Malaysian Journal of Mathematical Sciences 15(3): 333–345 (2021)

Malaysian Journal of Mathematical Sciences

Journal homepage: https://einspem.upm.edu.my/journal

Outlier Detection Method in Crossed Gage Repeatability and Reproducibility (R&R) Random Effect Model

Saupi, A. A.[1] and Midi, H.[1,2*]

[1]*Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia,*
[2]*Department of Mathematics & Statistics, Faculty of Science, Universiti Putra Malaysia, Malaysia*

E-mail: habshah@upm.edu.my
Corresponding author

Received: 10 April 2020
Accepted: 9 June 2021

Abstract

Gage Repeatability and Reproducibility (R&R) is the popular method for assessing the capability of a measurement system. Appropriate action can be taken up to improve the quality of the data if measurement system shows incapable. Identification of outliers in measurement data related to manufacturing process is very important since it can affect the efficiency of the measurement system, which lead to misleading prediction and conclusion. Many work on the identification of outliers in linear regression has been explored. However, not much work is devoted to outlier detection method for measurement system data. It is now evident that the classical standardized residual method failed to correctly identify outliers because it is computed based on sample mean. Hence, we propose a new method, which we call robust standardized residual based on median as an alternative to the existing method to rectify the outlier in crossed Gage R&R. The performance of our proposed method is validate through simulation and real data. The results show that our proposed method outperformed the classical method in terms of successfully detect the outliers, without having masking and smaller swamping effects.

Keywords: outliers; robust standardized residual; crossed Gage R&R; masking; swamping.

1 Introduction

Measurement system analysis (MSA) is a rigorous assessment of measurement systems that is crucial in a manufacturing process. Under MSA, measurement accuracy and precision are observed. Measurement accuracy comprises of biasness and linearity of a gage while measurement precision focus on two types of variations which are part to part variation and gage variation. Part variation is the variability due to different items or units being measured while gage variation is due to measurement systems that consists of repeatability and reproducibility (R&R). This paper will focus only on the analysis of measurement's precision which is based on crossed Gage R&R, random effect model. Repeatability is the variation in the measurement system which is due to measurement device while reproducibility is the variation in the measurement system which is caused by differences between operators who record the measurements. The Gage R&R study employs Analysis of Variance (ANOVA) technique to analyse the variation associated with each component, subsequently further analysis is carry out to determine whether or not measurement system precision can be acceptable.

According to [1] and [6], the measurement system is considered acceptable, marginal and not acceptable if the variation of Gage R&R is less than 10%, between 10-30% and greater than 30%, respectively. Moreover, the part-to-part variation is anticipated to be large (more than 80%) for acceptable measurement system.

However, in the presence of outliers, measurement capabilities may be affected and lead to misleading conclusion. [15] highlighted that robust regression method is more reliable than classical least squares method in order to reduce the effect of outliers. Since outliers give an adverse effect on the outcome of any statistical analysis, it is crucial to employ diagnostics checking for outliers before further analysis on the data is carried out. [8] employed statistical behavior of residuals for detecting outliers in a two-way table. However, according to [5], their method is subjected to masking and swamping and has no associated inferential theory. In factorial experiment design, [9] proposed the dynamic variable and Daniel's diagram to detect outliers. This method also suffers from masking effects. [4] and [14] employed the Cook-Statistics method for outlier detection to deal with masking problem and autocorrelation error in designed experiment, respectively. However, no specific cutoff points were given to both methods to detect outliers. [11] claimed that diagnostics method for outlier detection is more powerful if the median absolute deviation is considered rather than the sample mean for standard crossover design. According to [12], median provides a good robust alternative to the sample mean because it is not much affected by outliers.

To date, no study has been focused on identifying outliers in crossed Gage R&R for random effect model. Hence, this inspire us to formulate a diagnostic method to identify outliers in measurement system analysis to fill up the gap in the literature.

2 The Effect of Outliers in Gage R&R Study

2.1 Introduction

Measurement plays an important role to enhance the quality and process performance. An adequate measurement system is desirable for reliable data whereas the variability presence in the measurement is due to their inadequacy of the measurement system itself. The measurement system must be validated through Measurement System Analysis (MSA), whether it is appropriate to be utilized for quality enhancement purposes. MSA that combined measurement accuracy and precision is assessed for the acceptability of the measurement system. In this study, the crossed Gage R&R random effect model is used to evaluate the precision of the measurement system. The Analysis of Variance (ANOVA) approach is being used for isolating the potential sources of variability under the measurement system and decided their adequacy following several acceptance criteria.

2.2 Crossed Gage Repeatability & Reproducibility Random Effect Model

Crossed Gage R&R is a factorial design associated with two or more operators are taking several measurements on each part. The part that being measured is not destroyed after the measurement are taken, hence it could be used repeatedly. In addition, operator and part are considered as independent factors that are used to obtain the response variables, Y_{ijk}. In this model, Y_{ijk} is referring to the measurement that were obtained by operators (incorporates with *i*-th level) replicate (incorporates with *k*-th level) the measurement on different parts (incorporates with *j*-th level).

The model for crossed Gage R&R random effect model can be expressed as follows (see[13]):

$$y_{ijk} = \mu + O_i + P_j + OP_{ij} + \epsilon_{ijk} \begin{cases} i = 1, ..., a, \\ j = 1, ..., b, \\ k = 1, ..., n, \end{cases} \tag{1}$$

where μ denotes as general mean. O_i, P_j, OP_{ij}, ϵ_{ijk} are the effects of the operators, parts, parts and operators interaction, and random errors, respectively. Each of the effects is assumed to be normally and independently distributed, such that $O_i \sim N(0, \sigma_o^2)$, $P_j \sim N(0, \sigma_p^2)$, $OP_{ij} \sim N(0, \sigma_{op}^2)$ and $\epsilon_{ijk} \sim N(0, \sigma_e^2)$.

The total sum of squares of model (1) can be partitioned as follows:

$$\begin{aligned} SS_{Total} &= SS_{Operator} + SS_{Part} + SS_{Operator*Part} + SS_{Error} \\ \sum\sum\sum(y_{ijk} - \bar{y}_{...})^2 &= bn\sum(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum(\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &+ n\sum\sum(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum\sum\sum(y_{ijk} - \bar{y}_{ij.})^2 \end{aligned} \tag{2}$$

where,

$$\begin{aligned} \bar{y}_{i..} &= \frac{\sum\sum y_{ijk}}{bn}, \\ \bar{y}_{.j.} &= \frac{\sum\sum y_{ijk}}{an}, \\ \bar{y}_{ij.} &= \frac{\sum\sum y_{ijk}}{n}, \\ \bar{y}_{...} &= \frac{\sum\sum\sum y_{ijk}}{abn}. \end{aligned}$$

Table 1 shows the analysis of variance table for crossed Gage R&R random effect model. Ideally, due to the assumption that operator is well-trained, thus the operator effects is expected to be insignificant [7]. According to [6], in contrast, the part effect variation is anticipated to be significant which indicates that parts are distinguishable between each other.

Table 1: Analysis of variance table for crossed Gage R&R random effect model.

Sources	Sum Square	df	Mean Square	F-test
Operator	$bn\sum(\bar{y}_{i..}-\bar{y}_{...})^2$	a-1	$MS_o=\frac{SS_{operator}}{a-1}$	$\frac{MS_o}{MS_{op}}$
Part	$an\sum(\bar{y}_{.j.}-\bar{y}_{...})^2$	b-1	$MS_p=\frac{SS_{part}}{b-1}$	$\frac{MS_p}{MS_{op}}$
Operator *Part	$n\sum\sum(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$	(a-1)(b-1)	$MS_{op}=\frac{SS_{operator*part}}{(a-1)(b-1)}$	$\frac{MS_{op}}{MS_e}$
Error	$\sum\sum\sum(y_{ijk}-\bar{y}_{ij.})^2$	ab(n-1)	$MS_e=\frac{SS_{error}}{ab(n-1)}$	
Total	$\sum\sum\sum(y_{ijk}-\bar{y}_{...})^2$	abn-1		

The estimated variances due to operator, part and interaction between operator and part are given as follows:

$$\begin{aligned}\hat{\sigma}^2_{operator}&=\frac{MS_o-MS_{op}}{bn},\\ \hat{\sigma}^2_{part}&=\frac{MS_p-MS_{op}}{an},\\ \hat{\sigma}^2_{operator*part}&=\frac{MS_{op}-MS_e}{n}.\end{aligned}$$

The formulation of gage, part-to-part and total variation denoted as $\hat{\sigma}^2_{gage}$, $\hat{\sigma}^2_{part-to-part}$ and $\hat{\sigma}^2_{total}$ in capabilities measure is as follow:

$$\hat{\sigma}^2_{gage}=\hat{\sigma}^2_{repeatability}+\hat{\sigma}^2_{reproducibility}$$

where,

$$\begin{aligned}\hat{\sigma}^2_{error}&=MS_e=\hat{\sigma}^2_{repeatability},\\ \hat{\sigma}^2_{reproducibility}&=\hat{\sigma}^2_{operator}+\hat{\sigma}^2_{operator*part},\\ \hat{\sigma}^2_{part-to-part}&=\hat{\sigma}^2_{part},\\ \hat{\sigma}^2_{total}&=\hat{\sigma}^2_{gage}+\hat{\sigma}^2_{part}.\end{aligned}$$

As already mentioned, measurement system analysis through Gage R&R is aimed to allocate two main components of variation which are gage variation and part-to-part variation. Based on these two components variations, the credibility of measurement system precision is justified whether or not it is acceptable. Thus, in this regard, several capability measures are defined as follows:

The percentage contribution, or known as %Contribution (denotes as %varcomp) of variance components is calculated through the estimation of each variance over total variance.

Percentage contribution of variance component of the gage:

$$\text{Percentage of Contribution of gage} = \frac{\hat{\sigma}^2_{gage}}{\hat{\sigma}^2_{total}} \times 100.$$

Percentage contribution of variance component part-to-part:

$$\text{Percentage of Contribution of part} = \frac{\hat{\sigma}^2_{part}}{\hat{\sigma}^2_{total}} \times 100.$$

Percentage of Study Variation of the gage:

$$\text{Percentage of studyvar of gage} = \frac{6\times\sqrt{\hat{\sigma}^2_{gage}}}{6\times\sqrt{\hat{\sigma}^2_{total}}} \times 100.$$

Percentage of Study Variation of the part-to-part:

$$\text{Percentage of studyvar of part} = \frac{6\times\sqrt{\hat{\sigma}^2_{part}}}{6\times\sqrt{\hat{\sigma}^2_{total}}} \times 100.$$

The MSA is assessed through acceptance criteria to measure the capability of the measurement system. According to [1] and [6], the percentage of contribution of gage variation and study variation of gage should be below 10%, to precisely indicate as an acceptable measurement system. However, if the values lie between 10% and 30%, it indicates that the measurement system is in a marginal state. The measurement system which exceeds 30% is considered unacceptable (see [1] and [16]). This is bad for the measurement system and it is advisable for measurement systems to be improvised and needs a fixation from operators, equipment, and procedure involved. On the other hand, high values (more than 80%) of the percentage of contribution of part variation and study variation of part signify that measurement system is acceptable.

2.3 Real Data to Show the Importance of Detecting Outliers for Gage Study

To see the effect of outliers on the gage study, Arc Welding data set were presented in Table 2. The data set were taken from [2] which consider 2 operators. Each operator is required to record measurements for five different parts repeatedly for three times. To see the effect of outliers, we purposedly contaminated the data with two outliers (bold and in parentheses) as shown in Table 2.

Table 2: Arc welding data.

		Part 1	**Part 2**	**Part 3**	**Part 4**	**Part 5**
Operator A	Trial 1	0.94	1.05 (**1.75**)	1.03	1.01	0.88
	Trial 2	0.94	1.05	1.02	1.04	0.86
	Trial 3	0.97	1.04	1.05	1.00	0.88
Operator B	Trial 1	0.90	1.03	1.03	1.02	0.87
	Trial 2	0.92	1.04	1.05	1.01	0.88
	Trial 3	0.91	1.01	1.06 (**2.46**)	0.98	0.87

Table 3: ANOVA of crossed Gage R&R random effect model.

Sources	df	Sum Square	Mean Square	F-test	p-value
Parts	4	0.129413 (**0.64648**)	0.0323533 (**0.161620**)	114.413 (**1.83382**)	0.000 (**0.155**)
Operator	1	0.001080 (**0.00901**)	0.0010800 (**0.00901**)	3.819 (**0.10227**)	0.062 (**0.752**)
Repeatability	24	0.006787 (**2.11519**)	0.0002828 (**0.088133**)		
Total	29	0.137280 (**2.77068**)			

Table 4: Component of variances.

Sources	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.0003359 (**0.088133**)	5.91 (**87.80**)
Repeatability	0.0002828 (**0.088130**)	4.98 (**87.80**)
Reproducibility	0.0000531 (**0.000000**)	0.94 (**0.00**)
Part-to-Part	0.0053451 (**0.012248**)	94.09 (**12.20**)
Total Variation	0.0056810 (**0.100381**)	100.00 (**100.00**)

Table 5: Components of standard deviation and 6* sd analysis.

Sources	StdDev(SD)	Study Var (6*SD)	%Study Var
Total Gage R&R	0.0183283 **(0.269872)**	0.109970 **(1.78123)**	24.32 **(93.70)**
Repeatability	0.0168160 **(0.269872)**	0.100896 **(1.78123)**	22.31 **(93.70)**
Reproducibility	0.0072903 **(0.000000)**	0.043742 **(0.00000)**	9.67 **(0.00)**
Part-to-Part	0.0731101 **(0.110670)**	0.438661 **(0.66402)**	97.00 **(34.93)**
Total Variation	0.0753725 **(0.316829)**	0.452235 **(1.90097)**	100.00 **(100.00)**

Table (3-5) exhibit the results of ANOVA, components variances and component of standard deviations for clean and contaminated data (bold and in parentheses). It is interesting to see the results of Table 3 when no outliers are present in the data. As expected, test on parts is significant (p-value < 0.05) and test on operators are not significant (p-value > 0.05) for clean data in Table 3.

It can be observed that outliers have changed the results of Table (3-5). The test on parts in Table 3 is no longer significant (p-value $= 0.155 > 0.05$) which indicates that the gage is not capable of distinguishing between different units. It can be observed from Table 4 that when there is no outlier, the performance of system Gage R&R is good evident by having high percentage of contribution of variation between part (94.09%). However, the results become poor (12.20%) when outliers are present in the data.

The results of Table 5 also indicate that Gage R&R is reliable when no outlier in the data because the percentage contribution to total standard deviation of Gage R&R is less than 30% which is 24.32%. The outliers have increased this percentage of variation to 93.70% which effect the performance of the measurement system Gage R&R.

We have seen that outliers have an adverse effect on the Gage R&R measuring system. Hence, it is very important to identify this outlier before any analysis is carry out.

3 Classical Standardized Residual (CSR)

It can be observed from equation (2) that the variation which is due to error is given by:

$$\sum\sum\sum(y_{ijk} - \bar{y}_{ij.})^2.$$

This implies that the residual denoted as e_{ijk} is given by:

$$\begin{aligned} e_{ijk} &= y_{ijk} - \hat{y}_{ijk} \\ &= y_{ijk} - [\bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})] \\ &= y_{ijk} - \bar{y}_{ij.} \end{aligned}$$

The classical mean square error (CMS_e) is defined as:

$$CMS_e = \frac{\sum\sum\sum(y_{ijk} - \bar{y}_{ij.})^2}{ab(n-1)}. \quad (3)$$

The standardized residual is given by:

$$StdRes = \frac{e_{ijk}}{\sqrt{CMS_e}}.$$

The StdRes is the commonly used measure for identifying outliers. According to [17], an observation is considered as outliers if the absolute value of StdRes is greater than 3. Hereafter, we will refer StdRes as Classical Standardized Residual and denoted as CSR.

4 Robust Standardized Residual (RSR)

The classical standardized residual (CSR) is the widely used method to identify outliers in the analysis of variance model. However, this method is not very successful in detecting outliers since the CSR is based on sample mean in its computation of residuals and mean squares errors. It can be observed from equation (3) that the formulation of CMS_e is based on classical mean which is not resistant to outliers. Hence, in this regard, robust standardized residual (RSR) which is not easily affected by outliers is formulated by incorporating median in the calculation of residuals and mean squares errors.

The median were employed in the formulation of the robust residual and it is given by:

$$e_{ijkr} = (y_{ijk} - y_{ij.median}).$$

Subsequently, we employ median instead of mean in the computational of the robust mean square error, RMS_e. Therefore, our proposed RMS_e is defined as follows:

$$RMS_e = \frac{\sum\sum\sum(y_{ijk} - (y_{ij.median}))^2}{ab(n-1)}.$$

Therefore, our proposed standardized residual, RSR can be defined as follows:

$$RSR = \frac{e_{ijkr}}{\sqrt{RMS_e}}.$$

Since the distribution of RSR is intractable, following [10] and [3], we suggest employing confidence bound type cut-off point for RSR which is median (RSR) + 3*NMAD (RSR), where NMAD(RSR) = Median|RSR- Median(RSR)|/0.6745. Any observation corresponds to RSR which is greater than the cut-off point is considered as an outlier.

5 Simulation Study

In this section, we report a simulation study that is designed to investigate the performance of our proposed method compared to the classical method. We consider 3 operators, 5 parts and varying the number of replications for k= 2,3,4. Each of the observation is generated as follows:

$$y_{i1k} \sim N(30, 0.1),$$

$$y_{i2k} \sim N(32, 0.1),$$
$$y_{i3k} \sim N(34, 0.1),$$
$$y_{i4k} \sim N(36, 0.1),$$
$$y_{i5k} \sim N(38, 0.1),$$

where i=1,2,3, j=1,2,3,4,5 and k=2,3,4. The contamination is created by replacing randomly good data with any maximum number in the data. The simulation is replicated 1000 times.

5.1 Simulation Results

Tables 6 and 7 show the percentage of correct detection of outliers, percentage of masking and percentage of swamping of CSR and RSR for k =2 and 3. The results for k= 4 is not shown due to space limitations. However, the results are consistent.

It is interesting to see from Tables 6 and 7 that the proposed method is successfully detected all the outliers planted in the data, regardless of the number of outliers and sample sizes. Despite having no masking effects, however, our proposed method tends to swamp few inliers.

Nonetheless, the percentage of swamping of RSR seems to be decreasing when more outliers are inserted in the simulated data. Overall, the RSR method shows a consistent performance in detecting the outliers.

On the other hand, the performance of CSR is very poor. It tends to detect few outliers in the data and somehow, this method failed to detect even a single outlier, especially in small data sets. The CSR method suffers badly from masking effects while shows smaller swamping effect. It can be observed that the masking effect is increasing and the percentage of correct detection is decreasing as the number of outliers increases.

Table 6: Percentage of correct detection of outlier, masking and swamping effect for sample size=2.

Design	No of Outlier	Classical Standardized Residual			Robust Standardized Residual		
		Correct Detection (%)	Masking (%)	Swamping (%)	Correct Detection (%)	Masking (%)	Swamping (%)
3(5)(2)	1	0	100	0	100	0	5.34
	2	0	100	0	100	0	6.47
	3	0	100	0	100	0	6.73
	4	0	100	0	100	0	5.85
	5	0	100	0	100	0	3.79
4(5)(2)	1	100	0	2.50	100	0	4.35
	2	14.30	85.70	0	100	0	5.52
	3	5.50	94.50	0.14	100	0	6.24
	4	0.10	99.10	0.01	100	0	6.31
	5	3.30	96.70	0.25	100	0	5.81
3(10)(2)	1	100	0	1.67	100	0	3.60
	2	100	0	0	100	0	4.06
	3	60.60	39.40	1.01	100	0	4.15
	4	47.20	52.80	0.72	100	0	3.59
	5	33.90	66.10	1.26	100	0	2.50
	6	32.30	67.70	1.69	100	0	1.25
	8	27.50	72.50	2.39	100	0	2.57
5(8)(2)	1	100	0	1.25	100	0	3.01
	2	100	0	0	100	0	3.64
	3	100	0	0.65	100	0	4.07
	4	71.30	28.70	1.19	100	0	4.22
	5	61.20	38.80	1.67	100	0	4.24
	6	42.10	57.90	1.38	100	0	4.02
	8	32.70	67.30	1.66	100	0	2.77
	10	26.90	73.10	1.80	100	0	1.41
6(10)(2)	1	100	0	0	100	0	2.51
	2	100	0	0	100	0	2.94
	3	100	0	0.51	100	0	3.24
	4	82.50	17.50	0.46	100	0	3.49
	5	75.90	24.10	0.99	100	0	3.61
	6	68.30	31.70	1.21	100	0	3.57
	8	54.60	45.40	1.55	100	0	3.40
	10	45.10	54.90	1.71	100	0	2.16
	12	41.00	59.00	2.16	100	0	1.48

Table 7: Percentage of correct detection of outlier, masking and swamping effect for sample size=3.

Design	No of Outlier	Classical Standardized Residual			Robust Standardized Residual		
		Correct Detection (%)	Masking (%)	Swamping (%)	Correct Detection (%)	Masking (%)	Swamping (%)
3(5)(3)	1	100	0	0	100	0	7.81
	2	100	0	1.65	100	0	6.90
	3	100	0	3.88	100	0	6.09
	4	63.50	36.40	3.44	100	0	5.11
	5	43.90	56.10	3.09	100	0	3.81
4(5)(3)	1	100	0	0	100	0	7.98
	2	100	0	1.36	100	0	7.31
	3	100	0	2.53	100	0	6.60
	4	100	0	4.06	100	0	5.92
	5	71.90	28.10	2.74	100	0	5.19
3(10)(3)	1	100	0	2.21	100	0	7.93
	2	100	0	0.83	100	0	7.48
	3	100	0	1.11	100	0	6.98
	4	100	0	1.76	100	0	6.43
	5	100	0	2.77	100	0	5.83
	6	86.10	13.90	2.89	100	0	5.09
	7	74.90	25.10	2.90	100	0	4.25
	8	66.50	33.50	2.80	100	0	3.28
	10	60.60	39.40	3.85	100	0	1.99
5(8)(3)	1	100	0	1.67	100	0	7.82
	2	100	0	1.42	100	0	7.51
	3	100	0	0.40	100	0	7.17
	4	100	0	1.30	100	0	6.77
	5	100	0	1.66	100	0	6.37
	6	97.10	2.90	2.28	100	0	5.99
	8	84.30	15.70	3.00	100	0	5.14
	10	70.40	29.60	3.24	100	0	4.09
6(10)(3)	1	100	0	1.11	100	0	7.68
	2	100	0	0.99	100	0	7.50
	3	100	0	0.50	100	0	7.24
	4	100	0	0.26	100	0	7.02
	5	100	0	0.22	100	0	6.81
	6	100	0	0.92	100	0	6.53
	8	96.20	3.80	1.74	100	0	5.94
	10	90.00	10.00	2.33	100	0	5.28
	12	80.50	19.50	2.65	100	0	4.52

5.2 Real Data Example

In order to further assess the performance of our proposed method, real data set is considered in Table 8. The Arc Welding data set taken from [2] is used whereby 2 operators, i=2, taking measurement to 5 different parts, j=5, with three samples, k=3. The data was randomly contaminated with two outliers located at observation number 4 and 24. From the results, the classical method only detect one outlier, meanwhile our proposed method successfully detect all outliers in the data. This shows that our proposed method outperformed the classical method.

Table 8: Numerical result for arc welding data.

Obs	CSR (cut off=3)	RSR (cut-off= 0.1546057)	Obs	CSR (cut off=3)	RSR (cut-off= 0.1546057)
1	0.03450670	0.00000000	16	0.03450670	0.02837979
2	0.03450670	0.00000000	17	0.03450670	0.02837979
3	0.06901340	0.08513936	18	0.00000000	0.00000000
4	1.62181497	**1.98658517**	19	0.01150223	0.00000000
5	0.79365413	0.00000000	20	0.04600894	0.02837979
6	0.82816083	0.02837979	21	0.05751117	0.05675958
7	0.01150223	0.00000000	22	1.66782390	0.05675958
8	0.04600894	0.02837979	23	1.59881050	0.00000000
9	0.05751117	0.05675958	24	**3.26663440**	**4.00155012**
10	0.02300447	0.00000000	25	0.05751117	0.02837979
11	0.08051564	0.08513936	26	0.02300447	0.00000000
12	0.05751117	0.02837979	27	0.08051564	0.08513936
13	0.02300447	0.00000000	28	0.01150223	0.00000000
14	0.04600894	0.05675958	29	0.02300447	0.02837979
15	0.02300447	0.00000000	30	0.01150223	0.00000000

6 Conclusions

In this article, we propose a diagnostic robust method (RSR) for identifying outliers for data used in measurement system analysis. The classical method (CSR) of identifying outliers is not effective as it badly suffers from masking effects. Robust standardized residual, RSR based on median outperforms the classical method in terms of correctly detect the outliers, which is 100% detection without having any masking but only swamp few inliers. The poor performance of classical method is due to the used of sample mean in obtaining MS_e which is sensitive to outliers. Hence it is unable to detect the outliers, having some masking and tendency to swamp clean observation as an outlier. RSR method is highly recommended to identity outliers in a data set to avoid misleading conclusion.

Actknowledgement The authors gratefully acknowledge various helpful comments and suggestions made by the reviewers.

Conflict of Interest The authors declare no conflict of interest.

References

[1] AIAG (2010). *Measurement systems analysis - reference manual. 4th ed.* Chrysler, Ford, General Motors Supplier Quality Requirements Task Force, Southfield, Michigan.

[2] T. Allen (2006). *Introduction to engineering statistics and six sigma*. Springer, London.

[3] A. Bagheri, M. Habshah & R. H. M. R. Imon (2012). A novel collinearity-influential observation diagnostic measure based on a group deletion approach. *Communications in Statistics-Simulation and Computation*, *41*(8), 1379–1396.

[4] L. Bhar, V. K. Gupta & R. Parsad (2013). Detection of outliers in designed experiments in presence of masking. *Statistics and Applications*, *11*(1 & 2), 147–160.

[5] D. Bradu & D. M. Hawkins (1982). Location of multiple outliers in two-way tables, using tetrads. *Technometrics*, *24*(2), 103–108.

[6] M. R. Dhawale & D. N. Raut (2013). Evaluating measurement capabilities by gauge R&R using ANOVA for reliability. *System*, *3*(3), 1143–1148.

[7] D. S. Ermer (2006). Appraiser variation in Gage R&R measurement. *Quality Progress*, *39*(5), 75–78.

[8] J. F. Gentleman & M. B. Wilk (1975). Detecting outliers in a two-way table: I. statistical behavior of residuals. *Technometrics*, *17*(1), 1–14.

[9] J. Goupy (2006). Factorial experimental design: Detecting an outlier with the dynamic variable and the Daniel's diagram. *Chemometrics and Intelligent Labroatory Systems*, *80*(2), 156–166.

[10] M. Habshah, M. R. Norazan & A. H. M. Rahmatullah Imon (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics*, *36*(5), 507–520.

[11] F. P. Lim, I. Mohamed, N. Daud & S. L. Goh (2016). Comparison of outlier detection methods in standard 2x2 crossover design. *Sains Malaysiana*, *45*(3), 499–506.

[12] R. A. Maronna, D. R. Martin & V. J. Yohai (2006). *Robust statistics theory and methods*. John Wiley and Sons, West Sussex.

[13] D. C. Montgomery & G. C. Runger (1993). Gauge capability and designed experiments. Part I: Basic methods. *Quality Engineering*, *6*(1), 115–135.

[14] S. Ojha & L. Bhar (2015). Detection of outliers in designed experiments with correlated errors. *Journal of the Indian Society of Agricultural Statistics*, *69*(1), 57–63.

[15] M. A. Rashid, H. Midi, W. Dhhan & J. Arasan (2021). Detection of outliers in high-dimensional data using nu-support vector regression. *Journal of Applied Statistics*, pp. 1–20. https://doi.org/10.1080/02664763.2021.1911965.

[16] O. Senvar & S. U. O. Firat (2010). An overview of capability evaluation of measurement systems and gauge repeatability and reproducibility studies. *International Journal of Metrology and Quality Engineering*, *1*(2), 121–127.

[17] J. P. Stevens (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, *95*(2), 334–344.