Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

Performance of Rand's C statistics in clustering analysis: an application to clustering the regions of Turkey

Sinan Saraçli*

*Correspondence:
ssaracli@aku.edu.tr
Department of Statistics, Faculty of Arts and Sciences, Afyon Kocatepe University, Afyonkarahisar, TR-03200, Turkey

Abstract

Purpose: When a clustering problem is encountered, the researcher must be aware that choosing an incorrect clustering method and distance measure may significantly affect the results of the analysis. The purpose of this study is to determine the best clustering method and distance measure in cluster analysis and to cluster the regions of Turkey on the basis of this result.

Methods: In hierarchical clustering, there are several clustering methods and distance measures. For comparison of the clustering methods and distance measures, Rand's C statistic is one of the best methods. Rand's comparative statistic C takes on values from 0.0 to 1.0 inclusive that may be used to compare two resultant clusterings produced by applying clustering methods to a data set with unknown structure or to assess the performance of a clustering method on a data set with known structure.

Results: In this study, the seven regions of Turkey are clustered by all the clustering methods and distance measures. Related with the social and economic indicators, the final cluster number is taken as three. Then, according to Rand's C statistics, all possible pairs of distance measures for all clustering methods in hierarchical clustering are compared, and the results are given in the related tables.

Conclusions: According to the results of all possible comparisons, Ward's method is found to be the best among others, and final clustering of the regions is applied according to Ward's clustering measure.

Keywords: Rand's C statistics; hierarchical clustering methods; distance measures

1 Introduction

The word 'classification' can be used in a broad sense to include various types of diagrams that indicate either the relative degrees of similarities or the lines of descent [1]. The term *Cluster Analysis* encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. Clustering algorithms are often used to find homogeneous subgroups of entities depicted in a set of data [2].

Cluster analysis divides data into groups (clusters) that are meaningful, useful or both. If meaningful groups are the goal, then the clusters should capture the natural structure of the data. The sample characteristics are used to group the samples. Grouping can be arrived at either hierarchically partitioning the samples or non-hierarchically partitioning the samples. Thus, segmentation methods include probability-based grouping of observations and cluster (grouping)-based observations. They include hierarchical (tree-based

method) and non-hierarchical (agglomerative) methods. A good clustering method will produce high quality clusters with high intra-class similarity and low inter-class similarity. Classes, or conceptually meaningful groups of objects that share common characteristics, play an important role in how people analyze and describe the world [3].

A general question that researchers face in many areas of inquiry is how to organize observed data into meaningful structures, that is, how to develop taxonomies. In other words, cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Given the above, cluster analysis can be used to discover structures in data without providing an explanation/interpretation. In other words, cluster analysis simply discovers structures in data without explaining why they exist [4].

As Rand [5] mentioned in his study, many intuitively appealing methods had been suggested for clustering data; however, interpretation of their results had been hindered by the lack of objective criteria. For this purpose, he developed C statistics which depends on a measure of similarity between two different clusterings of the same set of data, and the measure essentially considers how each pair of data points is assigned to each clustering.

Rand [5] developed a comparative statistic C which takes on values from 0.0 to 1.0 inclusive that may be used to compare two resultant clusterings produced by applying clustering methods to a data set with unknown structure or to assess the performance of a clustering method on a data set with known structure. When C is equal to 1.0, there is a perfect agreement in the comparison. However, the meaning of a C value between 0.0 to 1.0 is not clear. Thus, a means of attaching statistical significance to the values of the C statistic is needed.

Ferreira and Hitchcock [6] compared the performance of four major hierarchical methods (single linkage, complete linkage, average linkage and Ward's method) for clustering functional data. They used the Rand index to compare the performance of each clustering method.

2 Method

According to Rand [5], the simple computational form for c is, for given N points, $X_1, X_2, \ldots, X_N$, and two clusterings of them $Y = \{Y_1, \ldots, Y_{K1}\}$ and $Y' = \{Y'_1, \ldots, Y_{K2}\}$,

$$c\big(Y, Y'\big) = \sum_{i<j}^{N} \gamma_{ij} \Big/ \binom{N}{2}, \tag{1}$$

where

$$\gamma_{ij} = \begin{cases} 1 & \text{if there exist } k \text{ and } k' \text{ such that both } X_i \text{ and } X_j \text{ are in both } Y_k \text{ and } Y'_{k'}, \\ 1 & \text{if there exist } k \text{ and } k' \text{ such that } X_i \text{ is both } Y_k \text{ and } Y'_{k'} \text{ while } X_j \text{ is in} \\ & \text{neither } Y_k \text{ and } Y'_{k'}, \\ 0 & \text{otherwise.} \end{cases}$$

For a given pair of clusterings Y and Y' of the same N points, arbitrarily number the clusters in each clustering and let n_{ij} be the number of points simultaneously in the ith

Table 1 Life expectancy, education, income and human development indexes for the regions of Turkey

Region	Life expectancy index	Education index	Income (GDP) index	Human development index
Central Anatolia	0.724	0.777	0.692	0.731
Mediterranean	0.753	0.755	0.802	0.770
Eastern Anatolia	0.705	0.622	0.378	0.568
Southeastern Anatolia	0.727	0.600	0.474	0.600
Aegean	0.736	0.759	0.817	0.771
Marmara	0.748	0.803	0.946	0.832
Blacksea	0.713	0.735	0.614	0.687

cluster of Y and the jth cluster of Y'. Then the similarity between Y and Y' is as follows:

$$c(Y, Y') = \frac{[\binom{N}{2} - [(\frac{1}{2})\{\sum_i(\sum_j n_{ij})^2 + \sum_j(\sum_i n_{ij})^2\} - \sum\sum n_{ij}^2]]}{\binom{N}{2}}. \tag{2}$$

In clustering analysis it is known that the most used seven distance measures are squared Euclidean, cityblock, Minkowski, cosine, customized, correlation (Pearson) and Chebychev, and there are seven clustering methods, which are average, centroid, complete, median, single, Ward and weighted method.

2.1 Data set

To apply and see the results of Rand's C statistics, regions of Turkey are considered with their life expectancy index, education index and income indexes as an illustrative example. The provinces of Turkey are organized into seven census-defined regions, which were originally defined at the First Geography Congress in 1941 [7]. They are CA: Central Anatolia, M: Mediterranean, A: Aegean, MA: Marmara, EA: Eastern Anatolian, SEA: Southeastern Anatolia, BS: Blacksea. Human development index is calculated by considering the life expectancy index, education index and income indexes that express three common characteristics of regions. Human development index is accepted as an important criterion to determine the development levels of the countries [8]. The related index values for each region of Turkey are given in Table 1 (Source: [9]).

In hierarchical cluster analysis, there are two final clusters at the end. Because it is hard to see the efficiencies of the distance measures and clustering methods for two clusters, the final cluster number is considered as three and the results are observed, whether the regions join the same cluster or not, for all clustering methods and distance measures. Table 2 shows the results of these analyses. For each measure, there are 21 results, which are calculated with all the possible combinations of seven regions with two groups. Then all the clustering methods are compared, and according to these comparisons, together in both, separate in both and mixed groups and Rand's C statistics are calculated. The results are given in Table 3. Because there are seven distance measures and seven clustering methods, after all the possible combinations, there are 147 results given in Table 3.

3 Results and discussion

According to Table 3, when all the clustering methods and distance measures are examined, mixed results range from one to seven. Related with this result, the Rand's C statistics,

Table 2 Clustering results of regions according to all distance measures and clustering methods

Distance measure	Clustering method	CA M	CA A	CA MA	CA EA	CA SEA	CA BS	M A	M MA	M EA	M SEA	M BS	A MA	A EA	A SEA	A BS	MA EA	MA SEA	MA BS	EA SEA	EA BS	SEA BS
Squared Euclidean distance	Between	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Nearest	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Within	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Furthest	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Centroid	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Median	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Ward's	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
Cosine	Between	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Nearest	X	X	X	X	X	X	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	✓	✓
	Within	✓	✓	✓	X	X	X	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Furthest	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Centroid	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Median	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Ward's	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
Pearson correlation	Between	X	X	X	✓	X	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	X	X
	Nearest	X	X	X	✓	X	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	✓	X
	Within	X	X	X	✓	X	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	✓	X
	Furthest	X	X	X	✓	X	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	✓	X
	Centroid	X	X	X	✓	X	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	✓	X
	Median	X	X	X	✓	X	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	✓	X
	Ward's	X	X	X	✓	X	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	✓	X
Customized	Between	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Nearest	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Within	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Furthest	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Centroid	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Median	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Ward's	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X

Table 2 (*Continued*)

Distance measure	Clustering method	CA M	CA A	CA MA	CA EA	CA SEA	CA BS	M A	M MA	M EA	M SEA	M BS	A MA	A EA	A SEA	A BS	MA EA	MA SEA	MA BS	EA SEA	EA BS	SEA BS
Minkowski	Between	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Nearest	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Within	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Furthest	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Centroid	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Median	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Ward's	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
Block	Between	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Nearest	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Within	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Furthest	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Centroid	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Median	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Ward's	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
Chebychev	Between	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Nearest	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Within	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Furthest	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X
	Centroid	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Median	✓	✓	✓	X	X	✓	✓	✓	X	X	✓	✓	X	X	✓	X	X	✓	X	X	X
	Ward's	X	X	X	X	X	✓	✓	✓	X	X	X	✓	X	X	X	X	X	X	✓	X	X

The symbol ✓ shows that the regions are in the same cluster and X shows that the regions are in a different cluster.

Table 3 Comparisons of clustering methods and Rand's C statistics

Clustering method	Distance measure							
	Squared Euclidean distance				Cosine			
	Together in both	Separate in both	Mixed	Rand's C	Together in both	Separate in both	Mixed	Rand's C
Between-Nearest	4	10	7	0.67	4	14	3	0.86
Between-Within	5	16	0	1.00	4	13	4	0.81
Between-Furthest	5	16	0	1.00	5	16	0	1.00
Between-Centroid	5	16	0	1.00	5	16	0	1.00
Between-Median	5	16	0	1.00	5	16	0	1.00
Between-Ward's	5	16	0	1.00	5	16	0	1.00
Nearest-Within	4	10	7	0.67	4	12	5	0.76
Nearest-Furthest	4	10	7	0.67	4	14	3	0.86
Nearest-Centroid	4	10	7	0.67	4	14	3	0.86
Nearest-Median	4	10	7	0.67	4	14	3	0.86
Nearest-Ward's	4	10	7	0.67	4	14	3	0.86
Within-Furthest	5	16	0	1.00	4	13	4	0.81
Within-Centroid	5	16	0	1.00	4	13	4	0.81
Within-Median	5	16	0	1.00	4	13	4	0.81
Within-Ward's	5	16	0	1.00	4	13	4	0.81
Furthest-Centroid	5	16	0	1.00	5	16	0	1.00
Furthest-Median	5	16	0	1.00	5	16	0	1.00
Furthest-Ward's	5	16	0	1.00	5	16	0	1.00
Centroid-Median	5	16	0	1.00	5	16	0	1.00
Centroid-Ward's	5	16	0	1.00	5	16	0	1.00
Median-Ward's	5	16	0	1.00	5	16	0	1.00

Clustering method	Distance measure							
	Pearson correlation				Customized			
	Together in both	Separate in both	Mixed	Rand's C	Together in both	Separate in both	Mixed	Rand's C
Between-Nearest	4	16	1	0.95	4	10	7	0.67
Between-Within	4	16	1	0.95	5	16	0	1.00
Between-Furthest	4	16	1	0.95	5	16	0	1.00
Between-Centroid	4	16	1	0.95	4	10	7	0.67
Between-Median	4	16	1	0.95	5	16	0	1.00
Between-Ward's	4	16	1	0.95	5	16	0	1.00
Nearest-Within	5	16	0	1.00	4	10	7	0.67
Nearest-Furthest	5	16	0	1.00	4	10	7	0.67
Nearest-Centroid	5	16	0	1.00	10	11	0	1.00
Nearest-Median	5	16	0	1.00	4	10	7	0.67
Nearest-Ward's	5	16	0	1.00	4	10	7	0.67
Within-Furthest	5	16	0	1.00	5	16	0	1.00
Within-Centroid	5	16	0	1.00	4	10	7	0.67
Within-Median	5	16	0	1.00	5	16	0	1.00
Within-Ward's	5	16	0	1.00	5	16	0	1.00
Furthest-Centroid	5	16	0	1.00	4	10	7	0.67
Furthest-Median	5	16	0	1.00	5	16	0	1.00
Furthest-Ward's	5	16	0	1.00	5	16	0	1.00
Centroid-Median	5	16	0	1.00	4	10	7	0.67
Centroid-Ward's	5	16	0	1.00	4	10	7	0.67
Median-Ward's	5	16	0	1.00	5	16	0	1.00

which show the agreement of the distance measures, are 0.67, 0.76, 0.81, 0.86, 0.95 and 1 respectively.

While the distance measure is 'Squared Euclidean', the *Nearest* clustering method is the worst method of all. If the distance measure is considered as 'Cosine', the clustering methods *Nearest* and *Within* perform a worse result than the other methods. While some consider 'Pearson correlation' as a distance measure in hierarchical clustering analysis, the

Table 3 (*Continued*)

Clustering method	Distance measure							
	Minkowski				Block			
	Together in both	Separate in both	Mixed	Rand's C	Together in both	Separate in both	Mixed	Rand's C
Between-Nearest	4	10	7	0.67	4	10	7	0.67
Between-Within	5	16	0	1.00	5	16	0	1.00
Between-Furthest	5	16	0	1.00	5	16	0	1.00
Between-Centroid	4	10	7	0.67	5	16	0	1.00
Between-Median	5	16	0	1.00	5	16	0	1.00
Between-Ward's	5	16	0	1.00	5	16	0	1.00
Nearest-Within	4	10	7	0.67	4	10	7	0.67
Nearest-Furthest	4	10	7	0.67	4	10	7	0.67
Nearest-Centroid	10	11	0	1.00	4	10	7	0.67
Nearest-Median	4	10	7	0.67	4	10	7	0.67
Nearest-Ward's	4	10	7	0.67	4	10	7	0.67
Within-Furthest	5	16	0	1.00	5	16	0	1.00
Within-Centroid	4	10	7	0.67	5	16	0	1.00
Within-Median	5	16	0	1.00	5	16	0	1.00
Within-Ward's	5	16	0	1.00	5	16	0	1.00
Furthest-Centroid	4	10	7	0.67	5	16	0	1.00
Furthest-Median	5	16	0	1.00	5	16	0	1.00
Furthest-Ward's	5	16	0	1.00	5	16	0	1.00
Centroid-Median	4	10	7	0.67	5	16	0	1.00
Centroid-Ward's	4	10	7	0.67	5	16	0	1.00
Median-Ward's	5	16	0	1.00	5	16	0	1.00
Within-Median	4	10	7	0.67	4	10	7	0.67
Within-Ward's	5	16	0	1.00	5	16	0	1.00
Furthest-Centroid	4	10	7	0.67	5	16	0	1.00
Furthest-Median	5	16	0	1.00	5	16	0	1.00
Furthest-Ward's	5	16	0	1.00	5	16	0	1.00
Centroid-Median	4	10	7	0.67	5	16	0	1.00
Centroid-Ward's	4	10	7	0.67	5	16	0	1.00
Median-Ward's	5	16	0	1.00	5	16	0	1.00

Clustering method	Distance measure			
	Chebychev			
	Together in both	Separate in both	Mixed	Rand's C
Between-Nearest	4	10	7	0.67
Between-Within	4	10	7	0.67
Between-Furthest	5	16	0	1.00
Between-Centroid	4	10	7	0.67
Between-Median	4	10	7	0.67
Between-Ward's	5	16	0	1.00
Nearest-Within	10	11	0	1.00
Nearest-Furthest	4	10	7	0.67
Nearest-Centroid	10	11	0	1.00
Nearest-Median	10	11	0	1.00
Nearest-Ward's	4	10	7	0.67
Within-Furthest	4	10	7	0.67
Within-Centroid	10	11	0	1.00
Within-Median	10	11	0	1.00
Within-Ward's	4	10	7	0.67
Furthest-Centroid	4	10	7	0.67
Furthest-Median	4	10	7	0.67
Furthest-Ward's	5	16	0	1.00
Centroid-Median	10	11	0	1.00
Centroid-Ward's	4	10	7	0.67
Median-Ward's	4	10	7	0.67

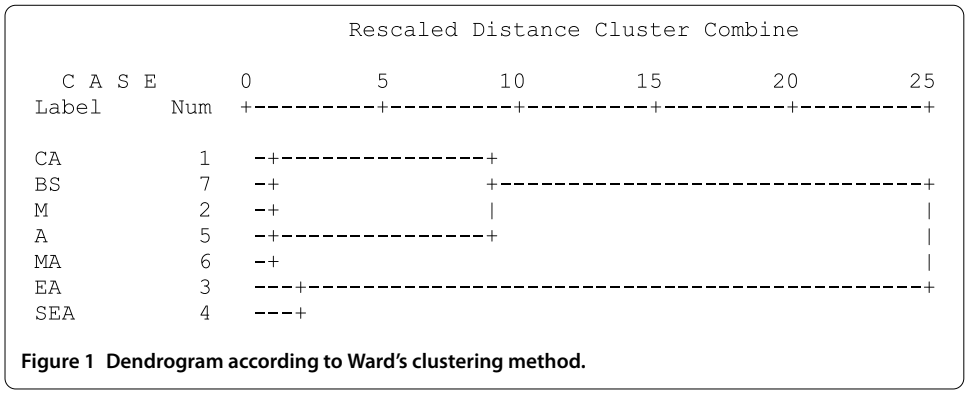

Figure 1 Dendrogram according to Ward's clustering method.

clustering method *Between* is not as suitable as other clustering methods. As it can be seen from Table 2, when the distance measure is considered as 'Minkowsky', the results of clustering methods *Nearest* and *Centroid* according to Rand's statistics are worse than in all other methods. The *Nearest* clustering method also shows the worst performance for the distance measure 'Block'. If the distance measure is considered as 'Customized', Rand's C statistics show that *Nearest* and *Centroid* clustering methods give the worst performances.

For the results of 'Chebychev' distance measure at least for one comparison, all of the clustering results vary for all clustering methods. So, it is really hard to say that any of the clustering methods show better performance than the others.

With respect to these results mentioned above, Ward's hierarchical clustering method is applied to the data set and the results of the analysis are also given in Figure 1.

According to the dendrogram given in Figure 1, at the first stage of the analysis, while Central Anatolia and Blacksea regions join the same cluster, Mediterranean, Aegean and Marmara regions join the other cluster. They connect to each other at the third stage. At the second stage, East Anatolia and South Eastern Anatolia regions join the same cluster and they combine with the other two clusters at the final stage according to Ward's hierarchical clustering method.

4 Conclusion

The earlier studies on comparing the clustering methods also confirm the results of this study. For example, in their study Kuiper and Fisher [10] compared six hierarchical clustering procedures. They used the Rand statistics and, according to their results, Ward's method was best of all. Blashfield [11] used Cohen's statistics to measure the accuracy of the clustering methods, and according to his results, Ward's method performed significantly better than the other clustering procedures. Hands and Everitt [12] also compared five hierarchical clustering techniques, and they found that Ward's method was the better overall than other hierarchical methods. According to Milligan and Cooper [13], Ward's method gave the best overall recovery. And in their study, Ferreira and Hitchcock [6] compared the performance of four major hierarchical methods according to Rand's criteria; and as a result, Ward's method was usually the best.

When there is a clustering problem, the researcher must be aware that choosing a wrong clustering method and distance measure may significantly affect the results of the analysis. For all the results given in related tables in this study, one can consider applying *Ward's* or *Median* clustering methods and keep away from applying the *Nearest* clustering method

for all distance measures to perform a hierarchical clustering analysis to obtain better results according to Rand's C statistics.

Competing interests
The author declares that he has no competing interests.

Acknowledgements
Dedicated to Professor Hari M Srivastava.
I would like to thank Professor Dr. İsmet Doğan for support and all statistical help. He is a lecturer in Afyon Kocatepe University, Faculty of Medicine, Department of Biostatistics, Afyonkarahisar/Turkey.

Received: 15 December 2012 Accepted: 8 March 2013 Published: 2 April 2013

References

1. Rohlf, FJ: Methods of comparing classifications. Ann. Rev. Ecolog. Syst. **5**, 101-103 (1974)
2. Tarpey, T: Clustering functional data. J. Classif. **20**, 93-114 (2003)
3. Tan, P, Steinbach, M, Kumar, V: Introduction to Data Mining. Addison-Wesley, Reading (2005)
4. Hill, T, Lewicki, P: STATISTICS: Methods and Applications. StatSoft, Tulsa (2007)
5. Rand, WM: Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. **66**, 846-850 (1971)
6. Ferreira, L, Hitchcock, DB: A comparison of hierarchical methods for clustering functional data. Commun. Stat., Simul. Comput. **38**, 1925-1949 (2009)
7. Yiğit, A: The studies of dividing Turkey into regions: an assessment for the past and the future developments. In: IV Ulusal Coğrafya Sempozyumu, Avrupa Birliği Sürecindeki Türkiye'de Bölgesel Farklılıklar, Ankara, Turkey, pp. 33-44 (2006)
8. Saraçlı, S, Yılmaz, V, Kaygısız, Z: Examining the geographical dispersion of the human development index in Turkey with multivariabled statistical techniques. In: 3 Ulusal Bilgi Ekonomi ve Yönetim Kongresi, Osmangazi University, Eskişehir, Turkey, 25-26 November (2004)
9. UNDP: Human Development Report. www.undp.org/hdro (2000)
10. Kuiper, FK, Fisher, LA: A Monte Carlo comparison of six clustering procedures. Biometrics **31**, 777-783 (1975)
11. Blashfield, RK: Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods. Psychol. Bull. **83**, 377-388 (1976)
12. Hands, S, Everitt, B: A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. Multivar. Behav. Res. **22**, 235-243 (1987)
13. Milligan, GW, Cooper, MC: A study of standardization of variables in cluster analysis. J. Classif. **5**, 181-204 (1988)

doi:10.1186/1029-242X-2013-142
Cite this article as: Saraçli: **Performance of Rand's C statistics in clustering analysis: an application to clustering the regions of Turkey.** *Journal of Inequalities and Applications* 2013 **2013**:142.