STATISTICS IN TRANSITION new series, September 2019
Vol. 20, No. 3, pp. 31–56, DOI 10.21307/stattrans-2019-023
Submitted – 20.07.2019; Paper ready for publication – 11.08.2019

OUTLIER DETECTION IN THE ANALYSIS OF NESTED GAGE R&R, RANDOM EFFECT MODEL

Mohammed Abduljaleel[1], Habshah Midi[2], Mostafa Karimi[3]

ABSTRACT

Measurement system analysis is a comprehensive valuation of a measurement process and characteristically includes a specially designed experiment that strives to isolate the components of variation in that measurement process. Gage repeatability and reproducibility is the adequate technique to evaluate variations within the measurement system. Repeatability refers to the measurement variation obtained when one person repeatedly measures the same item with the same Gage, while reproducibility refers to the variation due to different operators using the same Gage. The two factors factorial design, either crossed or nested factor, is usually used for a Gage R&R study. In this study, the focus is only on the nested factor, random effect model. Presently, the classical method (the method of analysing data without taking into consideration the existence of outliers) is used to analyse the nested Gage R&R data. However, this method is easily affected by outliers and, consequently, the measurement system's capability is also affected. Therefore, the aims of this study are to develop an identification method to detect outliers and to formulate a robust method of measurement analysis of nested Gage R&R, random effect model. The proposed methods of outlier detection are based on a robust mm location and scale estimators of the residuals. The results of the simulation study and real numerical example show that the proposed outlier identification method and the robust estimation method are the most successful methods for the detection of outliers.

Key words: measurement system analysis, mm location, nested Gage R&R, outlier, residuals.

1. Introduction and background

Control is a contentious word that on occasions can be identified with having power (Macintosh and Quattrone, 2010) or training oppression, but in a structural background it has been defined as the ability to create and monitor rules and regulations which should be followed (Ouchi and Maguire, 1975) or on the

[1] Ministry of Electricity, Baghdad, Iraq.
[2] Department of Applied and Computational Statistics, Institute for Mathematical Research, University Putra Malaysia, Serdang, Selangor 43400, Malaysia. E-mail: habshahmidi@gmail.com
[3] Department of Applied and Computational Statistics, Institute for Mathematical Research, University Putra Malaysia, Serdang, Selangor 43400, Malaysia. E-mail: mostafa.karimi.ir@gmail.com

opposing, has been seen as a routine, uninteresting task of observing, supervising, measuring and providing feedback (Reeves and Woodward, 1970). Whatever the definition, the concept is viewed by many as the central nervous system of the processes in every organization.

Montgomery (2007) clarified that the quality control system always has been an integral part of virtually all products and services. However, wakefulness of its importance and the introduction of formal methods for quality control and improvement have been an evolutionary development. An important part of the statistical quality control is the six sigma system (Smith, 1993). Six Sigma is a severe, focused and highly effective application of proven quality principles and techniques. Companies operating at six sigma typically spend less than 5 per cent of their profits fixing problems. In contrast with non-six sigma companies, these costs are often extremely high. Companies operating at three or four sigmas typically spend between 25 and 40 per cent of their profits fixing problems (Pyzdek and Keller, 2014). Based on (Kwak and Anbari, 2006), the authors showed that understanding the key features, impediments, and confines of the six sigma method allow organizations to better support their strategic directions and increasing needs for monitoring and training. Although Six Sigma provides assistance over prior approaches to quality management, it also creates new challenges for researchers and experts (Schroeder et al., 2008).

The important part of the six sigma quality is the measurement system analysis (MSA) used to isolate the variation among devices being measured from the error in the measurement system. The measurement system analysis has been the focus of substantial attention because of its ability to determine the level and range of variation in data. In a process that is important to a measurement system, some variation is likely to occur. The measurement system analysis is an important part of a study that is able to determine the amount of variation (Bourne et al., 2007).

To ensure that the measurement system variability is not adversely large, it is necessary to conduct the measurement system analysis (MSA). Such a study can be conducted in virtually any type of manufacturing industry. According to (He et al., 2011), MSA helps to measure the ability of a Gage or measuring device to produce data that support the analyst's decision-making requirements. Also, MSA is an important section of Six Sigma as well as of the ISO/TS 16949 standards. Burdick et al., (2003) showed that Gage repeatability and reproducibility (Gage R&R) is the most common study in MSA to assess the precision of measurement systems.

Awad et al., (2009) and Peruchi et al., (2013) showed that the repeatability represents the variability from the Gage or measurement tool when it is used to measure the same part (with the same operator or setup or in the same time period), whereas reproducibility reveals the variability arising from different operators, setups, or time periods. As stated by (Grejda et al., 2005; Parker et al., 2005; Piratelli-Filho et al., 2012), some works have used repeatability and/or reproducibility perceptions and ignored Gage R&R statistical analysis in matching measurement system variation to process variation. These studies comprise only Gage variability which are lacking to determine whether the measurement system is able to monitor a particular manufacturing process or not. In case the variation of the measurement system is small relative to the variation of the process, the measurement system is reflected as acceptable measurements. Furthermore, Gage R&R studies must be performed any time a process is adopted. This is because as the process variation decreases, a once-capable measurement system may now be unqualified. As identified by (Burdick et al., 2003; Wang and Chien, 2010), there are two methods commonly used in the analysis of a Gage R&R study, namely the analysis of variance ANOVA, X bar and R chart. Furthermore, analysts prefer the ANOVA method because it measures the operator-to-part interaction Gage error.

Larsen (2003) extended the univariate Gage R&R study to a common industrial test scenario where multiple features were tested on each device. Providing examples from an industrial application, the author showed that the total yield, false failures and missed false estimates could lead to improvements in the production test process and hence to lower production costs, and finally to customers receiving higher quality products. (Flynn et al., 2009) used regression analysis to analyse the qualified performance capability between two functionally equal but technically different automatic measurement systems. For such accurate measurements as repeatability and reproducibility, the authors found the "pass/fail" criteria for the unit being tested incorrect. Hence, they proposed a methodology based on principal components analysis (PCA) and MANOVA to examine whether there was a statistically significant difference among the measurement systems. He et al., (2011) proposed a PCA-based approach in MSA for the in-process monitoring of all instruments in multisite testing. The approach considers a defective instrument to be one whose statistical distribution of measurements differs significantly from the overall distribution across multiple test tools. Their approach can be implemented as an online monitoring procedure for test instruments so that, until a faulty instrument is identified, production goes continuously. Whereas, Parente et al., (2012) applied univariate and multivariate methods to evaluate the repeatability and reproducibility of the measurement of opposite phase chromatography (RP-HPLC) peptide profiles of excerpts from cheddar cheese. The ability to discriminate different samples was assessed according to the sources of variability in their measurement and analysis procedure. The authors showed that their study had an important impact on the design and analysis of experiments for summarizing of cheese proteolysis. Inferential statistical procedures helped them to analyse the relationships between design variables and proteolysis. In evaluating a measurement system's variation, the most an adequate technique, once an instrument is calibrated, is Gage repeatability and reproducibility Gage R&R (Hoffa and Laux, 2007). The primary purpose of a Gage study is to determine how much variation in the data is due to the measurement system, and whether the measurement system is accomplished by assessing the process performance. The first type of Gage R&R is crossed Gage R&R, which is developed to analyse data from typical measurement system studies. It adopts the most common approach to the appropriate measurement of data with an ANOVA model and evaluates different sources of variation in the measurement system using the variance components in the model. The second type of Gage R&R is the nested Gage R&R, which is developed to measure the system analysis when all operators in the system measure different parts.

2. Mathematical model and measurement guality of nested Gage R&R, random effect model

The nested design is the first option for destructive testing since each operator measures unique parts. If a part can be measured multiple times by different operators, then it is necessary to use the crossed design. In this study, we focus only on the nested design. For the nested experiment, as well as the part is nested within each operator, it is impossible to assess the operator and part interaction. The data of nested Gage R&R is represented as shown in Table 1.

Table 1. The experimental format for nested Gage R&R data

Operator $\boldsymbol{i}$	Parts $\boldsymbol{j}_{(\boldsymbol{i})}$	Replication	$\bar{\boldsymbol{y}}$	$\boldsymbol{S^2}$
	1	y_{111} y_{112} $\cdots\cdots y_{11n}$	$\bar{y}_{11}$	$s^2{}_{11}$
	. .		.	.
1	. .		.	.
	. .		.	.
	p	y_{1p1} y_{1p2} $\cdots\cdots y_{1pn}$	$\bar{y}_{1p}$	$S^2{}_{1p}$
	1	y_{211} y_{212} $\cdots\cdots y_{21n}$	$\bar{y}_{21}$	$s^2{}_{21}$
	2 .		.	.
2	. .		.	.
	. .		.	.
	p	y_{2p1} y_{2p2} $\cdots\cdots y_{2pn}$	$\bar{y}_{2p}$	$S^2{}_{2p}$
.				
.				
.				
.				
.				
	1	y_{o11} y_{o12} $\cdots\cdots y_{o1n}$	$\bar{y}_{o1}$	$s^2{}_{o1}$
	2 .		.	.
o	. .		.	.
	. .		.	.
	p	y_{op1} y_{op2} $\cdots\cdots y_{opn}$	$\bar{y}_{op}$	$S^2{}_{op}$

The analysis of variance, random effect model of nested Gage R&R, is represented in Equation 1.

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,3 \ldots\ldots\ldots . o \\ j = 1,2,3, \ldots\ldots\ldots . p \\ k = 1,2,3, \ldots\ldots\ldots n \end{cases} \tag{1}$$

where
μ is the overall mean
τ_i is the effect for the i_{th}operator, $\tau_i \sim^{iid} N(0, \sigma^2{}_\tau)$

$\beta_{j(i)}$ is the effect of the j_{th} part nested within the i_{th}operator, $\beta_{j(i)} \sim^{iid} N(0, \sigma^2{}_{\beta})$
ε_{ijk} is random error where $\varepsilon_{ijk} \sim^{iid} N(0, \sigma^2)$

The total variation and the total degree of freedom of the nested design random effect model can be partitioned into three components as follows:

$$ss_{Total} = ss_{operator} + ss_{part(operator)} + ss_{error}$$

$$\sum\sum\sum\left(y_{ijk} - \bar{y}_{...}\right)^2 = \sum\sum\sum(\bar{y}_{i..} - \bar{y}_{...})^2 + \sum\sum\sum\left(\bar{y}_{ij.} - \bar{y}_{i..}\right)^2 + \sum\sum\sum(y_{ijk} - \bar{y}_{ij.})^2$$

$$\sum\sum\sum\left(y_{ijk} - \bar{y}_{...}\right)^2 = pn\sum_{i}(\bar{y}_{i..} - \bar{y}_{...})^2 + n\sum_{I}\sum_{j}(\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum\sum\sum(y_{ijk} - \bar{y}_{ij.})^2$$

Partitioning of Degree of Freedom:

$$opn - 1 = (o-1) + o(p-1) + op(n-1)$$

$$opn - 1 = o - 1 + op - p + opn - op$$

where:

$$y_{i..} = \sum_j \sum_k y_{ijk}; \quad \bar{y}_{i..} = \frac{y_{i..}}{np}$$

$$y_{ij.} = \sum_k y_{ijk}; \quad \bar{y}_{ij.} = \frac{y_{ij.}}{n}$$

$$y_{...} = \sum\sum\sum y_{ijk}; \quad \bar{y}_{...} = \frac{y_{...}}{opn}$$

The Expected Mean Squares of the nested Gage R&R random effect model are presented in Table 2.

Table 2. Expected Mean Squares

Mean Squares	**Degree Of Freedom**	**Expected Mean Squares**
MS_o	$o-1$	$\sigma^2 + np\sigma^2{}_{\tau} + n\sigma^2{}_{\beta(i)}$
$MS_{P(O)}$	$o(p-1)$	$\sigma^2 + n\sigma^2{}_{\beta(i)}$
MSE	$op(n-1)$	σ^2

The nested experiment calculations for a total sum of squares (SS_{Total}), the sum of squares of the operator (SS_O), the sum of squares of the part nested within operator ($SS_{part(operator)}$) and the sum of a square of error (SSE) are shown in Table 3.

Table 3. Analysis of variance table for the random effect model for Gage R&R study nested design

Source	S.S	D.F	MS	F
$operator$	$ss_{operator} = pn\sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$o-1$	$MSo = \frac{ss_{operator}}{o-1}$	$F_O = \frac{MSo}{MS_{P(o)}}$
$Parts_{(Operator)}$	$ss_{part(o)} = n\sum_i\sum_j (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$o(p-1)$	$MS_{p(o)} = \frac{ss_{part}}{o(p-1)}$	$F_{P(O)} = \frac{MS_{p(o)}}{MSE}$
$Error$	$SSE = \Sigma\Sigma\Sigma(y_{ijk} - \bar{y}_{ij.})^2$	$op(n-1)$	$MSE = \frac{ss_{error}}{op(n-1)}$	
$\boldsymbol{Total}$	$ss_{Total} = \Sigma\Sigma\Sigma(y_{ijk} - \bar{y}_{...})^2$	$opn-1$		

As we mentioned previously, nested Gage R&R is a measurement system analysis whereby the variation in the system is due to repeatability and reproducibility. Repeatability is a variation from a measurement instrument and, on the other hand, reproducibility is a variation from the operators using the instrument (Erdmann et al., 2009).

In this design, the interest is to test for the operator effect and the part (operator) effect. The test on the operator effect is expected to be non-significant that implies operators have no difficulty in making consistent measurements.

The part (operator) is anticipated to be significant, which indicates the ability of the Gage/instrument to distinguish between units of measurement.

The following hypothesis and test statistics in Equation 2 are used to test the operator's effect:

$H_0: \sigma^2{}_O = 0$ (No significant difference between the operator's effects)
$H_1: \sigma^2{}_O > 0$ (Significant difference between the operator's effects)

$$\text{Test statistic: } F = \frac{MS_O}{MS_{P(o)}} \tag{2}$$

The estimated variance of the operator's effect can be formulated as follows:

$E(MS_O) - E\big(MSP(o)\big) = (\sigma^2 + np\sigma^2{}_\tau + n\sigma^2{}_{\beta(i)})$- $(\sigma^2 + n\sigma^2{}_{\beta(i)})$

$MS_O - MS_{P(O)} = \hat{\sigma}^2 + np\hat{\sigma}^2{}_\tau + n\hat{\sigma}^2{}_{\beta(i)} - \hat{\sigma}^2 - n\hat{\sigma}^2{}_{\beta(i)}$

Therefore:

$$\hat{\sigma}^2{}_\tau = \frac{MS_O - MS_{P(O)}}{pn} \tag{3}$$

$\hat{\sigma}^2{}_\tau$ is the estimated variance for the operator denoted as $\hat{\sigma}^2{}_{operator}$.

The following Equation (4) is used to test the part's effect

$H_0: \sigma^2{}_{\beta(i)} = 0$ (No significant difference between the parts)

$H_1: \sigma^2{}_{\beta(i)} > 0$ (Significant difference between the parts)

$$\text{Test statistic: } F = \frac{MS_{p(O)}}{MSE} \tag{4}$$

With a similar approach as in Equation (3) and Equation (4), the estimated variance of the parts' effect can be formulated as follows:

$$E\left(MS_{P(O)}\right) - E(MSE) = \sigma^2 + n\sigma^2{}_{\beta(i)} - \sigma^2$$

$$MS_{P(O)} - MSE = \hat{\sigma}^2 + n\hat{\sigma}^2{}_{\beta(i)} - \hat{\sigma}^2 = n\hat{\sigma}^2{}_{\beta(i)}$$

$$\hat{\sigma}^2{}_{\beta(i)} = \frac{MS_{P(O)} - MSE}{n} \tag{5}$$

$\hat{\sigma}^2{}_{\beta(i)}$ is the estimated variance for the part (operator) denoted as $\hat{\sigma}^2{}_{\text{parts(operator)}}$.

$$\hat{\sigma}_{part(operator)} = \sqrt{\hat{\sigma}^2{}_{parts(operator)}} \tag{6}$$

$$\hat{\sigma}^2{}_e = MSE \tag{7}$$

$$\hat{\sigma}^2{}_{repeatability} = MSE = \hat{\sigma}^2;\ \hat{\sigma} = \sqrt{MSE}$$
$$\hat{\sigma}^2{}_{reproducibility} = \hat{\sigma}^2{}_O$$

If the value of the variance components is less than 0, treat them as equal to 0 because variance cannot be negative.

The estimated variance for Gage R&R is given by:

$$\hat{\sigma}^2{}_{Gage\ R\&R} = \hat{\sigma}^2{}_{repeatability} + \hat{\sigma}^2{}_{reproducibility} = \hat{\sigma}^2 + \hat{\sigma}^2{}_o \tag{8}$$

$$\text{The standard deviation of Gage R\&R} = \sqrt{\hat{\sigma}^2{}_{Gage\ R\&R}} \tag{9}$$

Estimated variance of Total variation= $\hat{\sigma}^2{}_{Gage\ R\&R} + \hat{\sigma}^2{}_{part(operator)}$

$$\text{The estimated standard deviation of Total variation} = \sqrt{\hat{\sigma}^2{}_{Gage\ R\&R} + \hat{\sigma}^2{}_{parts(operat}} \tag{10}$$

In the study of Gage R&R design, the Gage capability can be measured by using the precision-to-tolerance ratio (or P/T ratio), as follows:

$\frac{P}{T} = \frac{6\hat{\sigma}_{Gage\ R\&R}}{USL-LSL}$ where $\hat{\sigma}_{GageR\&R}$ is the standard deviation of Gage R&R as stated in Equation (9).

USL and LSL are the upper and lower specification limits of the product under study (given in each nested Gage R&R data). If the P/T ratio is 0.1 or less, this indicates acceptable Gage capability (Headquarters, 2015). But there are clear dangers in relying too much on the P/T ratio, in some nested Gage R&R data. For example, the ratio may be made randomly small by increasing the width of the specification tolerance (Stevens, 2013). As such other measures are employed such as using the percentage contribution of the variance component of the total variation of Gage R&R and also the percentage contribution of the variation component of part (operator).

$$\text{Percentage contribution of Total variation of Gage R\&R} = \frac{\text{variation of Gage R\&R}}{Total\ variation} \quad (11)$$

The percentage contribution of variation of Gage R&R measures the contribution of the nested Gage R&R in the total variation. The small value of per cent contribution of Gage R&R means adequate Gage. % contribution $<$ 30% indicates the measurement system is capable.

Another important measure is by using percentage contribution of variance component of part (operator) as follows:

$$\text{Percentage contribution of part (operator)} = \frac{\text{the variance of part (operator)}}{Total\ variation} \quad (12)$$

A high percentage of contribution indicates good measurement, which implies that the measurement system can distinguish between parts. Most of the total variation in the measurement is due to differences between parts, which is desirable.

An equivalent measure of Gage capability is by using the percentage contribution of standard deviation namely:

$$\%\text{contribution of total sd of Gage R\&R} = \frac{standard\ deviation\ of\ Gage\ R\&R}{standard\ deviation\ of\ Total} \quad (13)$$

$$\%\text{contribution of total 6 sd of Gage R\&R} = \frac{6\ standard\ deviation\ of\ Gage\ R\&R}{6\ standarad\ deviation\ of\ Total} \quad (14)$$

$$\%\text{contribution of sd of part(operator)} = \frac{standard\ deviation\ of\ part(operator)}{\text{standard deviation of Total}} \quad (15)$$

3. Methodology

This section presents the methodology of this study. The study proposes a method to identify outliers in nested Gage R&R, namely, a method based on a highly efficient estimator which has a high breakdown point. Moreover, to reduce the negative effect of the outliers, a robust estimation method has been presented to obtain a reliable measurement with low variation after detecting the outliers. Then, two numerical examples and the analysis of the data are presented. Moreover, the simulation study is illustrated.

3.1. Data analysis: how outliers affect the analysis of Gage R&R

In this section, a numerical nested example is presented to show the effect of outliers on the nested Gage R&R measurements.

This data is taken from (Excel, 2013), which described an industrial application whereby heat treating of parts is inducted to perform a Gage R&R analysis on the hardness tester. For the reason of measuring the hardness, a piece of the product is cut, prepared and tested. That piece was altered, so it cannot be retested. It is assured that the parts within operators are homogeneous. For this process, three operators are included in the Gage R&R study, that is operator A, B and C. Each operator is needed to test two parts. But there are not always enough measurements for each operator to test parts from each operator. Based on that, a nested design has been used. Three operators are used and five parts from each operator and two measurements from each part have been taken. The total number of measurements is 30. Table 4 shows the collected data of industrial section.

To see the effect of outliers on the variability's measurements, we purposely contaminate the data with a certain number of outliers. The outliers are created by replacing one observation of each operator by $(maximum\ value\ of\ observation + 10\ sd\ in\ each\ operator)$. The outlier is represented in bold, in Table 4.

Table 4. Nested Gage R&R numerical example of three different operators from the industrial section

Operators	Parts									
A	1		2		3		4		5	
	1	2	1	2	1	2	1	2	1	2
	33.4	33.2	32.4	31.7	34.4	34.5	33.9	34.5	34.5	**35.7 (47.4)**
B	6		7		8		9		10	
	1	2	1	2	1	2	1	2	1	2
	32.5	32.1	32.1	32.3	**35.1 (48.1)**	34.7	32.4	33.1	34.8	34.9
C	11		12		13		14		15	
	1	2	1	2	1	2	1	2	1	2
	32.6	32.7	32.3	32.1	**34.9 (47.3)**	34.7	33.0	33.2	31.6	30.9

Nested ANOVA or nested Gage R&R table is represented to show the significance of the parts and the operators and the table of variance components to measure the Gage variation, part-to-part variation and the total variation as shown in Table 5-A. The components of variance and standard deviations contribution are shown in Table 5-B and Table 5-C, respectively. Five useful graphs for the interpretation of the experimental results are displayed in Figure 3 and Figure 4.

Table 5. Gage R&R (nested) for results without and with outliers

A. Nested ANOVA

Source	df	SS		MS		F		p	
		Without outlier	With outlier	Without outlier	With outlier	Without outlier	With outlier	Without outlier	With outlier
Operator	2	5.256	4.741	2.628	2.371	0.793	0.083	0.475	0.921
Part (operator)	12	37.766	342.178	3.133	28.514	26.024	1.687	$10e^{-5}$	0.168
Repeatability	15	1.911	253.455	0.127	16.897				
Total	29	46.932	600.374						

B. Components of variance analysis

Gage R&R Source	Var Comp		%Contribution (of Var Comp)	
	Without outlier	With outlier	Without outlier	With outlier
Total variation of Gage R&R	0.127	16.897	7.41	74.42
Repeatability	0.127	16.897	7.41	74.42
Reproducibility	$10e^{-5}$	$10e^{-5}$	$10e^{-5}$	$10e^{-5}$
$part\ to\ part\ or\ part_{(operator)}$	1.593	5.808	92.61	25.58
Total variation	1.721	22.705	100.00	100.00

C. Components of standard deviation and 6 × standard deviation analysis

Gage R&R Source	StdDev (SD)		Study Var (6 × SD)		%Study Var (%SV)	
	Without outlier	With outlier	Without outlier	With outlier	Without outlier	With outlier
Total s.d of Gage R&R	0.356	4.111	2.141	24.663	27.21	86.27
Repeatability	0.356	4.111	2.141	24.663	27.21	86.27
Reproducibility	$10e^{-5}$	$10e^{-5}$	$10e^{-5}$	$10e^{-5}$	$10e^{-5}$	$10e^{-5}$
$part\ to\ part\ or\ part_{(operator)}$	1.262	2.411	7.573	14.461	96.23	50.58
Total variation	1.311	4.765	7.871	28.591	100.00	100.00

*Specification tolerance (upper specification limit-lower specification limit=8). From the example and process standard deviation is 2.5.

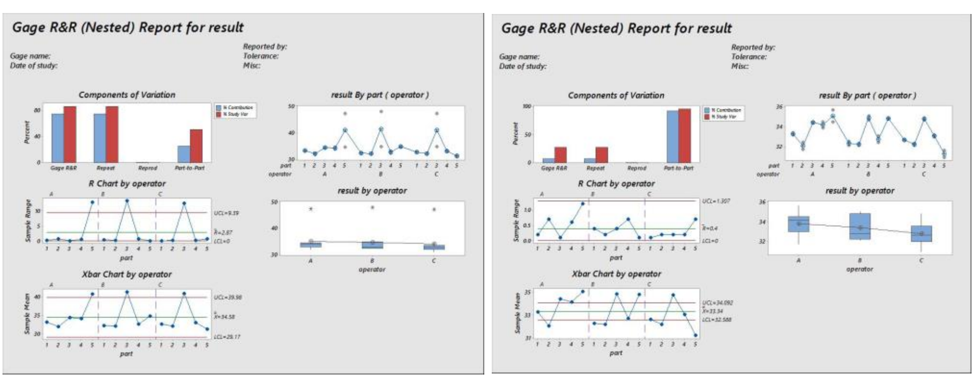

Figure 1. Nested Gage R&R: components of variation, result by part (operator), R chart by the operator, X bar chart by the operator, result by part (operator), and result by the operator without outlier (left) and with outliers (right).

It can be observed from Table 5-A, when there are no outliers there is a significant difference between parts when nesting within operators ($p - value < 0.05$). This results indicate that the Gage is capable of distinguishing between different units. The test on operator suggests that the operator has no difficulty of making consistent measurements ($p - value > 0.05$). These two conditions are desired. Gage R&R studies quantify this by determining the % Gage R&R value. Based on these results, the hardness tester (Gage R&R) is responsible for about 7% of the total variation. This test method appears to be very reliable because the % contribution of Gage R&R variation is less than 30%.

The percentage of contribution for the difference between parts, when nesting within operators ($part - to - part = 96.23$) as shown in Table 5-B, is high, which is close to 100%. The higher percentage contribution for the parts indicates good performance of system Gage R&R. These results can be seen from graphics. The components of the variation graph are placed in the upper left corner in Figure 3 and that means reliable data. Also in Table 5-B, the reproducibility is 0 because all the variations are due to the Gage variation and part-to-part (the part when nesting within operator variation) not due to the interaction between operators and parts.

Most of the variations are due to part-to-part (parts nested within operators) variation, with a low percentage of variation due to errors in the measurement system of Gage R&R at the $\bar{x}$ chart-located in the lower left corner in Figure 1. Most of the points in the $\bar{x}$ chart are outside the control limits when the variation is mostly due to part-to-part variation, 27 points of 30 outside the control limits, i.e. about 90%, the Gage is capable (should be more than 75% is outside the control limits) (Headquarters, 2015).

$$\text{The percentage of contribution for total sd of Gage R\&R } = \frac{(6 \times \text{SD})}{Total\ variation}$$

$= 2.141/7.871 = 27.2\%$. This means that the “spread” of the Gage R&R takes up to 27.2% of the total spread. This result implies that the Gage is acceptable (the Gage variation and spread should be less than 30%) (Headquarters, 2015).

Now, let us focus on the results with outliers of Table 5-A (with outliers). It can be seen that the part (operator) effect is not significant ($p - value > 0.05$). This indicates that the Gage cannot distinguish between different units, which is not desirable.

From Table 5-B (with outlier’s columns), in particular, the percentage of contribution for the difference between parts ($part - to - part = 25.58$) is much smaller than the percentage of contribution to the variation of the measurement system 92.61 when there is no outlier. It is noticed that the % contribution of variation of hardness tester (Gage R&R) has increased to 74.42% of the total variation. This indicates that the Gage is not capable in the presence of outliers.

Figure 1 shows that the components of the variation graph are placed in the upper left corner. Most of the variations are due to errors in the measurement variation Gage R&R, with a low percentage of variation due to the part-to-part (parts nested within operators) variation.

We have seen the effect of outliers on the measurement variation of the nested Gage R&R data in the numerical example just presented.

3.2. Outlier identification method

We have seen in the previous section that an outlier has a negative effect on the nested Gage R&R analysis. In this situation, it is very crucial to detect an outlier in the nested Gage R&R model. To the best our knowledge no such work has been devoted to identifying an outlier in the nested Gage R&R, random effect model.

Fearn (2001) and Walsh (2016) developed Cochran's C test to decide if a single estimate of variance (or a standard deviation) is significantly larger than a group of variances. Rousseeuw and Mia Hubert (2011) developed a modified Rousseeuw and Mia Hubert method to identify outliers in univariate data. Tukey (2011) also discussed the method of identifying outliers in such type of data. Bagheri and Midi (2011) noted that the traditional approach of identifying outliers in univariate data is by using T statistics, $T = \frac{x-\bar{x}}{s}$.

3.2.1. Rousseeuw and Mia Hubert method

Rousseeuw and Mia Hubert (2011) proposed a method to detect outliers in a univariate data as presented in the following steps:

- Compute the median of all observations in a data set
- Calculate $MAD = 1.483\, median\ \left|observation - median\ \right|$
- Calculate $z_i = \frac{observation - median}{MAD}$
- Any value of $|z_i| > 2.5$ is considered as an outlier

This method is denoted as Z_{RM}

3.2.2. Tukey method

Hubert and Vandervieren (2008) defined the Tukey method in the following steps:

- Compute Interquartile Range, $\text{IQR} = Q_3 - Q_1$

where:

$Q_3 = X[3n/4]$
$Q_1 = X[n/4]$
$x\ =\ variable\ of\ observation$
$n =\ sample\ size$

An observation is detected as an outlier if it lies outside the following interval

$$[Q_1 - 1.5\ \text{IQR},\ Q_3 + 1.5\ \text{IQR}]$$

This method is denoted as Int_T.

3.3. Proposed method of outlier detection in nested Gage R&R data

As already mentioned, no specific test is developed to identify outliers in nested Gage R&R. The Rousseeuw and Hubert method and the Tukey method is designed for univariate data. Hence, it can be adopted in the formulation of the detection of outliers in nested Gage R&R, with slight modification.

Instead of using the observed value of x as noted in the Rousseeuw and Hubert method and the Tukey method, the residuals can be computed in this regard. It can be observed that MAD and IQR are used as the scale estimator in the Rousseeuw and Hubert and the Tukey method, respectively.

Even though this estimator is resistant to outliers, its weakness is that it is not reliable under normality assumption (Lee et al., 2007). Another shortcoming of this method is that the use of the median is not very reliable because it has low efficiency under normal errors (Mazlina and Habshah, 2015).

As such, we propose to formulate a new test measure, which is based on highly efficient mm estimator, which has a high breakdown point.

The proposed method is summarized as follows:

Step 1: Perform Analysis of the variance method to nested Gage R&R, random effect model.

Step 2: Compute the fitted value as follows:

Referring to Equation 3.1; $E\left(y_{ijk}\right) = \mu$ because:

$\tau_i \sim^{iid} N(0, \sigma^2{}_{\tau}), \beta_{j(i)} \sim^{iid} N\left(0, \sigma^2{}_{\beta(\tau)}\right)$ and $\varepsilon_{ijk} \sim^{iid} N(0, \sigma^2)$

Hence, the fitted value is written as $\hat{y}_{ijk} = \hat{\mu} = \bar{y}_{...}$ where $\bar{y}_{...} = \frac{\sum\sum\sum y_{ijk}}{opn}$

Step 3: Compute the residual (e_{ijk}) of each observation as follows:

$$e_{ijk} = y_{ijk} - \bar{y}_{...}$$

Let e_{11}, e_{21},……, e_{opn} be opn residuals represented by $r_1, r_2, \ldots\ldots, r_{opn}$

The location-scale model can be written as follows:

$$r_i = \mu + \sigma\varepsilon_i$$

The mm location and scale of r_i is computed in three steps:

Step i: Use robust S estimator to obtain the initial consistent estimator μ_0 and scale σ_0.

Step ii: Compute m estimate of the scale of the residuals from the initial estimates of the location.

Step iii: Using m estimation method, compute the location and scale of the mm estimates

denoted as $\hat{\mu}_{mm}$ and $\hat{S}_{mm}$.

Step 4: Compute the mm location and scale estimates of the residuals. The mm location and
the scale estimates are chosen because according to [18] they has high breakdown points and high efficiency under normal errors.

Step 5: Compute $T_{mm} = \frac{e_{ijk} - \hat{\mu}_{mm}}{\hat{s}_{mm}}$.

Step 6: Any value of $|T_{mm}| > 2.5$ is declared as an outlier.

3.3.1. Simulation study

In order to assess the performance of our proposed method a simulation study is firstly carried out by considering three operators, five parts and two replicates. For each operator $i = 1,2,3; j = 1,2,3,4,5; k = 1,2$.

y_{i1j} is generated from $N(30,0.1)$

y_{i2j} is generated from $N(32,0.1)$

y_{i3j} is generated from $N(34,0.1)$

y_{i4j} is generated from $N(36,0.1)$

y_{i5j} is generated from $N(38,0.1)$

The above process is repeated for various number of operators, parts and samples; such as (3 operators, 5 parts and 2 samples), (3 operators, 10 parts and 2 samples), (4 operators, 5 parts and 2 samples), (5 operators, 6 parts and 2 samples), (5 operators, 8 parts and 2 samples), (6 operators, 10 parts and 2 samples), (10 operators, 12 parts and 2 samples). For each design layout, the data is then contaminated by replacing good observation with a certain number of outliers. The outliers are created by taking the maximum value of each data set +3, 5 and10 standard deviation. The same process is repeated for samples equal to 4. Since in practice it is expensive to collect data, it is not recommended to have more than 5 sample sizes. The proposed method is evaluated based on the number of correct detection of outliers. The number of iterations for each design layout is equal to 1000. The results are presented in Tables (6, 7, 8, 9, 10 and 11).

It can be observed from all tables that the T_{mm} is very successful in detecting outliers in the data set compared to the other two methods. The Rousseeuw and Tukey methods become very poor as the number of outliers increases. Both methods suffer from the masking effect. It is very interesting to see that our proposed method is capable of identifying the correct outliers with no masking effect.

Table 6. Percentage of correct detection of an outlier for 3 standard deviations, $n = 2$, the Rousseeuw method and the Tukey method

Operator (part) and two samples	Number of outliers	Proposed Method. Number of correct detection	Rousseeuw method in percentage	Tukey method in percentage
3(5)	0	100	55.6	73.8
	1	100	53.4	43.8
	2	100	44.6	41.6
	3	100	36.3	19.4
3(10)	0	100	37.7	72.2
	1	100	33.5	52.8
	2	100	32.1	38.3
	3	100	30.9	24.3
	4	100	29.8	12.4
	5	100	19.3	7.8
	6	100	0	0
4(5)	0	100	53.5	72.7
	1	100	45.8	47.4
	2	100	51.2	29.3
	3	100	35.4	19.7
	4	100	23.1	11.9
5(6)	0	100	39.3	74.9
	1	100	36.4	56.6
	2	100	32.2	36.7
	3	100	30.9	23.3
	4	100	30.1	10.3
	5	100	18.5	7.2
	6	100	0	0
5(8)	0	100	48.8	71.4
	1	100	42.6	59.6
	2	100	41.1	45.3
	3	100	40.3	30.7
	4	100	39	15.5
	5	100	26.8	8
	6	100	1.7	0.7
	7	100	0	0
	8	100	0	0
6(10)	0	100	35.4	65.3
	1	100	31.8	61.7
	2	100	29	50
	3	100	28.5	37.1
	4	100	21.9	19.2
	5	100	24.1	13
	6	100	9.5	2.4
	12	100	0	0
10(12)	0	100	43.7	64
	1	100	18.5	45.5
	2	100	10.2	42.9
	3	100	8.3	40.6
	4	100	7.9	38.5
	5	100	7.4	33.5
	6	100	6.6	14.8
	7	100	1.1	0.4
	8	100	0	0
	9	100	0	0
	10	100	0	0
	11	100	0	0
	12	100	0	0
	24	100	0	0

Table 7. Percentage of correct detection of an outlier for 3 standard deviations, $n = 4$, the Rousseeuw method and the Tukey method

Operator (part) and two samples	Number of outliers	Proposed method Number of correct detection	Rousseeuw method in percentage	Tukey method in percentage
3(5)	0	100	49.3	74.9
	1	100	36.3	29.9
	2	100	27.8	16.6
	3	100	13.6	11.5
	4	100	5.5	6.7
	5	100	3.7	3.5
	6	100	0	0
3(10)	0	100	49	73.1
	1	100	34.7	37.2
	2	100	35.1	19.2
	3	100	24.3	9.8
	4	100	11.2	5.6
	5	100	5.1	3.3
	6	100	0.2	0.7
	12	100	0	0
4(5)	0	100	48.8	71.4
	1	100	39.3	29.4
	2	100	31.4	15.9
	3	100	15.6	9.9
	4	100	7.3	7.6
	5	100	2.7	4.9
	6	100	0.3	0.4
	7	100	0	0
	8	100	0	0
5(8)	0	100	47.7	70.7
	1	100	42.6	39.7
	2	100	40.2	21.6
	3	100	26.1	10.1
	4	100	12.6	5
	5	100	6.7	2.3
	6	100	3.2	1.9
	12	100	0	0
6(10)	0	100	43.7	64
	1	100	33.1	42.6
	2	100	29.1	28.8
	3	100	25.9	12.9
	4	100	21.1	6.5
	5	100	12.6	2.7
	6	100	11.8	1.9
	12	100	0	0

Table 8. Percentage of correct detection of an outlier for 5 standard deviations, $\boldsymbol{n = 2}$, the Rousseeuw method and the Tukey method

Operator (part) and two samples	Number of outliers	Proposed method. Number of correct detection	Rousseeuw method in percentage	Tukey method in percentage
3(5)	0	100	58.6	73.8
	1	100	56.9	65.8
	2	100	55.8	53.3
	3	100	46.4	38.4
3(10)	0	100	58.4	68.5
	1	100	58.1	66.5
	2	100	57.9	66.5
	3	100	54.7	64.2
	4	100	50.8	61.9
	5	100	27.8	14.9
	6	100	0	0
4(5)	0	100	53.5	72.7
	1	100	49.8	69.3
	2	100	52	59.1
	3	100	43.5	41.4
	4	100	34.3	22.7
5(6)	0	100	39.3	74.9
	1	100	35.3	74.2
	2	100	30.6	65.6
	3	100	28.9	50.6
	4	100	27	29.1
	5	100	26.3	12.4
	6	100	0	0
5(8)	0	100	48.8	71.4
	1	100	33.1	70.8
	2	100	31.9	69.6
	3	100	30.6	58.8
	4	100	29.7	38.3
	5	100	26.9	32.8
	6	100	23.1	29.6
	7	100	19.5	21.7
	8	100	16.3	18.5
6(10)	0	100	35.4	65.3
	1	100	30.6	61.3
	2	100	20.1	59.6
	3	100	17.7	58.7
	4	100	15.2	40.2
	5	100	13.7	33.2
	6	100	9.8	26.9
	12	100	6.4	15.6
10(12)	0	100	43.7	64
	1	100	28.6	55.4
	2	100	24.9	52.7
	3	100	22.1	51
	4	100	19.3	48.5
	5	100	18.1	43.9
	6	100	15.4	40.3
	7	100	13.9	29.5
	8	100	10.6	21.4
	9	100	6.3	18.2
	10	100	4.5	13.9
	11	100	2.8	5.4
	12	100	1.3	4.6
	24	100	0	0

Table 9. Percentage of correct detection of an outlier for 5 standard deviations, $\boldsymbol{n = 4}$, the Rousseeuw method and the Tukey method

Operator(part) and two samples	Number of outliers	Proposed method. Number of correct detection	Rousseeuw method in percentage	Tukey method in percentage
3(5)	0	100	64.1	72.3
	1	100	32.7	33.5
	2	100	18.7	16.6
3(10)	0	100	63.5	72
	1	100	43.7	46.2
	2	100	30.6	29.3
	3	100	22.1	20.3
	4	100	19.8	15.4
	5	100	16.1	12.9
	6	100	10.5	8.7
	12	100	0	0
4(5)	0	100	62.8	72.3
	1	100	44.3	46.1
	2	100	26.6	25.8
	3	100	22.9	21.3
	4	100	18.9	20.8
	5	100	16.2	18.9
	6	100	10.7	13.2
	7	100	0	0
	8	100	0	0
5(8)	0	100	62.2	70.6
	1	100	40.4	47.9
	2	100	33.4	27.3
	3	100	22.7	19.8
	4	100	14.5	12.3
	5	100	11.8	9.6
	6	100	8.7	5.4
	12	100	0	0
6(10)	0	100	42.7	59.5
	1	100	35.8	45.5
	2	100	31.4	30.9
	3	100	24.7	21.7
	4	100	20.7	14.1
	5	100	16.2	9.8
	6	100	12.2	6.4
	7	100	6.3	4.4
	8	100	3.4	4.1
	9	100	1.7	2.9
	10	100	1.2	0.9
	11	100	0.4	0.7
	12	100	0.3	0.6
	24	100	0	0

Table 10. Percentage of correct detection of an outlier for 10 standard deviations, $n = 2$, the Rousseeuw method and the Tukey method

Operator (part) and two samples	Number of outliers	Proposed method Number of correct detection	Rousseeuw method in percentage	Tukey method in percentage
3(5)	0	100	58.6	73.8
	1	100	58	71.2
	2	100	57.4	62.1
	3	100	54.5	52.4
3(10)	0	100	55.4	72.6
	1	100	54.5	71
	2	100	54.3	64.9
	3	100	52.4	60.6
	4	100	44.9	58.5
	5	100	36.3	51.2
	6	100	28.5	49.8
4(5)	0	100	53.5	72.7
	1	100	52.8	71.8
	2	100	51.9	68.4
	3	100	50.6	55.4
	4	100	42.5	35.7
5(6)	0	100	39.3	74.9
	1	100	37.3	74.7
	2	100	34.3	72.2
	3	100	32.3	61.9
	4	100	30.4	40.9
	5	100	29.5	33.1
	6	100	0	0
5(8)	0	100	48.8	71.4
	1	100	24.9	76.2
	2	100	22.5	74.4
	3	100	20.4	68.1
	4	100	19.3	49.5
	5	100	0	0
	6	100	0	0
	7	100	0	0
	8	100	0	0
6(10)	0	100	35.4	65.3
	1	100	34.6	64.2
	2	100	32.8	61.9
	3	100	29.5	59.7
	4	100	24.6	56.7
	5	100	22.1	52.1
	6	100	19.8	44.9
	12	100	0	0
10(12)	0	100	43.7	64
	0	100	45.3	66.9
	1	100	42.9	63.7
	2	100	37.3	59.8
	3	100	28.5	52.4
	4	100	21.8	49.6
	5	100	19.6	38.5
	6	100	16.5	32.4
	7	100	12.5	30.9
	8	100	9.8	28.4
	9	100	7.2	25.1
	10	100	5.4	22.9
	11	100	3.7	21.4
	12	100	2.1	17.6
	24	100	0	0

Table 11. Percentage of correct detection of an outlier for 10 standard deviations, $\boldsymbol{n = 4}$, the Rousseeuw method and the Tukey method

Operator (part) and two samples	Number of outliers	Proposed method Number of correct detection	Rousseeuw method in percentage	Tukey method in percentage
3(5)	0	100	39.3	74.9
	1	100	38.9	66.6
	2	100	36.8	54.5
	3	100	33.7	40.2
	4	100	29.5	38.4
	5	100	14.6	30.6
	6	100	10.9	25.4
3(10)	0	100	40.5	72.8
	1	100	56.1	54.5
	2	100	50.9	40.2
	3	100	44.5	38.6
	4	100	25.1	35.4
	5	100	23.9	32.4
	6	100	21.2	29.6
	12	100	0	0
4(5)	0	100	48.8	71.4
	1	100	52.2	50.5
	2	100	47.4	33.6
	3	100	32.1	33.9
	4	100	20.6	25.3
	5	100	2.5	6.8
	6	100	1.3	4.1
	7	100	0	0
	8	100	0	0
5(8)	0	100	47.7	70.7
	1	100	46.1	54.3
	2	100	34.3	36.9
	3	100	40.2	43.9
	4	100	28.5	33.7
	5	100	21.9	30.4
	6	100	19.1	28.5
	7	100	12.9	21.4
	8	100	0	0
6(10)	0	100	43.7	64
	1	100	41.6	54.6
	2	100	38.4	45.3
	3	100	36.8	42.1
	4	100	35.2	38.8
	5	100	32.1	36.9
	6	100	29.4	34.6
	7	100	27.5	31.2
	8	100	24.3	29.5
	9	100	21.6	25.9
	10	100	18.7	23.1
	11	100	13.6	21.8
	12	100	10.2	17.3
	24	100	0	0

3.3.2. Proposed method on numerical example

To show the superiority of the proposed method in outlier detection in nested Gage R&R study, a numerical example is presented. This data has been taken from (Erdmann et al., 2009), described in the health section. This features of the quality improvement were to take measurements for a body temperature of patients. The measurement of the temperature has been taken using an ear thermometer. The normal body temperature for any individual range from 35 °C, which is a lower specification limit (LSL), to 40 °C, which is an upper specification limit (USL). The quality of the temperature measurement is assessed through a Gage R&R study. The nurses handling the ear thermometer may cause some variation. The other group of variation for the experiment involved different healthy persons. A single ear thermometer is used by all the nurses. Each patient is measured in the right and left ear. The experiment has been assumed to involve 3 nurses (operators) and each nurse measures 10 different healthy persons, four times. Table 12 shows the collected data of the health section.

To see the effect of outliers on the variability's measurements, we purposely contaminate the data with a certain number of outliers. The outliers are created by replacing one observation of each operator with $(maximum\ value\ of\ observation + 10\ sd\ in\ each\ operator)$. The outlier is represented in bold, in Table 12.

Table 12. Data of the nested Gage R&R experiment showing 3 different operators, 10 parts, and 4 samples in each part

Patients	Operator 1 Jolan				Operator 2 Mariska				Operator 3 Paula			
	1		2		1		2		1		2	
	r	1	r	1	r	1	r	1	r	1	r	1
1	37.3	37.5	37.3	37.5	37.5	37.7	37.3	37.6	37.5	37.6	37.4	37.5
2	37	37.3	36.7	36.8	37.5	37.3	37.4	37.2	37.4	37.4	37.3	37.1
3	36.4	37	37.3	37	37.5	37.3	37.4	37.1	37.6	37.4	37.2	37
4	37.6	37.5	37.6	37.4	37.5	37.5	37.5	37.7	37.7	37.6	37.6	37.5
5	36.7	37.6	**37.8 (41.8**)	37.5	37.9	37.5	37.6	37.6	37.9	37.6	**37.9 (41.2)**	37.8
6	37.5	37.7	37.6	37.3	**38.4 (41.6)**	38	37.8	37.8	37.6	37.9	37.8	37.8
7	37	36.9	37.1	37.3	37.1	37.3	37.4	37.5	37.2	37.4	37.1	37.2
8	37.7	37.4	37.6	37.4	37.6	37.5	37.5	37.1	37.5	37.4	37.2	36.9
9	36.4	36.5	37.6	36.1	37.1	36.9	36.7	36.8	37	36.4	36.9	36.8
10	37.2	37.4	37	37.3	37.1	37.2	37.2	37.2	37.1	37.2	37	37.3

The residuals r_i, the z_i of the Rousseeuw and Mia Hubert method, the interval of the Tukey method and our proposed T_{mm} method are presented in [Table 13 A and Table 13 B].

Table 13. A. Residuals r_i, z_i, the interval of Tukey and T_{mm}

No	Residuals r_i	z_i	Tukey interval	T_{mm}
1	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
2	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
3	-0.7634017	-3.3715	(-0.46317,4.3683)	-1.37166387
4	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
5	-0.5395861	-2.3601	(-0.46317,4.3683)	-0.96954743
6	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
7	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
8	0.20646592	1.01146	(-0.46317,4.3683)	0.370840721
9	-0.7634017	-3.3715	(-0.46317,4.3683)	-1.37166387
10	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
11	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
12	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
13	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
14	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
15	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
16	0.20646592	1.01146	(-0.46317,4.3683)	0.370840721
17	-0.3903757	-1.6858	(-0.46317,4.3683)	-0.7014698
18	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
19	-0.6887965	-3.0344	(-0.46317,4.3683)	-1.23762506
20	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
21	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
22	-0.5395861	-2.3601	(-0.46317,4.3683)	-0.96954743
23	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
24	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
25	3.26527912	14.8348	(-0.46317,4.3683)	5.866432121
26	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
27	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
28	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
29	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
30	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
31	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
32	-0.4649809	-2.0229	(-0.46317,4.3683)	-0.83550861
33	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
34	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
35	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
36	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
37	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
38	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
39	-0.9872173	-4.383	(-0.46317,4.3683)	-1.77378031
40	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
41	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
42	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
43	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
44	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
45	0.35567632	1.68577	(-0.46317,4.3683)	0.63891835
46	3.11606872	14.1605	(-0.46317,4.3683)	5.598354492
47	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
48	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
49	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
50	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
51	0.20646592	1.01146	(-0.46317,4.3683)	0.370840721
52	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
53	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
54	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
55	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
56	0.43028152	2.02293	(-0.46317,4.3683)	0.772957164
57	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
58	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
59	-0.3903757	-1.6858	(-0.46317,4.3683)	-0.7014698
60	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335

Table 13. B. Residuals r_i, z_i, the interval of Tukey and T_{mm}

No	Residuals r_i	z_i	Tukey interval	T_{mm}
61	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
62	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
63	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
64	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
65	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
66	0.28107112	1.34862	(-0.46317,4.3683)	0.504879535
67	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
68	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
69	-0.5395861	-2.3601	(-0.46317,4.3683)	-0.96954743
70	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
71	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
72	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
73	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
74	0.20646592	1.01146	(-0.46317,4.3683)	0.370840721
75	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
76	0.28107112	1.34862	(-0.46317,4.3683)	0.504879535
77	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
78	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
79	-0.4649809	-2.0229	(-0.46317,4.3683)	-0.83550861
80	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
81	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
82	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
83	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
84	0.20646592	1.01146	(-0.46317,4.3683)	0.370840721
85	0.35567632	1.68577	(-0.46317,4.3683)	0.63891835
86	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
87	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
88	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
89	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
90	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
91	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
92	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
93	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
94	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
95	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
96	0.35567632	1.68577	(-0.46317,4.3683)	0.63891835
97	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
98	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
99	-0.7634017	-3.3715	(-0.46317,4.3683)	-1.37166387
100	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
101	-0.0173497	3.4E-15	(-0.46317,4.3683)	-0.03127572
102	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454
103	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
104	0.13186072	0.67431	(-0.46317,4.3683)	0.236801906
105	2.81764792	12.8119	(-0.46317,4.3683)	5.062199233
106	0.28107112	1.34862	(-0.46317,4.3683)	0.504879535
107	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
108	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
109	-0.3903757	-1.6858	(-0.46317,4.3683)	-0.7014698
110	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
111	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
112	-0.2411653	-1.0115	(-0.46317,4.3683)	-0.43339217
113	-0.3157705	-1.3486	(-0.46317,4.3683)	-0.56743098
114	0.05725552	0.33715	(-0.46317,4.3683)	0.102763091
115	0.28107112	1.34862	(-0.46317,4.3683)	0.504879535
116	0.28107112	1.34862	(-0.46317,4.3683)	0.504879535
117	-0.1665601	-0.6743	(-0.46317,4.3683)	-0.29935335
118	-0.3903757	-1.6858	(-0.46317,4.3683)	-0.7014698
119	-0.4649809	-2.0229	(-0.46317,4.3683)	-0.83550861
120	-0.0919549	-0.3372	(-0.46317,4.3683)	-0.16531454

It can be observed from Tables [13 A and Table 13 B] that our proposed method can detect the 3 outliers that are purposely placed in the data set. However, the Rousseeuw method detects 8 outliers and the Tukey method detects 15 outliers.

5 Conclusion

Outliers have an adverse effect on the analysis of the nested Gage R&R measurements and give a misleading conclusion. Therefore, they should be detected at the outset before further analysis is carried out. Once the outlier is detected, the management should find out whether this outlier was caused by the parts or operators handling the equipment or it is a true error from random variation. Proper action should be taken if those outliers are due to operators or parts. As such, it is very crucial to have an efficient method of identifying outliers. We propose a new method, T_{mm} in this regard. The simulation study and the numerical example clearly show that our proposed method is able to successfully identify an outlier with no masking effect. Nonetheless, the other two methods are not performing well and suffer from masking effect.

REFERENCES

MACINTOSH, N. B., QUATTRONE, P., (2010). Management accounting and control systems: An organizational and sociological approach: John Wiley Sons.

BOURNE, M., MELNYK, S., FAULL, N., FRANCO-SANTOS, M., KENNERLEY, M., MICHELI, P., GRAY, D., (2007). Towards a definition of a business performance measurement system. International Journal of Operations and Production Management, 27(8), pp. 784–801.

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(12), pp. 14602–14608.

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), p. 342.

AWAD, M., ERDMANN, T. P., SHANSHAL, Y., BARTH, B., (2009). A measurement system analysis approach for hard-to-repeat events. Quality Engineering, 21(3), pp. 300–305.

PERUCHI, R. S., BALESTRASSI, P. P., DE PAIVA, A. P., FERREIRA, J. R., DE SANTANA CARMELOSSI, M., (2013). A new multivariate gage R&R method for correlated characteristics. International Journal of Production Economics, 144(1), pp. 301–315.

GREJDA, R., MARSH, E., VALLANCE, R., (2005). Techniques for calibrating spindles with nanometer error motion. Precision engineering, 29(1), pp. 113–123.

PARKER, D. H., ANDERSON, R., EGAN, D., FAKES, T., RADCLIFF, B., SHELTON, J. W., (2005). Weighing the world's heaviest telescope at eight points with corrections for lifting perturbations. Precision engineering, 29(3), pp. 354–360.

PIRATELLI-FILHO, A., FERNANDES, F. H. T., ARENCIBIA, R. V., (2012). Application of virtual spheres plate for AACMMs evaluation. Precision engineering, 36(2), pp. 349–355.

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), p. 342.

WANG, F.-K., CHIEN, T.-W., (2010). Process-oriented basis representation for a multivariate gauge study. Computers Industrial Engineering, 58(1), pp. 143–150.

FLYNN, M. J., SARKANI, S., MAZZUCHI, T. A., (2009). Regression analysis of automatic measurement systems. IEEE Transactions on Instrumentation and Measurement, 58(10), pp. 3373–3379.

PARENTE, E., PATEL, H., CALDEO, V., PIRAINO, P., MCSWEENEY, P. L., (2012). RP-HPLC peptide profiling of cheese extracts: a study of sources of variation, repeatability and reproducibility. Food Chemistry, 131(4), pp. 1552–1560.

HOFFA, D. W., LAUX, C. M., (2007). Gauge R&R: an effective methodology for determining the adequacy of a new measurement system for micron-level metrology.

ERDMANN, T. P., DOES, R. J., BISGAARD, S., (2009). Quality quandaries* a gage R&R study in a hospital. Quality Engineering, 22(1), pp. 46–53.

NO, A., COMMITTEE, A. M., (2015). Using the Grubbs and Cochran tests to identify outliers. Analytical Methods, 7(19), pp. 7948–7950.

STEVENS, N. T., STEINER, S. H., BROWNE, R. P., & MACKAY, R. J., (2013). Gauge R&R studies that incorporate baseline information. IIE Transactions, 45(11), pp. 1166–1175.

EXCEL SPC FOR EXCEL, (2013). Access date: 1 January 2017 "Destructive Gage R&R Analysis" https://www.spcforexcel.com/knowledge/measurement-systems-analysis/destructive-gage-rr-analysis.

RELIASOFT CORPORATION., EXPERIMENT DESIGN AND ANALYSIS REFERENCE, (2015), Worldwide Headquarters, p. 379.

LEE, E. F., CZABOTAR, P. E., SMITH, B. J., DESHAYES, K., ZOBEL, K., COLMAN, P. M., FAIRLIE, W. D., (2007). Crystal structure of ABT-737 complexed with Bcl-x L: implications for selectivity of antagonists of the Bcl-2 family. Cell death and differentiation, 14(9), p. 1711.

MIDI, H., ABU BAKAR, N. M., (2015). The Performance of Robust-Diagnostic F In The Identification Of Multiple High Leverage Points. Pakistan Journal of Statistics, 31(5).

FEARN, T., THOMPSON, M., (2001). A new test for 'sufficient homogeneity'. Analyst, 126(8), pp. 1414–1417.

WALSH, S. J., MACSIK, Z., WEGRZYNEK, D., KRIEGER, T., BOULYGA, S., (2016). Model diagnostics for detecting and identifying method repeatability outliers in precision studies: application to a homogeneity study under a two-stage nested ANOVA. Journal of Analytical Atomic Spectrometry, 31(3), pp. 686–699.

ROUSSEEUW, P. J., HUBERT, M., (2011). Robust statistics for outlier detection. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), pp. 73–79.