 3. Variation components, univariate gauge indices and bootstrap confidence interval for case 2.

Source	R_a	R_y	R_z	R_q	R_t	BLCL	BUCL
Part-to-part	0.444	1.564	1.383	0.456	1.696		
Measurement system	0.082	0.646	0.428	0.111	0.643		
Total Variation	0.452	1.693	1.448	0.470	1.813		
%R&R	18.22	38.18	29.52	23.66	35.47	22.66	35.36

The first two steps of the proposed procedure of bootstrap confidence intervals have already been conducted, as seen in Table 3. Then, 2000 bootstrap samples were generated from the *%R&R* indices. After that, the bootstrap confidence interval [22.66 and 35.36%] based on BC_a method was built using Equations 12 at 14. Finally, variance-covariance matrices and standard deviation based on scores of principal components for manufacturing process (part-to-part), measurement system, and total variation were estimated and stored in Table 4.

Table 4. Variation components and multivariate gauge indices for case 2.

Source	Manova					PCA		
	G	WA_t	WG_t	WA_{ms}	WG_{ms}	PC_1	PC_2	WPC
	1.045	0.621	0.839	1.021	0.604			
	0.621	0.916	0.873	0.684	0.926			
P	0.839	0.873	1.016	0.894	0.869	2.045	0.822	8.656
	1.021	0.684	0.894	1.019	0.668			
	0.604	0.926	0.869	0.668	0.939			
	0.035	0.047	0.044	0.042	0.044			
	0.047	0.153	0.099	0.070	0.132			
MS	0.044	0.099	0.092	0.065	0.089	0.619	0.262	2.743
	0.042	0.070	0.065	0.059	0.065			
	0.044	0.132	0.089	0.065	0.133			
	1.080	0.668	0.883	1.063	0.647			
	0.668	1.069	0.972	0.754	1.059			
T	0.883	0.972	1.108	0.959	0.958	2.137	0.863	9.080
	1.063	0.754	0.959	1.078	0.732			
	0.647	1.059	0.958	0.732	1.071			
$\%R\&R_m$	44.64	29.30[a]	29.12[a]	30.92[a]	30.23[a]	28.97[b]	30.31[b]	30.21[b]

[a]evaluation index within the confidence interval based on Manova; [b]evaluation index within the confidence interval based on PCA.

$\%R\&R_m$ indices based on Manova were calculated by extracting eigenvalues from variance-covariance matrices, in Table 4, using Equations 2 at 4. $\%R\&R_m$ indices using PCA method were obtained by standard deviation related to either scores or weighted scores of principal components, according to Equation 6. Figure 1 also shows the multivariate evaluation indices and the BC_a confidence intervals estimated from case 2. Based on Manova method, the result using *G* index ($\%R\&R_m$ = 44.64%) showed that simple geometric mean was unable to estimate the multivariate index within the bootstrap confidence interval [22.66 and 35.36%]. Nevertheless, the weighted approaches (WA_t, WG_t, WA_{ms} and WG_{ms} indices) presented satisfactory results to classify the measurement system. Using PCA method, similar performance was observed for estimating the multivariate evaluation indices. PC_1 and PC_2 represented 98.6% explanation of the original variables and estimated the multivariate index within the confidence interval. As seen in Table 4, *WPC* index has also been effective classifying the measurement system.

Case 3: simulated data analysis

Peruchi et al. (2013) presented a simulation study for multivariate GR&R using the same setup in Majeske (2008). The authors simulated 15 scenarios considering several correlation structures for *Y*s and different types of measurement systems. Assessing this dataset using ANOVA method, univariate indices were estimated to four quality characteristics at each scenario. Table 5 shows the *%R&R* indices and the bootstrap confidence interval obtained by the proposed procedure.

Table 5. Simulation study scenarios, univariate gauge indices and bootstrap confidence interval for the case 3.

Scenario			Univariate (*%R&R*)				BCI	
S	MS	Corr.	Y_1	Y_2	Y_3	Y_4	*BLCL*	*BUCL*
S1	UN	VL	49.9	39.3	38.3	34.1	36.20	47.25
S2	UN	L	42.2	55.5	44.3	39.8	41.00	52.70
S3	UN	M	40.8	52.4	42.6	36.9	38.85	49.95
S4	UN	H	45.3	33.2	41.2	47.8	35.20	46.15
S5	UN	VH	31.1	34.9	37.8	41.1	32.78	39.55
S6	MA	VL	15.8	14.1	13.7	10.2	11.08	14.95
S7	MA	L	18.6	27.2	21.3	24.1	19.95	25.72
S8	MA	M	15.5	23.7	17.0	14.6	15.20	22.02
S9	MA	H	13.2	10.3	13.6	16.9	11.12	16.08
S10	MA	VH	15.2	19.0	19.7	20.9	16.15	20.42
S11	AC	VL	8.4	6.3	4.9	5.3	5.25	7.88
S12	AC	L	5.6	4.6	6.7	5.4	4.85	6.38
S13	AC	M	6.2	9.6	6.6	5.9	6.08	8.85
S14	AC	H	5.7	4.5	5.9	7.3	4.85	6.95
S15	AC	VH	6.5	7.6	8.6	9.2	6.78	8.90

Using Equations 2 at 4 and 6, multivariate gauge indices were also estimated for each scenario. Table 6 presents these indices obtained by Manova and PCA methods.

Table 6. Comparison of gauge indices for multivariate measurement system in case 3.

Scenarios	Manova					PCA			
S	G	WA_t	WG_t	WA_{ms}	WG_{ms}	PC_1	PC_2	PC_3	WPC
S1	10.78	31.62	18.28	48.84	48.51	52.24	19.55	15.32	39.71[b]
S2	13.30	36.50	30.28	45.42[a]	44.94[a]	53.84	10.48	20.91	52.84
S3	11.32	38.27	31.45	45.76[a]	45.51[a]	47.79[b]	11.21	17.65	47.87[b]
S4	28.15	42.95[a]	42.20[a]	44.33[a]	44.14[a]	44.38[b]	8.94		44.03[b]
S5	64.09	35.81[a]	35.79[a]	36.05[a]	35.94[a]	36.10[b]			36.11[b]
S6	4.97	9.62	6.48	15.67	15.38	18.46	6.92	2.79	12.65[b]
S7	10.04	19.86	14.68	27.15	26.87	24.97[b]	2.77	8.37	26.98
S8	5.40	15.38[a]	13.24	18.02[a]	17.90[a]	19.71[b]	6.01	11.78	19.86[b]
S9	14.31[a]	14.37[a]	14.35[a]	14.44[a]	14.43[a]	14.17[b]	6.01		14.00[b]
S10	47.23	16.95[a]	16.94[a]	17.33[a]	17.12[a]	18.63[b]			18.63[b]
S11	4.08	5.00	4.41	6.75[a]	6.49[a]	6.41[b]	6.39[b]	4.59	4.10
S12	2.01	4.70	3.53	6.18[a]	6.13[a]	6.69	1.15	2.25	6.89
S13	2.28	6.07	5.32	7.05[a]	7.00[a]	7.87[b]	1.95	5.75	8.04[b]
S14	7.22	6.58[a]	6.56[a]	6.65[a]	6.63[a]	5.99[b]	3.34		5.91[b]
S15	39.35	7.78[a]	7.78[a]	8.18[a]	7.89[a]	7.92[b]			7.92[b]

[a]evaluation index within the confidence interval based on Manova; [b]evaluation index within the confidence interval based on PCA.

Figure 2 presents the multivariate evaluation indices and BC_a confidence intervals of simulated scenarios with unacceptable measurement systems. Indices obtained by both Manova with simple geometric mean (*G* index) and PCA with individual analysis of principal components (PC_1, PC_2 and/or PC_3 indices) have represented the worst estimates. Effectiveness was observed only in one (S9) and three (S5, S10 and S15) scenarios, respectively. Weighted Manova using eigenvalues extracted from total variation matrix determined moderate effectiveness. WA_t and WG_t estimated the multivariate evaluation index within BCI in seven (S4, S5, S8, S9, S10, S14 and S15) and six (S4, S5, S9, S10, S14 and S15) scenarios, respectively.

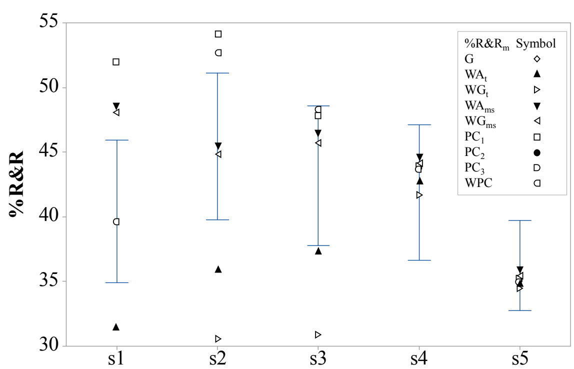

Figure 2. Multivariate gauge indices and bootstrap confidence intervals for S1-S5 scenarios; Source: the author.

In this simulation study, the most effective approaches, in estimating the evaluation index of the measurement system, were weighted Manova based on eigenvalues extracted from measurement system matrix (WA_{ms} and WG_{ms} indices) and weighted principal components (*WPC*). According to 95% bootstrap confidence interval, WA_{ms}, WG_{ms} and *WPC* have failed only on three (S1, S6 and S7), three (S1, S6 and S7) and four (S2, S7, S11 and S12) scenarios, respectively.

Results and discussion

Taking into account the aforementioned results, Table 7 summarizes the performance of multivariate methods for distinct types of measurement systems and several correlation structures among quality characteristics. Comparing the multivariate indices to the bootstrap confidence interval, weighted approaches based on WA_{ms}, WG_{ms} and *WPC* have presented the best performances. WA_{ms} and WG_{ms} weight the $\sqrt{\lambda_{ms}/\lambda_t}$ ratio with the explanation percentage of the eigenvalues extracted from measurement system matrix ($\hat{\Sigma}_{ms}$), using Equations 3 and 4. As seen in Table 7, this strategy showed better estimates than G, WA_t and WG_t indices. Accordingly, WPC weights each principal component with their respective eigenvalues using Equation 7. Table 7 determines that evaluating each principal component individually is inadequate.

Table 7. Overview of multivariate analyses of measurement systems.

Cases 1, 2 and 3			Manova					PCA	
MS	Corr.	Evidence	G	WA_t	WG_t	WA_{ms}	WG_{ms}	PC_i	WPC
	VL	S1							✔
	L	S2				✔	✔		
UN	M	S3				✔	✔		✔
	H	S4 and Case 2		✔	✔	✔	✔		✔
	VH	S5		✔	✔	✔	✔	✔	✔
	VL	S6 and Case 1							✔
	L	S7							
MA	M	S8		✔		✔	✔		✔
	H	S9 and Case 2		✔	✔	✔	✔		✔
	VH	S10		✔	✔	✔	✔	✔	✔
	VL	S11				✔	✔		
	L	S12				✔	✔		
AC	M	S13				✔	✔		✔
	H	S14		✔	✔	✔	✔		✔
	VH	S15		✔	✔	✔	✔	✔	✔

Nevertheless, it is essential to highlight that low or very low correlation structures among characteristics deserve special attention. In such multivariate scenario, even weighted approaches had presented poor performance. Therefore, practitioners should estimate the multivariate index carefully by using both Manova and PCA methods in order to ensure that the measurement system was properly classified. Furthermore, additional indices such as ‘ndc’ (number of distinct categories) and *%P/T* (percentage of precision-to-tolerance) may be calculated with the aim of determining properly the contribution of variation due to repeatability and reproducibility.

Conclusion

This article has investigated the multivariate analysis of measurement systems through repeatability and reproducibility studies. The main contribution of this research was to develop an extensive comparison of multivariate GR&R studies using Manova and PCA methods. Differently from previous works (Peruchi et al., 2013; 2014), better estimates for confidence intervals were provided by bias-corrected and accelerated bootstrap procedure (BC_a). The result analyses have shown that weighted approaches were the most effective strategies for estimating the evaluation index in multivariate measurement systems. As seen in Table 7, multivariate gauge indices using WA_{ms}, WG_{ms} and *WPC* obtained success in 13, 13 and 12 scenarios, respectively. Even though in few scenarios these strategies have failed, the estimates were quite close to the bootstrap confidence limits. Further study can be extended to other multivariate indices such as 'ndc' and *%P/T*. Moreover, expanded GR&R and nested GR&R applied to multivariate processes deserve special attention in future researches.

Acknowledgements

For their support in this research, the authors would like to thank the Brazilian agencies: Fapemig, Fapeg, CNPq and Capes.

References

Al-Refaie, A., & Bata, N. (2010). Evaluating measurement and process capabilities by GR&R with four quality measures. *Measurement, 43*(6), 842-851.

Automotive Industry Action Group. (2010). *Measurement systems analysis: reference manual* (4th ed.). Southfield, MI: Chrysler.

Awad, M., Erdmann, T. P., Shanshal, Y., & Barth, B. (2009). A measurement system analysis approach for hard-to-repeat events. *Quality Engineering, 21*(3), 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*(4), 342-354.

Burdick, R. K., Borror, C. M., & Montgomery, D. C. (2005). *Design and analysis of gauge R&R studies: Making decisions with confidence intervals in random and mixed ANOVA models*. Alexandria, EG: ASA-SIAM Series on Statistics and Applied Probability.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York City, NY: Chapman & Hall.

Erdmann, T. P., Does, R. J., & Bisgaard, S. (2009). Quality quandaries*: A gage R&R study in a hospital. *Quality Engineering, 22*(1), 46-53.

Kaija, K., Pekkanen, V., Mäntysalo, M., Koskinen, S., Niittynen, J., Halonen, E., & Mansikkamäki, P. (2010). Inkjetting dielectric layer for electronic applications. *Microelectronic Engineering, 87*(10), 1984-1991.

Li, M. H. C., & Al-Refaie, A. (2008). Improving wooden parts' quality by adopting DMAIC procedure. *Quality and Reliability Engineering International, 24*(3), 351-360.

Majeske, K. D. (2008). Approval criteria for multivariate measurement systems. *Journal of Quality Technology, 40*(2), 140-153.

Majeske, K. D. (2012). Two-sample tests for comparing measurement systems. *Quality Engineering, 24*(4), 501-513.

Pereira, R. B. D., Peruchi, R. S., Paiva, A. P., Costa, S. C., & Ferreira, J. R. (2016). Combining Scott-Knott and GR&R methods to identify special causes of variation. *Measurement, 82*, 135-144.

Peruchi, R. S., Balestrassi, P. P., Paiva, A. P., Ferreira, J. R., & Carmelossi, M. S. (2013). A new multivariate gage R&R method for correlated characteristics. *International Journal of Production Economics, 144*(1), 301-315.

Peruchi, R. S., Paiva, A. P., Balestrassi, P. P., Ferreira, J. R., & Sawhney, R. (2014). Weighted approach for multivariate analysis of variance in measurement system analysis. *Precision Engineering, 38*(3), 651-658.

Polini, W., & Turchetta, S. (2004). Test protocol for micro-geometric wear of sintered diamond tools. *Wear, 257*(3), 246-256.

Senol, S. (2004). Measurement system analysis using designed experiments with minimum α–β Risks and n. *Measurement, 36*(2), 131-141.

Wang, F. K. (2013). An assessment of gauge repeatability and reproducibility with multiple characteristics. *Journal of Testing and Evaluation, 41*(4), 651-658.

Wang, F. K., & Chern, H. L. (2012). Confidence intervals for two-dimensional data with circular tolerances in a gauge R&R study. *Quality & Quantity, 46*(1), 55-69.

Wang, F. K., & Chien, T. W. (2010). Process-oriented basis representation for a multivariate gauge study. *Computers & Industrial Engineering, 58*(1), 143-150.

Wang, F. K., & Li, E. Y. (2003). Confidence intervals in repeatability and reproducibility using bootstrap method. *Total Quality Management, 14*(3), 341-354.

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*(4), 486-500.

Wehrens, R., Putter, H., & Buydens, L. M. (2000). The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems, 54*(1), 35-52.

Woodall, W. H., & Borror, C. M. (2008). Some relationships between gage R&R criteria. *Quality and Reliability Engineering International, 24*(1), 99-106.

Wu, C. W., Pearn, W. L., & Kotz, S. (2009). An overview of theory and practice on process capability indices for quality assurance. *International Journal of Production Economics, 117*(2), 338-359.

Received on September 26, 2015.
Accepted on March 17, 2016.