Improved Approximation Scales for Unreplicated Factorial Experiments

F. Aboukalam¹ · M. Alharbi¹ · M. Ishaq Bhatti²

Received: 13 April 2022 / Accepted: 14 September 2022 / Published online: 25 September 2022
© The Author(s) 2022

Abstract
Assessing the sizes of active contrasts in un-replicated factorial and fractional factorial experiments by quick and powerful methods are required in analyzing the big data in various research areas of Human endeavors. One of the old methods based on Lenth (1989) is being used in some statistical and data analytical applications which is fast and less efficient. We propose a new class of tests which are simpler, faster, and more powerful using the location median-function ($\psi_{\text{med}}(x)$) after being skipped one and/or two times. An empirical study of simulation experiments to compute the critical values, sizes and powers using various sample sizes demonstrate the superiority of our methods. The proposed methods are illustrated in examples which can be employed in various fields of research in conducting data analytics using high computing power and machine learning.

Keywords  Un-replicated factorial experiments · Redescending M-estimators for scale · Sizes · Powers · Data analytics

Abbreviations

| Abbreviation | Description |
|--------------|-------------|
| Med          | Median      |
| IoT          | Information of thing |
| DOE          | Design of experiments |
| S            | Scale       |
| $S_0$        | Initial robust scale estimate |

¹ Department of Statistics and Operations Research, College of Sciences, King Saud University, P. O. Box 2455, Riyadh 11451, Saudi Arabia
² SP Jain School of Global Management, Sydney and La Trobe Business School, Melbourne, NSW 2141, Australia
1 Introduction

In recent years, social and physical scientists are interested in searching a quick and powerful data analytics techniques based on computing technologies, information of thing (IoT), machine learning to analyze big data. For example, Salmaso et al. [16], suggested a two-step workflow where design of experiments (DOE) is conducted prior to the usual big data analytics and machine learning modeling phase. They show their DOE on an industrial application. Among others who pointed out the use of un-replicated factorials in the DOE and other areas are Deepa et al. [9] who give an interesting review of literature of the subject whereas Guerra-Zubiaga and Luong [10] address on the energy consumption parameter analysis of industrial robots using factorial DOE without due regards being given to factors and factor interactions which do exist in designing the DOE methodologies with powerful and speedy tests. This paper fills this gap in the literature and present some applications with various examples to demonstrate the proposed tests, sizes and tests power are better than their competitors like Lenth [14] and Aboukalam [2].
It is common practice in data analytics that the ‘response’ in many practical experiments may be affected by multiple factors, but sometimes either the most factors or some factors interactions are not active in various research areas related to big data (see for example [5, 16], Tsai [17]. Hassan et al. (2022)\(^1\) and Wosiak et al. [18] and references there in Deepa et al. (2022)). Considering the experimentation cost, time, effort, and/or limitation of data resources, missing values in the data, the researchers usually apply for estimating and testing main and interaction effect of each factor at ‘two levels’ and do not replicate experiments. Therefore, researchers employ un-replicated \(2^p\) factorial experiments and use only one observation per treatment combination. There are many methods in the literature for analyzing unreplicated \(2^p\) factorial experiments. For example, some of these methods are discussed in Lenth [14], Hamada and Balakrishnan [12], Loughin and Noble [15] and Haaland, O’Connell (1995) and Aboukalam [2] among others. In general, Lenth’s method seems popular and are being used by various researcher. Aboukalam and his associates improved Lenth’s method (see Aboukalam and Al-Shiha [1], Aboukalam [2] and Al-Shiha and Aboukalam [3] using two kinds of powers and size as numerical criteria for illustrating the comparison. Due to recent revolution in computing power in the presence of IoT and machine learning these methods are no more powerful as was reported in earlier literature.

Now, due to advancement in computing power and freely availability of computing software’s like R, the speed of quickness and complex computation is not an issue. This paper attempts to demonstrate how to use both powerful statistical tools to obtain better results. We employed Huber’s skew symmetric function, \(\psi\), used in the location step to find a scale \(S\) which solves the one step equation of the scale-part of \(M\)-estimate\(^2\) given below:

\[
S^2 = S_0^2 \sum_{j=1}^{n} \left[ \psi\left( \frac{x_j}{S_0} \right) \right]^2 \left[ \frac{1}{(n-1) \times B} \right]
\]

(1)

where \(S_0\) is an initial robust scale estimate and the constant \(B\) is chosen to be equal to \(E(\psi^2)\) to make the estimate consistent at the normality. The quick powerful methods of Al-Shiha and Aboukalam [3] and Aboukalam [2] used \(\psi_{med}(x) = Sign(x)\) (signal of x) the location median-function after being skipped one time and two times, respectively. Here, we aim at performing some useful approximations on the Eq. (1) with \(\psi_{med}(x)\) to obtain innovative forms in terms of quickness and powerful results. The improved forms use \(SKM1\) and \(SKM2\) due to that one and two skipping to \(\psi_{med}(x)\) are done, respectively. The improved form \(SKM2\) enables us to generate a class of three competitive quick methods \(SKM2(k \times A)\) \(k = 1, 2, 3\) given that \(A\) is a

\(^1\) Applied on studying the effects of catalyst loading, catalyst selection, porous transport layer type and conductive additive content related water quality issue also BA et al. [5] applied to multifactor unreplicated factorial experiments. Bhatti and Al Shanfari (2017) in clustering whereas Wosiak et al. [18] used measuring the influence on the energy consumption of a vertical electrolyser under natural convection.

\(^2\) For detail understanding and basics of this readers are referred to see [13], page 147.
constant. This task was not possible with the old form used in Aboukalam [2]. The results of \textit{SKM1} and \textit{SKM2(A)} are to update the results of Al-Shiha and Aboukalam [3] and Aboukalam [2], and then compared with the old results without approximations. An empirical comparison showed that both results are comparable, which gives us confidence that approximations are successful and safe.

This paper clarifies that the ranking of the quick methods under study in terms of quality from lowest to highest is; Lenth, \textit{SKM1}, \textit{SKM2(A)}, \textit{SKM2(2A)} and \textit{SKM2(3A)}. The superiority of the new methods \textit{SKM2(2A)}, \textit{SKM2(3A)} over Lenth may sometimes equals several times that of \textit{SKM2(A)} over Lenth. Thus, the added value of these two new quick methods is remarkably high. Moreover, the study is considered as a list, in one place, of all quick methods of the author competing with Lenth’ method. Eventually, necessary proofs are needed, and numerical simulation experiments are conducted under some given level of significances and four selected sample sizes often used in the field to compare the methods. Tables of critical points, sizes and powers are empirically computed.

### 2 Theoretical Objectives

Suppose that $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n$ are the best linear unbiased estimators (BLUE) of the main and interaction effects obtained from an un-replicated $2^p$ factorial experiment in the standard order, where the effects are denoted by $\beta_1, \beta_2, \ldots, \beta_n$. Under the usual assumptions about the experimental errors: normality, independence, and a common unknown variance $\sigma^2$, the estimates $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n$ are independently normally distributed random variables with a common unknown variance $\tau^2 = \sigma^2 / 2^p$ and means $\beta_1, \beta_2, \ldots, \beta_n$ respectively. The statistical inference problem may be stated formally as follows. There are $n$ normal distributions with unknown means $\beta_1, \beta_2, \ldots, \beta_n$, and a common unknown variance $\tau^2$. From each distribution, a single observation $\hat{\beta}_j (j = 1, 2, \ldots, n)$ is obtained. The objective is to infer, if any of the $\beta_j$ is different from zero under a level of significance. In other words, the objective is to determine whether the hypothesis $H_o : \beta_j = 0$ can be rejected under a significance level $\alpha$. This procedure is done on the basis of the observation $\hat{\beta}_j$ and an estimate of the scaler, call it $S$. The hypothesis $H_o$ is rejected if the $j$-th standardized absolute estimate $SAE_j = \frac{|\hat{\beta}_j|}{S}$ is larger than a critical point $cr(\alpha, n)$. As a result, once a good scale estimate has been chosen, those active effects $\beta_j$ corresponding to large values of $SAE_j$ could be identified. Lenth [14] proposed the easy expression of scale:

\[
S_L = 1.5 \times Med\left\{ |\hat{\beta}_j| : |\hat{\beta}_j| \leq 2.5 \times S_0 \right\}
\]

(2)

given that $S_0 = 1.5 \times Med\{ |\hat{\beta}_j| \}$ is an initial scale.

We aim at producing new effective and quick forms of scales.


3 Skipped Median (SKM) scale

Let \( \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_n \) be \( n \) estimated effects from the scale model \( F\left(\hat{\beta}/\tau\right) \). \( S_0 \) be an initial robust scale estimate, like \( S_0 = 1.5 \times \text{Med}\left|\hat{\beta}_j\right| \). The Eq. (1) of one step scale part of M-estimates is given by:

\[
S^2 = S_0^2 \sum_{j=1}^{n} \left[ \psi\left(\frac{\hat{\beta}_j}{S_0}\right)\right]^2 \left[ \frac{1}{(n-1) \times B} \right]
\]

Indeed, we can produce simple forms of scales in using the location median-function:

\[
\psi_{\text{med}}(x) = \text{sign}(x)
\]

after being skipped one time (SKM1) at \( \mp 2.5 \) as following:

\[
\psi_{\text{SKM1}}(x) = \text{sign}(x) \times 1_{[0,2.5]}(\left|x\right|)
\]

or two times (SKM2) at \( \mp 1 \) and \( \mp 2.5 \) as following:

\[
\psi_{\text{SKM2}}(x) = \text{sign}(x) \left[ 1_{[0,1]}(\left|x\right|) + \frac{1}{2} \times 1_{(1,2.5]}(\left|x\right|) \right]
\]

The chosen points \( \mp 2.5 \) are taken to facilitate the comparison with \( S_L \), the scale of Lenth.

Next, it is to approximate the expression (3) with \( \psi_{\text{SKM1}}(x) \) and \( \psi_{\text{SKM2}}(x) \).

3.1 Skipped Median Scale One Time (SKM1)

If \( \psi_{\text{SKM1}} \) is used in expression (3), then:

\[
s^2_{\text{SKM1}} = S_0^2 \sum_{j=1}^{n} \left[ \psi_{\text{SKM1}}\left(\frac{\hat{\beta}_j}{S_0}\right)\right]^2 \left[ \frac{1}{(n-1) \times B} \right]
\]

Assuming that: \( n_0 = \#\hat{\beta}_j;\left|\hat{\beta}_j\right| > 2.5S_0 \), then it is clear that:

\[
\sum_{j=1}^{n} \left[ \psi_{\text{SKM1}}\left(\frac{\hat{\beta}_j}{S_0}\right)\right]^2 = n - n_0,
\]

So:

\[
s^2_{\text{SKM1}} = \left( 1.5 \times \text{Med}\left|\hat{\beta}_j\right| \right)^2 \left( \frac{n - n_0}{n} \right) \left[ \frac{n}{(n-1) \times B} \right]
\]
As simplifying is aimed, we neglect the constant \( \frac{n}{(n-1) \times B} \) and nothing is affected except the critical point, hence:

\[
s_{SKM1} = 1.5 \times Med\left|\hat{\beta}_i\right| \times \left(\frac{n - n_0}{n}\right)^{1/2}.
\]

But as \( \left(\frac{n - n_0}{n}\right)^{1/2} \approx \left(1 - \frac{n_0}{2n}\right) \), then \( s_{SKM1} \) will take the simpler form:

\[
S_{SKM1} \approx Med\left|\hat{\beta}_i\right| \times (1.5 - A_0 \times n_0),
\]

where \( A_0 = \frac{0.75}{n} \).

### 3.2 Skipped Median Scale Two Times (SKM2)

If \( \psi_{SKM2} \) is used in expression (3), then:

\[
s^2_{SKM2} = S_0^2 \sum_{i=1}^{n} \left[ \psi_{SKM2} \left( \frac{\hat{\beta}_i}{S_0} \right) \right]^2 \left[ \frac{1}{(n-1) \times B} \right]
\]

Assuming that: \( n_0 = \#\left|\hat{\beta}_i\right| > 2.5 \times S_0 \) and \( n_{00} = \#\left|\hat{\beta}_i\right| > S_0 \), then it is true:

\[
\sum_{i=1}^{n} \left[ \psi_{SKM2} \left( \frac{\hat{\beta}_i}{S_0} \right) \right]^2 = (n - n_{00}) + \left(\frac{n_{00} - n_0}{4}\right) = n - \frac{3n_{00} + n_0}{4}
\]

So:

\[
s^2_{SKM2} = \left(1.5 \times Med\left|\hat{\beta}_i\right|\right)^2 \left(1 - \frac{3n_{00} + n_0}{4n}\right) \left[ \frac{n}{(n-1) \times B} \right]
\]

Also, neglecting \( \frac{n}{(n-1) \times B} \) affects only the critical point, hence:

\[
s_{SKM2} = 1.5 \times Med\left|\hat{\beta}_i\right| \times \left(1 - \frac{3n_{00} + n_0}{4n}\right)^{1/2}
\]

But as \( \left(1 - \frac{3n_{00} + n_0}{4n}\right)^{1/2} \approx \left(1 - \frac{3n_{00} + n_0}{8n}\right) \), then \( s_{SKM2} \) will take the simpler form:

\[
s_{SKM2} \approx Med\left|\hat{\beta}_i\right| \times \left[1.5 - A \times (3n_{00} + n_0)\right],
\]

where \( A = \frac{1.5}{8n} = \frac{0.1875}{n} \).

Overall, we propose the class of scales \( s_{SKM2(k \times A)} \) with three degrees \( k = 1, 2, 3 \) of multiplying the constant \( A \):

\[
s_{SKM2(A)} = Med\left|\hat{\beta}_i\right| \times \left[1.5 - A \times (3n_{00} + n_0)\right]
\]
The values 2 and 3 of the factor \( k \) can be seen as reflection of some suitable shapes of the used \( \psi \) – function, or even no worry if they are very successful proposals.

Finally, to help the experimenter remember all the aforesaid quick scales we propose to read \( s_{SKM_1} \) and \( s_{SKM_2(A)} \) using the unified form:

\[
s_{SKM} = \text{Part of} \ 1.5 \times \text{Med of Entire} \{ \hat{\beta}_j \}.
\]

whereas Lenth’s scale is read using the parallel way:

\[
S_L = \text{Entire} \ 1.5 \times \text{Med of Part of} \{ \hat{\beta}_j \}.
\]

4 Simulation Studies and Results

4.1 Critical Points of the Test Statistics

Scales are usually compared under similar probabilities (=\( \alpha \)) of rejecting the null hypothesis \( H_0 : \beta_j = 0 \) given that all the \( n \) effects \( \beta_j \) are not active, that is \( H_0 \) is true. To fulfill this requirement, we should compute the empirical critical point, \( cr(\alpha, n) \), of the test statistics under \( H_0 \). The value of \( cr(\alpha, n) \) is \( 100(1-\alpha)\% \) quantile point of the distribution of the test statistic \( SAE_j \). The reference distribution function of this statistic is built empirically based on several runs \( (nr=10,000) \) of samples of size \( n \) from the standard normal. Hence, the empirical distribution includes \( nr \times n \) values of the aforesaid test statistic. For some values of \( \alpha \) and four selected sample sizes \( n \), the empirical critical points of Lenth, \( SKM_1 \), \( SKM_2(A) \), \( SKM_2(2A) \) and \( SKM_2(3A) \) test statistics are given in Table 1 in the appendix.

4.2 Numerical Criteria; Sizes and Powers

These five techniques, like Lenth, \( SKM_1 \), \( SKM_2(A) \), \( SKM_2(2A) \) and \( SKM_2(3A) \) will be assessed and compared in this study using numerical criteria. We intend to contribute toward the concept of power, defined as the probability of declaring an effect as active given that it is an active effect. The Power of the proposed test is accompanied with the size of the test. Here, the size we mean the probability of declaring an effect as active given that it is inactive. As is well known that the size must be close to the significance level \( \alpha \) when \( H_0 \) is true. Here we consider, \( \alpha \) is an error premium.
that is paid in the null case, and many experimenters may not care about it in the non-null case if it stays within the volume $\alpha$. In general, any technique would be the best if its powers are the superior than the competing tests and the sizes are the inferior.\(^3\) Moreover, the decline in Powers indicates that the test losses some real active effects. This usually happens if the band $S \times cr(\alpha, n)$ becomes lager or the scale $S$ has the deficiency of inflation. The increase in size indicates that the test losses some real inactive effects. This usually happens if the band becomes smaller or the scale $S$ has the deficiency of shrinkage. Powers should be first looked at. Moreover, if two techniques have quite comparable powers, then the better is the technique whose sizes is lower. Because for this, the method with lower sizes will cause less false active alarms. Consequently, the further investigations will cost less.

The relevant powers and sizes were empirically assessed since $nr=10,000$ runs of standard normal random samples of sizes $n=15, 31, 63$ and 127. Let $na$ be the number of active effects that are planted in the introduced sample. In each run, the first $na$ effects, $\{\beta_j, j = 1, 2, \ldots, na\}$, were violated and made active effects and the rest $(n-na)$ were kept on as inactive effects. The selected number ‘na’ was given the values 1, 2, 4, 6 and 7 as $n=15$; the values 3, 6, 9, 12 and 15 as $n=31$; the values 6, 12, 18, 24 and 31 as $n=63$; and eventually the values 12, 24, 36, 48 and 63 as $n=127$. The effect $\beta_j$ is made active if a shift, say $\Delta$, is added to the observation: $\beta_j (i.e., \hat{\beta}_j + \Delta)$. The value of the shift $\Delta = 0$ considers the null case, and we propose the other values of $\Delta (= 2, 3, 4, 5$ and 6). The simulation, like the authors’ 2001, 2006, 2007 papers, cover only the violation in the shift $\Delta$. This kind of violation is considered in the field.

To assure the correctness of the critical points, the simulation has been conducted for the null case $\Delta = 0$, the sizes and the powers of all methods come out close to the predefined level $\alpha$. The relatively high significance level $\alpha = 0.15, 0.20$ are often selected to offer the desired active effect more chances of appearing. Eventually, the empirical Sizes and Powers are defined by:

$$[\text{Size}]\{\text{Pow}\} = \frac{\text{Num of declared actives they are inactive actives in all runs}}{\text{inactives actives in all runs}} \frac{nr \times (n-na)}{nr \times na}$$

Indeed, we need global ideas about the sizes and powers apart of huge details about the figure $na$. Therefore, we propose to increase $nr=50,000$, and in each run $j$ the value of $na_j$ is selected randomly from the set of all its possible values. The global powers ($GPow$) and the global sizes ($GSize$) are recomputed according to the laws:

\(^3\) See [6] for details when designing optimal tests in testing cluster effects, computing test size, critical values and the doing empirical power comparisons in search of uniformly most powerful tests. It was cited by Angelopoulos et al. [4] in various applications in clustering and used by Bhatti and Al Shanfari (2017).
4.3 Results Discussion

Figure 1 highlights that GPow of Lenth and SKM1 methods are comparable, but it is growing gradually over them with SKM2(A), SKM2(2A) and SKM2(3A) methods, respectively. The supreme of SKM2(3A) over Lenth may sometimes reach to a great success of 12 units, as in the case SKM2(3A), $n = 127$ and $\Delta = 5$. Whereas the supreme of the SKM2(A) over Lenth cannot exceed 4 unites. The GSize of SKM1, SKM2(A) and SKM2(2A) are lesser than this of the Lenth. For SKM2(3A), there is similarity in low shifts and the lesser is again otherwise. Briefly, the save from the deficiencies of the shrinkage or the inflation that may happen can be ranked from the lowest to the highest when using quick methods under study is as follows: Lenth, SKM1, SKM2(A), SKM2(2A) and SKM2(3A). The next applications highlight this property.
4.4 Applications

Application (1): Fig. 2, show the normal probability plot of the ordered absolute effects: 0.42, 0.73, 0.94, 0.98, 1.11, 2.73, 4.70, 5.10, 5.11, 5.58, 5.80, 6.65, 8.42. The data includes 15 effects where the last 7 of them are actives.

Fig. 1 (continued)

Fig. 2 The inflation deficiency of Lenth’s technique. Effects: −2.50, −1.19, −0.42, 0.73, 0.94, 0.98, 1.11, 2.73, 4.70, 5.10, 5.11, 5.58, 5.80, 6.65, 8.42. Lenth bands = ± 5.916, SKM1 bands = ± 5.852, SKM2(A) bands = ± 5.411, SKM2(2A) bands = ± 5.002, SKM2(3A) bands = ± 4.393
Here, only 2 active effects are detected, and the other 5 active effects are lost. SKM1 – scale:

\[ S_{SKM1} = 2.73 \times (1.5 - 0.05 \times 0) = 4.095, \]

Here, only 2 active effects are detected, and the other 5 active effects are lost. SKM2 (A)– scale:

\[ S_{SKM2(A)} = 2.73 \times [1.5 - 0.0125 \times (3 \times 7 + 0)] = 3.378, \]

Here, only 4 active effects are detected, and the other 3 active effects are lost. SKM2 (2A)– scale:

\[ S_{SKM2(2A)} = 2.73 \times [1.5 - 0.025 \times (3 \times 7 + 0)] = 2.662, \]

Here, only 6 active effects are detected, and the other 1 active effect is lost. SKM2 (3A)– scale:

\[ S_{SKM2(3A)} = 2.73 \times [1.5 - 0.0375 \times (3 \times 7 + 0)] = 1.945, \]

Here, all 7 active effects are detected, and nothing is lost. Figure 2 displays the bands and the inflation deficiency of Lenth in comparing with SKM1 and SKM2. Let’s consider another application.
Application (2): In Fig. 3, the normal probability plot of Example II of Box and Meyer [8] is done. The ordered absolute effects of this example are as follows: 0.03, 0.05, 0.13, 0.13, 0.13, 0.15, 0.15, 0.30, 0.37, 0.37, 0.40, 0.40, 0.42, 2.15, and 3.10. There are 15 effects, where the last two of them are likely to be real active effect.

The initial figures needed for all methods starts from:

\[
\begin{align*}
M ed \left| \hat{\beta}_j \right| &= 0.30 \quad & S_0 = 1.5 \times 0.30 = 0.45, \\
n_0 &= \left\{ \left| \hat{\beta}_j \right| : \left| \hat{\beta}_j \right| > 2.5 \times S_0 = 1.125 \right\} = 2 \quad & n_{00} &= \left\{ \left| \hat{\beta}_j \right| : \left| \hat{\beta}_j \right| > S_0 = 0.45 \right\} = 2 \\
A_0 &= \frac{0.75}{n} = 0.05 \quad & A &= \frac{0.1875}{n} = 0.0125 \\
\end{align*}
\]

Lenth – scale:

\[
S_L = 1.5 \times M ed \left\{ \left| \hat{\beta}_j \right| : \left| \hat{\beta}_j \right| \leq 2.5 \times S_0 = 1.125 \right\} = 1.5 \times 0.15 = 0.225,
\]

Lenth – bands = ± \( S_L \times cr(0.15, 15) = ± 0.225 \times 1.444741 = ± 0.3251 \).

Here, 7 effects are detected as actives and 5 of them are false.

SKM1 – scale:

\[
S_{SKM1} = 0.30 \times (1.5 - 0.05 \times 2) = 0.42,
\]

SKM1 – bands = ± \( S_{SKM1} \times cr(0.15, 15) = ± 0.42 \times 1.429107 = ± 0.6002 \).

Here, 2 effects are detected as actives, and nothing is false.

SKM2 (A)– scale:

\[
S_{SKM2(A)} = 0.30 \times [1.5 - 0.0125 \times (3 \times 2 + 2)] = 0.42,
\]

Fig. 3  The shrinkage deficiency of Lenth’s technique. Effects: −0.37, −0.15, −0.05, −0.03, 0.13, 0.13, 0.15, 0.15, 0.30, 0.37, 0.40, 0.40, 0.42, 2.15, 3.10, Lenth bands = ± 0.3251, SKM1 bands = ± 0.6002, SKM2(A) bands = ± 0.6728, SKM2(2A) bands = ± 0.7328, SKM2(3A) bands = ± 0.8130
SKM2 (2A)–bands = ± $S_{SKM2(2A)} \times cr(0.15, 15) = \pm 0.42 \times 1.601896 = \pm 0.6728$.

Here, 2 effects are detected as actives, and nothing is false.

SKM2 (2A)–scale:

$$S_{SKM2(2A)} = 0.30 \times [1.5 - 0.025 \times (3 \times 2 + 2)] = 0.39,$$

SKM2 (2A)–bands = ± $S_{SKM2(2A)} \times cr(0.15, 15) = \pm 0.39 \times 1.878985 = \pm 0.7328$.

Here, 2 effects are detected as actives, and nothing is false.

SKM2 (3A)–scale:

$$S_{SKM2(3A)} = 0.30 \times [1.5 - 0.0375 \times (3 \times 2 + 2)] = 0.36,$$

SKM2 (3A)–bands = ± $S_{SKM2(3A)} \times cr(0.15, 15) = \pm 0.36 \times 2.258468 = \pm 0.8130$.

Here, 2 effects are detected as actives, and nothing is false. Figure 3 displays the bands and the shrinkage deficiency of Lenth in comparing with SKM1 and SKM2.

5 Concluding Remarks

This paper proposes a new class of simpler, faster, efficient, and more powerful method of approximations using the location median-function ($\psi_{mod}(x)$). The speed of approximation is achieved by skipping one and/or two times the simulation process. We observe that model based on SKM2(3A)-technique is far better than its competitors recorded in the literature including Lenth technique. The proposed technique is quite simpler, faster, and free from the deficiencies of the inflation and/or the shrinkage that may exist. The comparative analysis with previous studies is done in presenting mathematical proof with simulated tables of the critical values, sizes, and powers. The proposed methods are illustrated in examples to demonstrate its application in data analytics.

Appendix

See Tables 1 and 2 here.
Table 1  Critical values $cr(\alpha, n)$ for Lenth, SKM1, SKM2 (A), SKM2 (2A) and SKM2 (3A)

| Critical points | $n / \alpha$ | 0.05 | 0.10 | 0.15 | 0.20 |
|-----------------|-------------|------|------|------|------|
| Lenth           | 15          | 2.175109 | 1.708962 | 1.444741 | 1.259765 |
|                 | 31          | 2.065653 | 1.683632 | 1.448644 | 1.273831 |
|                 | 63          | 2.01142 | 1.665788 | 1.446368 | 1.280277 |
|                 | 127         | 1.981849 | 1.652366 | 1.441615 | 1.280456 |
| SKM1            | 15          | 2.128298 | 1.682562 | 1.429107 | 1.24662 |
|                 | 31          | 2.029886 | 1.660282 | 1.433915 | 1.261612 |
|                 | 63          | 1.993603 | 1.649207 | 1.430479 | 1.266952 |
|                 | 127         | 1.967636 | 1.640926 | 1.430609 | 1.270705 |
| SKM2 (A)        | 15          | 2.389733 | 1.888494 | 1.601896 | 1.393617 |
|                 | 31          | 2.29744 | 1.873511 | 1.614379 | 1.421202 |
|                 | 63          | 2.246882 | 1.860413 | 1.61225 | 1.427282 |
|                 | 127         | 2.22694 | 1.855757 | 1.61733 | 1.435525 |
| SKM2 (2A)       | 15          | 2.862901 | 2.232305 | 1.878985 | 1.62734 |
|                 | 31          | 2.713685 | 2.189577 | 1.875339 | 1.647283 |
|                 | 63          | 2.625892 | 2.161185 | 1.871249 | 1.653958 |
|                 | 127         | 2.580188 | 2.14718 | 1.868548 | 1.657729 |
| SKM2 (3A)       | 15          | 3.56442 | 2.734363 | 2.258468 | 1.922252 |
|                 | 31          | 3.31491 | 2.641947 | 2.247257 | 1.958603 |
|                 | 63          | 3.170222 | 2.592427 | 2.234016 | 1.969145 |
|                 | 127         | 3.090701 | 2.557701 | 2.219304 | 1.9669 |
Table 2  Global Sizes and Powers for Lenth, SKM1, SKM2 (A), SKM2 (2A) and SKM2 (3A) for $\alpha = 0.15$, and selected $n = 15, 31, 63, 127$

| N   | Shift | 2       | 3       | 4       | 5       | 6       | 2       | 3       | 4       | 5       | 6       |
|-----|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| n=15| Lenth | 0.048   | 0.037   | 0.045   | 0.061   | 0.079   | 0.419   | 0.606   | 0.788   | 0.913   | 0.969   |
|     | SKM1  | 0.045   | 0.031   | 0.031   | 0.033   | 0.034   | 0.416   | 0.607   | 0.786   | 0.912   | 0.970   |
|     | SKM2(A) | 0.047  | 0.031   | 0.030   | 0.030   | 0.030   | 0.419   | 0.621   | 0.811   | 0.931   | 0.980   |
|     | SKM2(2A) | 0.050  | 0.036   | 0.036   | 0.037   | 0.039   | 0.412   | 0.625   | 0.827   | 0.943   | 0.987   |
|     | SKM2(3A) | 0.057  | 0.044   | 0.046   | 0.048   | 0.051   | 0.410   | 0.634   | 0.846   | 0.959   | 0.993   |
| n=31| Lenth | 0.043   | 0.035   | 0.046   | 0.063   | 0.080   | 0.419   | 0.610   | 0.770   | 0.880   | 0.951   |
|     | SKM1  | 0.042   | 0.029   | 0.030   | 0.034   | 0.036   | 0.420   | 0.608   | 0.768   | 0.882   | 0.952   |
|     | SKM2(A) | 0.042  | 0.029   | 0.028   | 0.029   | 0.029   | 0.423   | 0.620   | 0.790   | 0.906   | 0.969   |
|     | SKM2(2A) | 0.046  | 0.034   | 0.036   | 0.039   | 0.040   | 0.421   | 0.633   | 0.814   | 0.930   | 0.981   |
|     | SKM2(3A) | 0.052  | 0.043   | 0.048   | 0.055   | 0.060   | 0.420   | 0.646   | 0.842   | 0.953   | 0.990   |
| n=63| Lenth | 0.041   | 0.033   | 0.045   | 0.064   | 0.081   | 0.420   | 0.607   | 0.747   | 0.833   | 0.918   |
|     | SKM1  | 0.040   | 0.028   | 0.030   | 0.034   | 0.036   | 0.423   | 0.606   | 0.745   | 0.837   | 0.920   |
|     | SKM2(A) | 0.041  | 0.028   | 0.027   | 0.029   | 0.029   | 0.426   | 0.618   | 0.768   | 0.869   | 0.947   |
|     | SKM2(2A) | 0.044  | 0.033   | 0.035   | 0.038   | 0.040   | 0.424   | 0.636   | 0.795   | 0.898   | 0.969   |
|     | SKM2(3A) | 0.047  | 0.040   | 0.047   | 0.055   | 0.061   | 0.423   | 0.653   | 0.828   | 0.932   | 0.984   |
| n=127| Lenth | 0.041   | 0.033   | 0.046   | 0.06   | 0.084   | 0.426   | 0.612   | 0.738   | 0.796   | 0.877   |
|     | SKM1  | 0.039   | 0.028   | 0.030   | 0.034   | 0.037   | 0.424   | 0.604   | 0.731   | 0.796   | 0.878   |
|     | SKM2(A) | 0.040  | 0.027   | 0.027   | 0.028   | 0.028   | 0.424   | 0.616   | 0.753   | 0.830   | 0.914   |
|     | SKM2(2A) | 0.042  | 0.032   | 0.034   | 0.037   | 0.039   | 0.427   | 0.636   | 0.785   | 0.870   | 0.950   |
|     | SKM2(3A) | 0.048  | 0.038   | 0.046   | 0.056   | 0.0630  | 0.431   | 0.659   | 0.822   | 0.912   | 0.976   |
Acknowledgements  We are thankful to two anonymous referees and the handling editor for constructive comments which have improved the quality of the paper. We take all responsibilities of any errors.

Author Contributions  The following authors and their initials are used. Fayz Aboukalam (FA), Maryam Alharbi (MA), M. Ishaq Bhatti (MIB). “FA carried out the initial model based on earlier work, participated in the sequence alignment, and drafted the manuscript. MA carried out simulation under supervision of FA and MIB. All three authors (FA, MA and MIB) participated in the design of the study and performed the statistical analysis. All authors read, reviewed the final draft, and approved the final manuscript in revised format to be submitted to JSTA.”

Funding  Not applicable.

Availability of Data and Material  Its simulated Empirical study, the data can be provided on request.

Declarations

Competing interests  Not applicable.

Ethics Approval and Consent to Participate  Not applicable.

Consent for Publication  All authors give consent for publication.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Aboukalam, M.A.F., Al-Shiha, A.A.: A robust analysis for unreplicated factorial experiments. Comput. Stat. Data Anal. 36(1), 31–46 (2001)
2. Aboukalam, F.: More on quick analysis of unreplicated factorial designs avoiding shrinkage and inflation deficiencies. Int. J. Reliabil. Appl. 7, 167–175 (2006)
3. Al-Shiha, A.A., Aboukalam, M.A.F.: Quick and easy analysis of unreplicated factorial designs avoiding shrinkage deficiency. J. Stat. Theory Appl. 6, 35–43 (2007)
4. Angelopoulos, P., Koukouvinos, C., Skountzou, A.: Clustering effects in unreplicated factorial experiments. Commun. Stat. Simul. Comput. 42(9), 1998–2007 (2013)
5. Ba, I., Bl, A., Gm, O.: A proposed method of identifying significant effects in un-replicated factorial experiments. Annals. Comput. Sci. Series 17, 2 (2019)
6. Bhatti, M.I.: Cluster effects in mining complex data. Nova Science Publishers, Incorporated, NY, USA (2012)
7. Bhatti, M.I., Al-Shanfari, H.: Econometric analysis of model selection and model testing. Routledge (2017)
8. Box, G.E., Meyer, R.D.: An analysis for unreplicated fractional factorials. Technometrics 28, 11–18 (1986)
9. Deepa, N., Pham, Q.V., Nguyen, D.C., Bhattacharya, S., Prabadevi, B., Gadekallu, T.R., Pathirana, P.N.: A survey on blockchain for big data: approaches, opportunities, and future directions. Future Gener. Comput. Syst. Appear 131, 209 (2022)
10. Guerra-Zubiaga, D.A., Luong, K.Y.: Energy consumption parameter analysis of industrial robots using design of experiment methodology. Int. J. Sustain. Eng. 14(5), 996–1005 (2021)
11. Haaland, P.D., Connell, O., M. A.: Inference for effect-saturated fractional factorials. Technometrics 37, 82–93 (1995)
12. Hamada, M., Balakrishnan, N.: Analyzing unreplicated factorial experiments: A review with some new proposal. Stat. Sin. 8, 1–41 (1998)
13. Huber, P.J.: Robust Statistics. Wiley, New York (1981)
14. Lenth, R.V.: Quick and easy analysis of unreplicated factorials. Technometrics 31, 469–473 (1989)
15. Loughin, T.M., Noble, W.: A permutation test for effects in an unreplicated factorial design. Technometrics 39, 180–190 (1997)
16. Salmaso, L., Pegoraro, L., Giancristofaro, R.A., Ceccato, R., Bianchi, A., Restello, S., Scarabottolo, D.: Design of experiments and machine learning to improve robustness of predictive maintenance with application to a real case study. Commun. Stat. Simul. Comput. 51(2), 570–582 (2022)
17. Tsai, S.F.: Analyzing dispersion effects from replicated order-of-addition experiments. J. Qual. Technol. 1, 1–18 (2022)
18. Wosiak, G., da Silva, J., Sena, S.S., Carneiro-Neto, E.B., Lopes, M.C., Pereira, E.: Investigation of the influence of the void fraction on the energy consumption of a vertical electrolyser under natural convection. J. Environ. Chem. Eng. 10(3), 107577 (2022)