Comprehensive Evaluation of Reference Values of Parametric and Non-Parametric Effect Size Methods for Two Independent Groups

Ayşegül Yabacı Tak[1,*] and Ilker Ercan[2]

[1]Bezmialem Vakıf University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Istanbul, Turkey

[2]Bursa Uludag University, Faculty of Medicine, Department of Biostatistics, Bursa, Turkey

Abstract: In the field of health and other sciences, effect size (ES) provides a scientific approach to the effectiveness of treatment or intervention. The p-value indicates whether the statistical difference depends on chance, while ES gives information about the effectiveness of the treatment or intervention, even if the difference is not significant. For this reason, ES has become a very popular measure in recent years. It depends on which ES will be used based on the distribution of data and the number of groups. In this study, parametric and non-parametric ES were evaluated for two independent groups.

When the literature was examined, there were no studies aimed at evaluating the reference values of the parametric and non-parametric ES methods used for two independent groups. In this study, the reference values of parametric and non-parametric ES methods for two independent groups were re-evaluated by a simulation study. As a result, the very small reference value of parametric ES methods was determined differently from the literature. It has been seen that the reference values of non-parametric ES methods are valid in cases where the skewness is low, and new reference values have been proposed at the varying skewness level.

Keywords: Effect Size, Parametric Effect Size, Non-Parametric Effect Size, Two Independent Groups.

1. INTRODUCTION

Effect Size (ES) is easy to calculate and understand in many disciplines. ES apart from statistical significance scientifically explains the magnitude of treatment or effectiveness. In recent years, various definitions have emerged with the popularity of ES. To determine the size of the effect of an intervention, three different aspects of the effect must be determined. These are effect size dimension, effect size measure, and effect size value [1]. The main idea of the effect size dimension is information that can be measured. For example, in a situation where the relationship between two variables is investigated, ES can be determined as the correlation coefficient. Another aspect of effect size is the effect size measure. It can be summarized as the mathematical index used to determine the size of the relevant effect. An example of this is the Cohen d effect size formula. After determining the effect size dimension and effect size measure, a concrete value of the size of the intervention should be obtained, which is called the effect size value. In light of this information, the definition of effect size is made as a real value obtained from the effect size measure based on statistics or parameters [1].

In the health sciences, the statistical significance of a treatment or drug between placebo and treatment groups may vary depending on the sample size. Sometimes, because of the large or small sample size, insignificant effects can be interpreted as significant, and significant effects can be interpreted as insignificant. Because ES is generally independent of the sample size, it shows the ES of the treatment or drug, even if the difference between the groups in clinical studies is not significant. Various effect size methods have been proposed according to the number of groups and the distribution of the data. In this article, we discussed Cohen d, Hedge g, Glass delta, Cliff delta, Vargha and Delaney A (VDA), and Glass Rank-Biserial Correlation coefficient (r_{rb}), which are frequently used parametric and non-parametric effect sizes for two independent groups. When the interpretations of these ES indexes in the literature were evaluated, Cohen [2] classified the effect sizes as small (d=0.2), medium (d=0.5), and large (d ≥ 0.8). The main reason for establishing the classes as small, medium, and large in the effect size interpretation is to determine the most appropriate sample size with power analysis in planning the study [3]. Based on the research findings in the literature, reference intervals were developed by Sawilowsky [4] as d=0.01 very small, d=0.20 small, d=0.50 medium, d=0.80 large, d=1.2 very large, and d=2.0 huge. The values corresponding to Cohen d reference intervals of other ES methods were presented by Cohen [5]. The interpretation of the effect sizes used in cases where two independent groups are not parametric is as follows: the values of 0.11, 0.28, and 0.43 for the cliff delta, 0.56, 0.64, and 0.71 for the VDA, and 0.10, 0.30, and 0.50 for the r_{rb} effect size criterion correspond to the small, medium, and large effect size values, respectively [6].

In the literature, it has been observed that there are limited studies on the evaluation of the similarities,

*Address correspondence to this author at the Bezmialem Vakıf University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Istanbul, Turkey; E-mail: aysegulyabaci@gmail.com

E-ISSN: 1929-6029/22

reference intervals, and performances of parametric and non-parametric ES methods. Therefore, this paper is aimed to re-evaluate the reference values of parametric and non-parametric effect size methods given in the literature for two independent groups and to evaluate the similarities of the methods. In addition, it is also aimed to compare and re-evaluate the reference values for the varying states of skewness (for skewness=0.5, 1.5, and 2) with the values in the literature for non-parametric effect size methods.

2. DATA ASSUMPTIONS FOR EFFECT SIZE

The most important assumptions in the dataset for calculating ES are the assumption of "normality" and "homogeneity of variances". As with statistical analysis methods, effect size methods are divided into parametric effect size methods and nonparametric effect size methods according to the normality of the data. When the assumption of normality is provided, it becomes easier to determine ES between groups, because most of the data are gathered around the mean value. When the assumption of the normal distribution is not provided, the data are mostly not gathered around the mean. Therefore, it would not be correct to use mean values in determining ES between groups. In this case, non-parametric effect size methods are used by taking into account the ordering of the observations. The other assumption is the homogeneity of variances. The assumption of homogeneity of variances between groups is important in parametric ES methods. The assumption assumes that the variances of independent groups with continuous variables are similar, equal, or equivalent. When these two assumptions are provided, parametric ES estimators are consistent and asymptotic estimators. When the assumptions are not provided, it is suggested to use non-parametric ES methods available in the literature.

3. CONVERSION OF ES METHODS

For a more universal interpretation, other standard ES indices can be used by converting them to Cohen d. The equations used for the conversion of non-parametric ES indices to Cohen d for two independent groups in the paper are given below [7].

- Conversion from Cohen d to r_{rb} can be done using Eq. (1):

$$r_{rb} = \frac{d}{\sqrt{d^2+a}} \quad (1)$$

a is a correction factor in cases where $n_1 \neq n_2$ and is calculated as given in Eq.(2). The correction factor (a) depends on the ratio n_1 to n_2 instead of the absolute values of these numbers.

$$a = \frac{(n_1+n_2)^2}{n_1 n_2} \quad (2)$$

- Conversion from r_{rb} to Cohen d can be done using Eq. (3):

$$r_{rb} = \frac{2r}{\sqrt{1-r^2}} \quad (3)$$

- Conversion from Cliff Delta (δ) to Cohen d: Eq.(4) is used to convert the Cliff delta effect size estimate to the Cohen d effect size estimate so that it does not overlap between two standard normal distributions, with Φ^{-1} the normal cumulative distribution is the inverse of the function.

$$d(\delta) = 2z\frac{-1}{\delta-2}\ , z_p \equiv \Phi^{-1}(p) = AUC^{-1}(p) \quad (4)$$

- Conversion from Cohen d to Cliff Delta (δ): Eq. (5) is used to convert the Cohen d effect size estimate to the Cliff delta (δ) effect size so that it does not overlap between the two standard normal distributions.

$$\delta(d) = \frac{2AUC(\frac{d}{2})-1}{AUC(\frac{d}{2})}\ , AUC(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\, dt \quad (5)$$

There is a linear relationship between Cliff delta and VDA effect size, which is converted as $VDA = \frac{(Cliff\ Delta+1)}{2}$.

4. MATERIAL AND METHOD

This study is aimed to compare the performance of commonly used ES methods and to re-evaluate reference values used in interpreting ES. R Studio program was used in the simulation study. Datasets are derived from the normal distribution for parametric data and the Fleishman distribution for non-parametric data [8, 9]. K- means clustering algorithm was used to evaluate the similarity of the methods and their reference values. The Chalinski-Harabasz (CH) index and Silhouette (S) index were used to determine the optimal number of clusters [10, 11]. In the simulation studies, simulti corrdata, effect size, eff size, rcompanion, ppclust, factoextra, dplyr, cluster, psych, ClusterR, gridExtra, and readxl R program packages were used [12-20].

4.1. K-Means Clustering Algorithm and Methods for Determining the Optimal Number of Clusters

Clustering algorithms are methods developed to divide the ungrouped data matrix into homogeneous subgroups according to certain characteristics. Clusters are homogeneous within themselves and heterogeneous among themselves. K-Means clustering algorithm was developed by MacQueen [21]. The method allows each data to belong to only one cluster. Therefore, it is a sharp and non-hierarchical clustering algorithm. The purpose of the algorithm is to divide d data points of n dimensions into k clusters, thus minimizing the sum of squares within the cluster. Results vary depending on the number of clusters. Therefore, cluster verification is required by re-running the algorithm for different cluster numbers or using the indexes suggested in the literature to determine the optimal number of clusters. In this study, Chalinski-Harabasz (CH) and Silhouette index (S), which are commonly used to determine the optimal number of clusters in the K-means clustering algorithm, were calculated.

4.2. Simulation Algorithm

The simulation algorithm for parametric and non-parametric ES methods for two independent groups is given step by step in sections 4.2.1 and 4.2.2.

4.2.1. Simulation Algorithm for Parametric ES Methods

Step 1: x and y group effect size reference intervals are very small $(d = 0.01)$, small $(d = 0.20)$, medium $(d = 0.50)$, large $(d = 0.80)$, very large $(d = 1.20)$, and huge $(d = 2)$ is generated from the normal distribution for $n = 1000$ and $t = 1000$ replicates by taking constant standard deviations (n: sample size, t: number of replicates).

Step 2: Parametric ES methods are applied to the input matrix. The input matrix obtained for the parametric methods is given in $\boldsymbol{Z_{parametric}}$ Eq. (6).

$$\boldsymbol{Z_{parametric}} = \begin{bmatrix} d_1 & g_1 & delta_1 \\ \vdots & \vdots & \vdots \\ d_{6000} & g_{6000} & delta_{6000} \end{bmatrix} \quad (6)$$

Step-3: The optimal number of clusters for the k-means clustering algorithm to be applied to the input matrix is determined and the most appropriate reference value for the methods is evaluated.

4.2.2. Simulation Algorithm for Non-Parametric ES Methods

Step 1: x and y group is derived from Fleishman distribution for $n = 1000$ and $t = 1000$ and skewness (γ_1)=$0.5, 1.5$ and 2. Reference values considered in data derivation are $0.11, 0.28$ and 0.43 for Cliff delta, $0.56, 0.64$ and 0.71 for VDA, and $0.10, 0.30$ and 0.50 for r_{rb} ES method.

Step-2: Cliff delta, VDA and r_{rb} methods are applied to the input matrix. The input matrix obtained for non-parametric ES methods is given in $\boldsymbol{Z_{non-parametric}}$ Eq. (7).

$$\boldsymbol{Z_{non-parametric}} = \begin{bmatrix} Cliff\ delta_1 & VDA_1 & r_{rb1} \\ \vdots & \vdots & \vdots \\ Cliff\ delta_{3000} & VDA_{3000} & r_{rb3000} \end{bmatrix} \quad (7)$$

Step-3: The optimal number of clusters for the k-means clustering algorithm to be applied to the input matrix is determined and the most appropriate reference value for the methods is evaluated.

4.3. Simulation Results

4.3.1. Simulation Results for Parametric ES Methods

CH and S index values were calculated to determine the optimal number of clusters in the $\boldsymbol{Z_{parametric}}$ matrix are given in Table **1**.

Table 1: CH and S Index Values Calculated for the $Z_{parametric}$ Matrix

Number of Clusters (k)	CH Index	S Index
k=2	17539.94	0.65
k=3	35943.04	0.75
k=4	71220.26	0.78
k=5	144086.52	0.79
k=6	223372.86	0.81

According to the CH and S index, the optimal number of clusters with the highest CH and S index (CH=223372.86, S=0.81) for the dataset was determined as k=6. The reference values obtained as a result of the k-means clustering algorithm applied for k=6 are given in Table **2**.

According to Table **1**, Cohen d, Hedge g, and Glass delta effect size methods are quite similar to each other in terms of reference values. The new reference values proposed as a result of the simulation study are the same as the reference values in the literature, except for the small effect size reference value. The new reference values determined according to the optimal number of clusters for parametric ES methods are; A very small effect size was obtained as 0.0441 for Cohen d, 0.0440 for Hedge g, and Glass delta. The small effect size was obtained as 0.2068 for the Cohen d, Hedge, Glass delta. The medium effect size was obtained as 0.5005 for Cohen d, 0.5003 for Hedge g, and 0.5005 for Glass delta. The large effect size was obtained as 0.8002 for Cohen d, 0.7999 for Hedge g, and 0.8004 for Glass delta. The very large effect size was obtained as 1.2003 for Cohen d, 1.1998 for Hedge g, and 1.2005 for Glass delta. The huge effect size was 2.000 for Cohen d, 1.9994 for Hedge g, and 20006 for Glass delta. In summary, 0.044 will mean very small, 0.20 small, 0.50 medium, 0.80 large, 1.20 very large, and 2 huge effects. The graphic obtained for the visual form of the cluster numbers is given in Figure **1**.

Table 2: Cluster Analysis Results for k=6 of Parametric ES Methods Applied in the $Z_{parametric}$ Matrix

Number of Clusters (k=6)	n=1000 t=1000	Cohen d	Hedge g	Glass Delta
Cluster 1 (Very Small) n=1057	Minimum	0.0001	0.0001	0.0001
	Maximum	0.1252	0.1251	0.1252
	Standard Deviation	0.0323	0.0323	0.0323
	Mean	0.0441	0.0440	0.0440
	Median	0.0382	0.0381	0.0379
Cluster 2 (Small) n=944	Minimum	0.1256	0.1255	0.1262
	Maximum	0.3510	0.3509	0.3529
	Standard Deviation	0.0422	0.0421	0.0421
	Mean	0.2068	0.2068	0.2068
	Median	0.2029	0.2029	0.2034
Cluster 3 (Medium) n=998	Minimum	0.3551	0.3550	0.3546
	Maximum	0.6480	0.6477	0.6433
	Standard Deviation	0.0481	0.0481	0.0484
	Mean	0.5005	0.5003	0.5005
	Median	0.4993	0.4991	0.4998
Cluster 4 (Large) n=1001	Minimum	0.6507	0.6504	0.6508
	Maximum	0.9528	0.9525	0.9657
	Standard Deviation	0.0500	0.0499	0.0509
	Mean	0.8002	0.7999	0.8004
	Median	0.7991	0.7988	0.8013
Cluster 5 (Very Large) n=1000	Minimum	1.0502	1.0498	1.0384
	Maximum	1.3558	1.3553	1.3731
	Standard Deviation	0.0520	0.0520	0.0546
	Mean	1.2003	1.1998	1.2005
	Median	1.1993	1.1989	1.2009
Cluster 6 (Huge) n=1000	Minimum	1.8133	1.8176	1.8092
	Maximum	2.1632	2.1623	2.2126
	Standard Deviation	0.0583	0.0583	0.0652
	Mean	2.0001	1.9994	2.0006
	Median	2.0000	1.9993	2.0007

4.3.2. Simulation Results for Non-Parametric ES Methods

The CH and S index values are related to determining the optimal number of clusters in the $\boldsymbol{Z}_{non-parametric}$ matrix for γ_1=0.5,1.5 and 2 are given in Table **3**.

The optimal number of clusters with the highest CH and S index was $k = 3$ for the $\boldsymbol{Z}_{non-parametric}$ matrix with γ_1=0.5,1.5 and 2 according to the CH and S index. The reference values obtained as a result of the k-means clustering algorithm applied for k=3 and the skewness of 0.5, 1.5, and 2 are given in Tables **4-6**.

4.3.2.1. Simulation Results for $\boldsymbol{\gamma_1}$=0.5

When non-parametric ES methods are clustered for k=3 and $\gamma_1 = 0.5$, small effect size was obtained as 0.0874 for Cliff delta, 0.5498 for VDA, and 0.1071 for r_{rb}. The medium effect size was obtained as 0.2461 for Cliff delta, 0.6230 for VDA, and 0.3184 for r_{rb}. The large effect size was obtained as 0.4043 for Cliff delta, 0.6944 for VDA, and 0.5280 for r_{rb}. It is seen that different results are obtained from the literature reference values within three non-parametric ES methods.

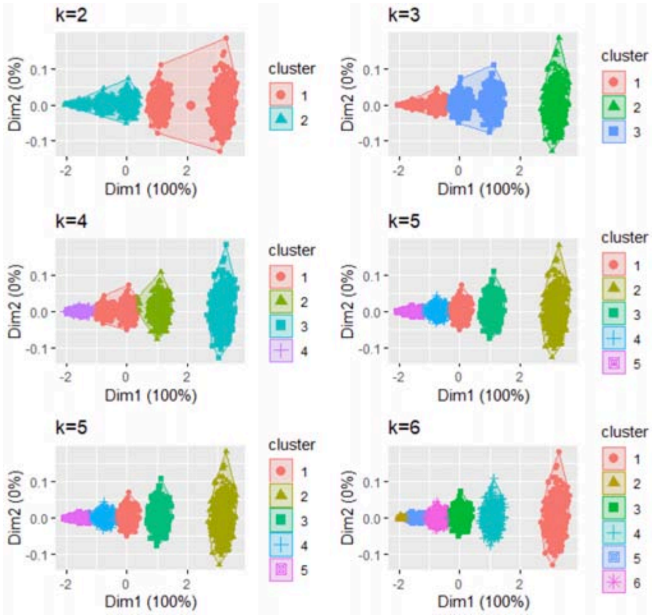


Figure 1: Visual form of cluster numbers of parametric ES methods.

Table 3: CH and S Index when γ_1=0.5, 1.5 and 2

γ_1	Number of Clusters (k)	CH Index	S Index
0.5	k=2	19376.13	0.66
	k=3	38819.39	0.73
1.5	k=2	10675.85	0.76
	k=3	68606.38	0.85
2	k=2	12308.90	0.78
	k=3	72726.57	0.85

Table 4: K-Means Clustering Analysis Results for γ_1=0.5 and k=3 of Non-Parametric ES Methods

Number of Clusters (k=3)	n=1000 t=1000 γ_1=0.5	Cliff Delta	VDA	r_{rb}
Cluster 1 (Small) n=1000	Minimum	0.0076	0.5098	0.0380
	Maximum	0.1659	0.5830	0.1740
	Standard Deviation	0.0260	0.0130	0.0224
	Mean	0.0874	0.5498	0.1071
	Median	0.0881	0.5503	0.1080
Cluster 2 (Medium) n=1000	Minimum	0.1668	0.5890	0.2500
	Maximum	0.3131	0.6555	0.3730
	Standard Deviation	0.0250	0.0125	0.0205
	Mean	0.2461	0.6230	0.3184
	Median	0.2470	0.6235	0.3200
Cluster 3 (Large) n=1000	Minimum	0.3265	0.6566	0.4700
	Maximum	0.4674	0.7262	0.5810
	Standard Deviation	0.0232	0.0117	0.0169
	Mean	0.4043	0.6944	0.5280
	Median	0.4048	0.6947	0.5280

Table 5: K-Means Clustering Analysis Results for γ_1=1.5 and k=3 of Non-Parametric ES Methods

Number of Clusters (k=3)	n=1000 t=1000 γ_1=1.5	Cliff Delta	VDA	r_{rb}
Cluster 1 (Small) n=1000	Minimum	0.0572	0.5365	0.0910
	Maximum	0.2106	0.6126	0.2200
	Standard Deviation	0.0253	0.0126	0.0217
	Mean	0.1373	0.5765	0.1601
	Median	0.1381	0.5759	0.1600
Cluster 2 (Medium) n=1000	Minimum	0.2460	0.6230	0.3180
	Maximum	0.3897	0.6935	0.4430
	Standard Deviation	0.0241	0.0121	0.0198
	Mean	0.3225	0.6612	0.3823
	Median	0.3226	0.6613	0.3820
Cluster 3 (Large) n=1000	Minimum	0.4015	0.6948	0.5120
	Maximum	0.5438	0.7656	0.6130
	Standard Deviation	0.0224	0.0113	0.0167
	Mean	0.4751	0.7307	0.5640
	Median	0.4756	0.7308	0.5630

4.3.2.2. Simulation results for $\boldsymbol{\gamma_1}$=1.5

When non-parametric ES methods are clustered for k=3 and $\gamma_1 = 1.5$, small effect size was obtained as 0.1373 for Cliff delta, 0.5765 for VDA, and 0.1601 for r_{rb}. The medium effect size was obtained as 0.3225 for Cliff delta, 0.6612 for VDA, and 0.3823 for r_{rb}. The large effect size was obtained as 0.4751 for Cliff delta, 0.7307 for VDA, and 0.5640 for r_{rb}. It is seen that the results obtained in the three non-parametric ES methods are different compared to the literature.

4.3.2.3. Simulation Results for $\boldsymbol{\gamma_1}$=2

When non-parametric ES methods are clustered for k=3 and $\gamma_1 = 2$, small effect size was obtained as 0.1701 for Cliff delta, 0.5937 for VDA, and 0.1926 for r_{rb}. The medium effect size was obtained as 0.3708 for

Table 6: K-Means Clustering Analysis Results for γ_1=2 and k=3 of Non-Parametric ES Methods

Number of Clusters (k=3)	n=1000 t=1000 γ_1=2	Cliff Delta	VA	r_{rb}
Cluster 1 (Small) n=1000	Minimum	0.0931	0.5554	0.1270
	Maximum	0.2385	0.6280	0.2530
	Standard Deviation	0.0251	0.0125	0.0215
	Mean	0.1701	0.5937	0.1926
	Median	0.1711	0.5941	0.1930
Cluster 2 (Medium) n=1000	Minimum	0.2895	0.6448	0.3580
	Maximum	0.4490	0.7171	0.4830
	Standard Deviation	0.0237	0.0119	0.0195
	Mean	0.3708	0.6854	0.4246
	Median	0.3715	0.6858	0.4250
Cluster 3 (Large) n=1000	Minimum	0.4495	0.7245	0.5430
	Maximum	0.5877	0.7879	0.6440
	Standard Deviation	0.0221	0.0111	0.0166
	Mean	0.5226	0.7546	0.5924
	Median	0.5231	0.7548	0.5930

Cliff delta, 0.6854 for VDA, and 0.4246 for r_{rb}. The large effect size was obtained as 0.5226 for Cliff delta, 0.7546 for VDA, and 0.5924 for r_{rb}.

The visual shapes of the clusters created according to the optimal number of clusters are given in Figures **2-4**.

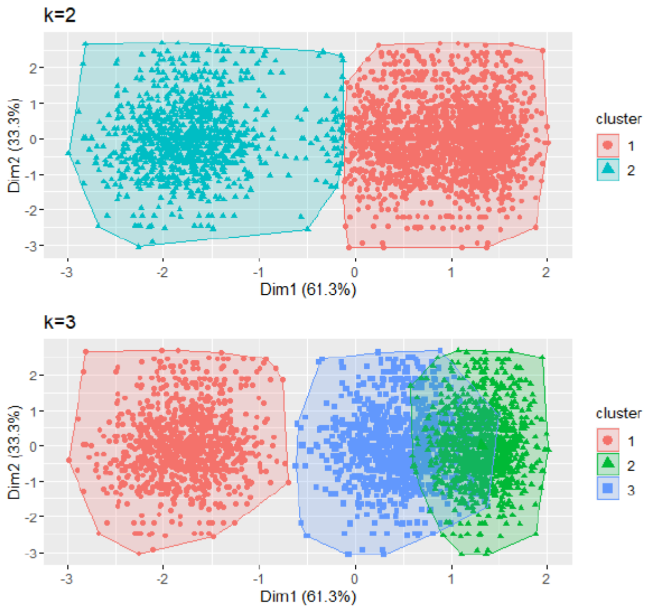


Figure 2: Visual shape of clusters with $\boldsymbol{\gamma_1 = 0.5}$.

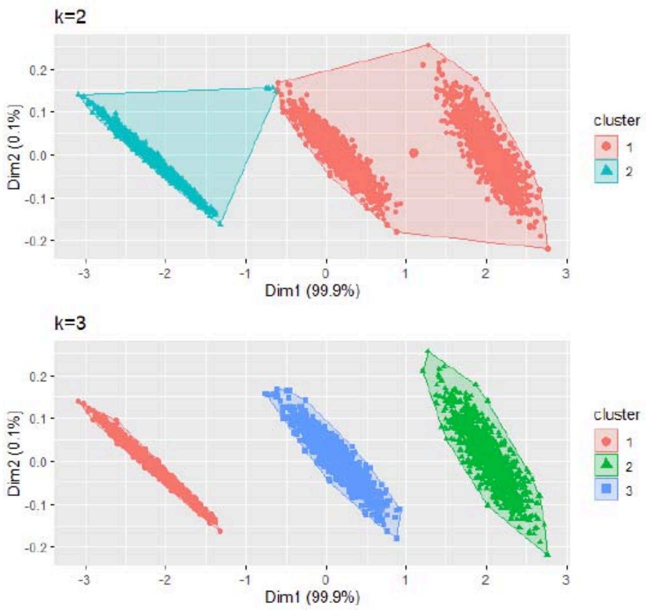


Figure 3: Visual shape of clusters with $\boldsymbol{\gamma_1 = 1.5}$.

As a result, the three clustered methods were not similar in terms of reference values, and the proposed new reference values differed from the reference values in the literature. In addition, as the skewness value increases, the reference values of three non-parametric ES methods also increase. It has been concluded that the reference values given for each method in the literature are valid when skewness is low, and as a result of the simulation study, these methods cluster better when $\gamma_1 = 1.5$ and 2. The change in reference values of non-parametric ES methods with the increase in skewness is given in Figure **5**.

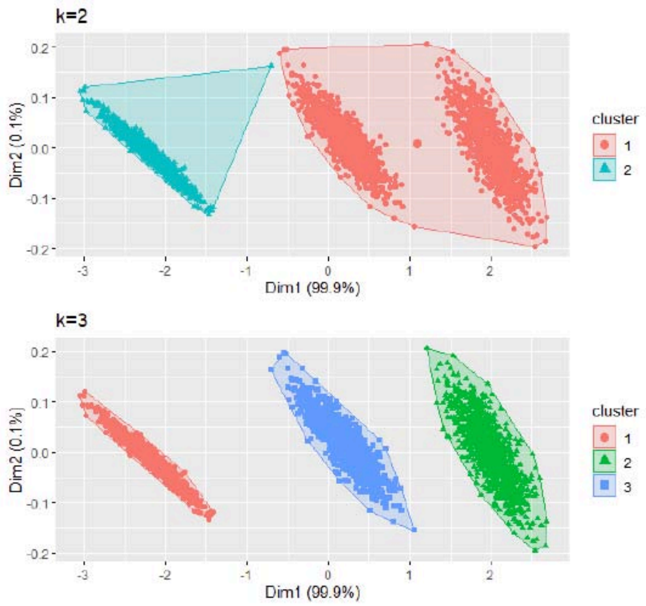
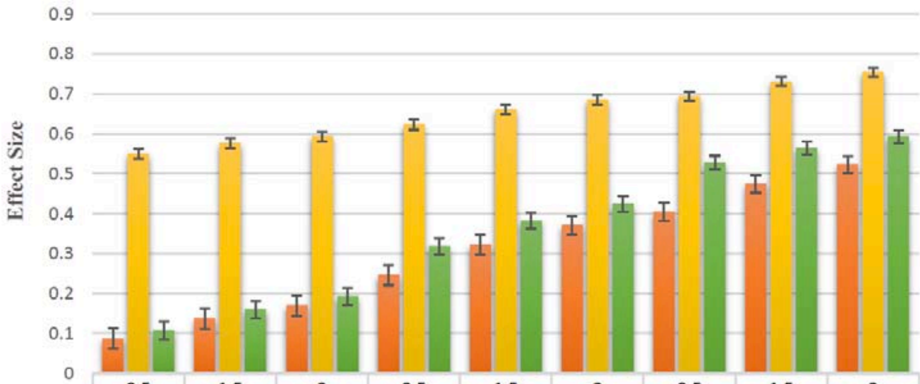


Figure 4: Visual shape of clusters with $\boldsymbol{\gamma_1 = 2}$.

5. CONCLUSION

In scientific studies, when the p-value is less than a determined significance level, the result is considered to be statistically significant. However, the p-value is a result that indicates whether this difference or relationship is due to chance. ES provides a scientific approach to the size of the intervention or effectiveness. For this reason, researchers' interest in ES methods has increased [22, 23]. ES methods used vary depending on whether the groups are dependent or independent and the number of groups in the experimental design. Cohen d, Hedge g, and Glass delta effect size methods have been proposed if the variables for the two independent groups provide the assumptions of normality and variance homogeneity [5, 24, 25]. In the case where the assumption of normality and homogeneity of variances is not provided, non-parametric effect size methods are proposed that take into account the rankings of the data instead of the mean. Cliff delta, VDA, r_{rb} are ES methods for two nonparametric independent groups [26-28].

Simulation applications for the evaluation of reference intervals of parametric and non-parametric ES methods have not been encountered in the literature. In this study, data were derived by keeping the standard deviations from the normal distribution based on the six reference value for n=1000 and t=1000. After applying parametric ES methods for two independent groups to the derived dataset, the optimal number of clusters was determined as k=6 according to



	0,5	1,5	2	0,5	1,5	2	0,5	1,5	2
	Small			Medium			Large		
Cliffdelta	0.0874	0.1373	0.1701	0.2461	0.3225	0.3708	0.4043	0.4751	0.5226
VDA	0.5498	0.5765	0.5937	0.6230	0.6612	0.6854	0.6944	0.7307	0.7546
Glass r	0.1071	0.1601	0.1926	0.3184	0.3823	0.4246	0.5280	0.5640	0.5924

Cliffdelta VDA Glass r

Figure 5: Change of reference values of non-parametric ES methods when skewness 0.5, 1.5 and 2.

the CH and S index. As a result of the k-means clustering algorithm applied for k=6, the reference value of 0.010, which is classified as very small in the literature, of Cohen d, Hedge g, and Glass delta effect size methods was obtained as 0.044 in this study. Small, medium, large, very large, and huge ES reference values were obtained as the same as 0.20, 0.50, 0.80, 1.20, and 2 reference values in the literature. Non-parametric data were derived from Fleishman distribution to evaluate the reference intervals of Cliff delta, VDA and r_{rb}, which are non-parametric ES methods. For the dataset that does not provide normal distribution, the reference values of the methods were examined according to the varying skewness and kurtosis values. Skewness=0.5 and kurtosis= -0.8161896, skewness=1.5 and kurtosis=2.4658850, skewness=2 and kurtosis=5.3377003 were evaluated. According to the CH and S index, the optimal number of clusters was determined to be k=3. According to the simulation results, it was seen that the reference values in the literature for all three methods were valid when the skewness and kurtosis were low, and the reference values increased as the skewness and kurtosis increased. It was also concluded that the methods clustered better when the skewness was 1.5 and 2.

As a result, in the study, the three parametric ES methods evaluated were similar to each other and a new reference value was determined for the very small ES with the simulation study. It was determined that the three non-parametric ES methods evaluated were different from each other in terms of reference values and the values in the literature were valid when the dataset was close to normal. The reference values of the three non-parametric effect size methods where the skewness is 0.5, 1.5, and 2 are suggested by the simulation study.

REFERENCES

[1] Kelley K, Preacher KJ. On effect size. Psychological Methods 2012; 17: 137. https://doi.org/10.1037/a0028086

[2] Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates 1988.

[3] Valentine JC, Cooper H. A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). Psychological Methods 2008; 13: 130. https://doi.org/10.1037/1082-989X.13.2.130

[4] Sawilowsky SS. New effect size rules of thumb. Journal of Modern Applied Statistical Methods 2009; 8: 26. https://doi.org/10.22237/jmasm/1257035100

[5] Cohen J. The statistical power of abnormal-social psychological research: a review. The Journal of Abnormal and Social Psychology 1962; 65: 145. https://doi.org/10.1037/h0045186

[6] Peng C-YJ, Chen L-T. Beyond Cohen's d: Alternative effect size measures for between-subject designs. The Journal of Experimental Education 2014; 82: 22-50. https://doi.org/10.1080/00220973.2012.745471

[7] Borenstein M, *et al*. Converting among effect sizes. Introduction to Meta-Analysis 2009; pp. 45-49.

[8] Fleishman AI. A method for simulating non-normal distributions. Psychometrika 1978; 43: 521-532. https://doi.org/10.1007/BF02293811

[9] Luo H. Generation of Non-normal Data: A Study of Fleishman's Power Method, ed, Department of Statistics, Uppsala University, 2011.

[10] Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics-theory and Methods 1974; 3: 1-27. https://doi.org/10.1080/03610927408827101

[11] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis Journal of Computational and Applied Mathematics 1987; 20: 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

[12] Ben-Shachar MS, *et al*. Package 'effectsize', 2021.

[13] Cebeci Z, *et al*. Package 'pplust'. CRAN, 2019.

[14] Fialkowski A. SimMultiCorrData: Simulation of Correlated Data with Multiple Variable Types. 2018, URL https://CRAN. R-project. org/package= SimMultiCorrData. R package version 0.2 1, pp. p251.

[15] Kassambara A, Mundt F. Package 'factoextra'. Extract and Visualize the Results of Multivariate Data Analyses 2017; 76.

[16] Mangiafico S, Mangiafico MS. Package ‘rcompanion’. Cran Repos 2017; 20: 1-71.

[17] Rogmann JJ, Rogmann MJJ. Package ‘orddom’, 2013.

[18] Torchiano M, Torchiano MM. Package ‘effsize’. Package “Effsize”[(accessed on 9 March 2021)] 2020.

[19] Wickham H, *et al*. Package ‘readxl’ 2019.

[20] Yarberry W. DPLYR, in CRAN Recipes, Springer, 2021; pp. 1-58. https://doi.org/10.1007/978-1-4842-6876-6_1

[21] MacQueen J. Some methods for classification and analysis of multivariate observations, in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA, 1967; pp. 281-297.

[22] Ellis P. Thresholds for interpreting effect sizes. Retrieved January 13, 2009; pp. 2014.

[23] Sullivan GM, Feinn R. Using effect size—or why the P value is not enough; Journal of graduate medical education 2012; 4: 279-282. https://doi.org/10.4300/JGME-D-12-00156.1

[24] ---, Primary, secondary, and meta-analysis of research. Educational Researcher 1976; 5: 3-8. https://doi.org/10.3102/0013189X005010003

[25] Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics 1981; 6: 107-128. https://doi.org/10.3102/10769986006002107

[26] Cliff N. Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin 1993; 114: 494. https://doi.org/10.1037/0033-2909.114.3.494

[27] Glass GV. A ranking variable analogue of biserial correlation: Implications for short-cut item analysis. Journal of Educational Measurement 1965; 2: 91-95. https://doi.org/10.1111/j.1745-3984.1965.tb00396.x

[28] Vargha A, Delaney HD. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. Journal of Educational and Behavioral Statistics 2000; 25: 101-132. https://doi.org/10.3102/10769986025002101

Received on 29-07-2022 Accepted on 11-09-2022 Published on 19-09-2022

https://doi.org/10.6000/1929-6029.2022.11.11