Working Paper Series

Adaptive Information Systems and Modeling in Economics and Management Science

An Examination Of Indexes For Determining The Number Of Clusters In Binary Data Sets

Andreas Weingessel
Evgenia Dimitriadou
Sara Dolnicar

Working Paper No. 29
January 1999

Working Paper Series

January 1999

SFB
'Adaptive Information Systems and Modelling in Economics and Management Science'

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

This piece of research was supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modelling in Economics and Management Science').

An Examination Of Indexes For Determining The Number Of Clusters In Binary Data Sets

Andreas Weingessel **Evgenia Dimitriadou**
Institut für Statistik, Wahrscheinlichkeitstheorie
und Versicherungsmathematik
Technische Universität Wien, Wiedner Hauptstraße 8–10/1071,
A-1040 Wien, Austria,
Email: *firstname.lastname*@ci.tuwien.ac.at
Sara Dolnicar
Institut für Tourismus und Freizeitwirtschaft
Wirtschaftsuniversität Wien
Augasse 2–6, A-1090 Wien, Austria,
Email: Sara.Dolnicar@wu-wien.ac.at

Abstract

An examination of 14 indexes for determining the number of clusters is conducted on artificial binary data sets being generated according to various design factors. To provide a variety of clustering solutions the data sets are analyzed by different non hierarchical clustering methods. The purpose of the paper is to present the performance and the ability of an index to detect the proper number of clusters in a binary data set under various conditions and different difficulty levels.

Keywords: Number of Clusters, Clustering Indexes, Binary Data, Artificial Data Sets.

1 Introduction

Clustering is the partitioning of a set of objects into groups so that objects within a group are "similar" and objects in different groups are "dissimilar". Thus, the purpose of clustering is to identify "natural" structures in a data set. In real life clustering situations a researcher in the beginning is confronted with crucial decisions like choosing the appropriate clustering method and selecting the number of clusters in the final solution. The latter is considered to be an unsolved problem of great significance as the result and the success of a research is depending on this decision. Numerous strategies have been proposed for finding the right number of clusters. In the literature such stopping criteria have a long history. Consequently, they have been expressed in several forms according to the need of every task they are about to be applied to. These criteria, also called indexes, can be generally categorized as dependent or combined with a clustering algorithm (e.g. Xu, 1997; Hall et al., 1973) ones, and also as independent, which can be applied to any algorithm. The latter category can be divided into two major ones. The first, external criterion analysis (see Milligan, 1981), uses an independently obtained partition that must be specified a priori or obtained from a clustering of a separate data set. Its main disadvantage is that in empirical data sets a priori information can not be always obtained. The second category, internal criterion analysis (see Milligan & Cooper, 1985), uses the information obtained from within the clustering process. The internal criterion measures represent the "goodness" of fit between the input data and the resulting cluster partition.

Monte Carlo evaluations of these criterion measures have been conducted by researchers in order to analyze and compare their performance (see Milligan & Cooper, 1985; Milligan, 1981, 1980).

In the analysis of real world data, one often encounters binary data, like for example in the analysis of psychological questionnaires. Imagine a questionnaire where respondents are asked to indicate their agreement with statements concerning their daily habits, e.g., "I want to rest and relax" or "I am getting stressed at work". The data generated for this piece of research are based on scenarios presuming certain answer tendencies and typical answer profiles within people clusters. Answer prototypes are constructed by defining the cluster preferences, i.e., relative frequencies of agreement for each variable and subsample. The term "variable" denotes directly observed answers in binary format, where 1 indicates that a person answered a certain question with "Yes".

Only few literature exists based on dealing and clustering with binary data sets. Binary sets are by origin a special category of data. Their simple structure makes certain computations more simple and faster. For these reasons it is worthy to work specifically and analyze the performance of the clustering indexes applied on such kind of data. The purpose of this paper is to perform this comparison of the indexes applied on such sets.

The paper is organized as follows. In Section 2, Section 3, and Section 4 there is a description of the artificial binary scenarios, the clustering algorithms used for the simulations, and the analyzing indexes, respectively. Section 5 demonstrates our experimental results and some comments on them. A summary of the paper is given in Section 6.

2 Binary Data Scenarios

12-dimensional binary data sets are used in the experiments. Each variable models a "yes/no" statement in a questionnaire. A structure in the data is introduced by creating 6 types with different answering behavior. Table 1 depicts the basic scenario. An "H" entry in the table means that a person has a high probability (0.8 in the experiments) to answer "yes" to this particular question, whereas "L" denotes a low probability (0.2) for a "yes". As can be seen in the table, the variables are grouped into 4 groups of several indicator variables each. Within these groups the answer probabilities are the same for each type. The variables are modeled to be independent for each type, a dependency is only added by the mixture of the types.

There are 3 design factors being varied in the experiments.

	G1			G2			G3			G4			
Type	I1	I2	I3	I1	I2	I3	I1	I2	I3	I1	I2	I3	n
1	H	H	H	H	H	H	L	L	L	L	L	L	1000
2	L	L	L	L	L	L	H	H	H	H	H	H	1000
3	L	L	L	H	H	H	H	H	H	L	L	L	1000
4	H	H	H	L	L	L	L	L	L	H	H	H	1000
5	L	L	L	H	H	H	L	L	L	H	H	H	1000
6	H	H	H	L	L	L	H	H	H	L	L	L	1000

Table 1: The Basic Scenario 1a6

Number of Indicators One level has 3 indicator variables for each group, the other one 5, 4, 2, and 1 indicator variables for the 4 groups.

Size of Clusters There are experiments with equal cluster sizes (1000 for each type) and with unequal cluster sizes (1000–300–700–3000–500–500 data points in the types 1–6) respectively.

Number of Clusters There are experiments with 3,4,5 and 6 clusters (as in Table 1). The 3 types of the 3 cluster scenarios correspond to the first 3 types from all the 6 cluster scenarios. Likewise, the 4 and 5 cluster scenarios are generated.

All three design factors are crossed, yielding 16 different data sets. For the description of the results the following naming convention is chosen for the data sets. Each scenario is described by XY(Z), where X is "1" for scenarios with an equal number of indicators, and "2" for scenarios with an unequal number. Y equals "a" for scenarios with equal cluster sizes and "b" for scenarios with unequal cluster sizes. Z finally gives the number of clusters (3, 4, 5 or 6) and is omitted if it is not necessary. In Table 2 an overview of the names is given.

	Equal Cluster Size	Unequal Cluster Size
Equal Nr. of Indicators	Scenario 1a	Scenario 1b
Unequal Nr. of Indicators	Scenario 2a	Scenario 2b

Table 2: The Names of the Cluster Scenarios

Note that, as in real-world situations, the types are not clearly separated, but there is an overlapping cluster structure in the data sets. Scenario 1a, for example, has a Bayes classification rate of 83%, that is only 83% of the data points can be assigned to the right cluster even if the cluster structure is a priori completely known.

3 Algorithm

Two algorithms, namely the $k$means (also known as LBG algorithm, see Linde et al. (1980)) and hard competitive learning (see Xu et al., 1993), are used for the experiments in order to support us with various clustering solutions, preventing in this way the dependency of the solutions on the clustering method.

The decision in favor of these partitioning algorithms was made under the following considerations. Due to computer development, huge data sets are often available. In the case of hierarchical methods, it is required to compute all pairwise distances. For n data points, the number of pairwise distances equals $n(n-1)/2$. For $n = 6000$ this gives already approximately 18 million distances, which makes computation infeasible.

In Dolnicar et al. (1998) it is shown (for the 6 cluster scenarios) that the $k$means and hard competitive learning algorithms are able to find the correct cluster structure, but they sometimes stick in a local minimum due to the random initialization of the cluster centers. To overcome these instabilities the following experimental setup is chosen. Cluster solutions are computed starting with 2 cluster centers up to 13 centers. The range was chosen so that it contains twice the number of clusters that there are in the data sets, so that the solution where every existing cluster might be split into two parts is still contained in the range of considered centers. For each of the different number of clusters the algorithms are repeated 10 times. The results with minimum within-sum-of-square are chosen and used to compute the index for this particular solution. From this vector the number of clusters is found as described in the next section. To ensure the stability of the results the above process is repeated 100 times for every scenario.

4 Indexes

In this paper 14 different indexes are compared. These indexes represent 14 internal criteria that can be computed independently from the clustering algorithm. Indexes applied only on hierarchical clustering methods (see Aldenderfer & Blashfield, 1996; Milligan & Cooper, 1985) are not used, because these methods are not suitable for large data sets (i.e., one is confronted with memory and time problems). Moreover, indexes computing measures similar to the ones from the hierarchical methods, (i.e., measures using pairwise distances; e.g. Gamma measure (Baker & Hubert, 1975), also the Point Biserial measure (Milligan, 1981)) are excluded from the study for the same reason. There are other criteria and indexes for determining the number of clusters being fuzzy (e.g. Yang & Yu, 1990), heuristic by depending on graphical methods (see for example Arratia & Lander, 1990; Andrews, 1972; Gengerelli, 1963), being valid for data sets belonging to specific distributions (e.g., Likelihood Ratio measure; Wolfe, 1970) or being applicable under certain conditions (e.g. the Cubic clustering criterion; Sarle, 1983). For achieving an objective result in this research, they are omitted.

After computing a particular index for a range of cluster numbers, one has to decide which cluster number to choose. In the simplest case one can consider as a solution that number of cluster where the index has its maximum (or minimum) value. However, this simple rule does not work in most cases. In the literature the index is often chosen in that way that the curve of the index values is plotted and the user chooses a particular number by visual inspection, often where the curve has an "elbow", i.e., a positive or negative "jump" of the index curve, or a local peak. In this paper the use of such a subjective criteria is omitted, but an objective one is being computed. Therefore, besides looking at the maximum (or minimum) value $\max_n i_n$ (where n is the number of clusters and i_n the index value for n clusters) of the index the following statistics were considered. Here, they are described for an index where maximum values are of interest.

- The maximum difference to the cluster at the left side $(\max_n(i_n - i_{n-1}))$. This is the part where the curve has its maximum increase.
- The maximum difference to the cluster at the right side $(\max_n(i_n - i_{n+1}))$. This is the part where the curve has its maximum decrease.
- The minimum value of the second differences
$(\min_n ((i_{n+1} - i_n) - (i_n - i_{n-1})))$.
This measures a positive "elbow".

The decision of which measure to use is made after computing all of them for all the data sets and taking that one which performed best on average.

The description of the indexes is categorized into 3 groups, based on the statistics mainly used to compute them.

The first group is based on the sum of squares within (SSW) and between (SSB) the clusters. These statistics measure the dispersion of the data points in a cluster and between the clusters respectively. These indexes are:

Calinski and Harabasz (see Calinski & Harabasz, 1974)
$(SSB/(k-1))/(SSW/(n-k))$, where n is the number of data points and k is the number of clusters.
The minimum value of the second differences is taken as the proposed number of clusters.

Hartigan (see Hartigan, 1975)
$\log(SSB/SSW)$
The minimum value of the second differences is taken as the proposed number of clusters.

Ratkowsky and Lance (see Ratkowsky & Lance, 1978)
$\text{mean}(\sqrt{(varSSB/varSST)})$, where $varSSB$ stands for the SSB for every variable and $varSST$ for the total sum of squares for every variable.
The maximum difference to the cluster at the right side is taken as the proposed number of clusters.

Ball and Hall (see Ball & Hall, 1965)
SSW/k, where k is the number of clusters.
The maximum value of the second differences to the cluster at the left side is taken as the proposed number of clusters.

The second group is based on the statistics of T, i.e., the scatter matrix of the data points, and W, which is the sum of the scatter matrices in every group. These indexes are:

Scott and Symons (see Scott & Symons, 1971)
$n \log(|T|/|W|)$, where n is the number of data points and $|\cdot|$ stands for the determinant of a matrix.
The maximum difference to the cluster at the left side is taken as the proposed number of clusters.

Marriot (see Marriot, 1971)
$k^2|W|$, where k is the number of clusters.
The maximum value of the second differences is taken as the proposed number of clusters.

TraceCovW (see Milligan & Cooper, 1985)
$\text{TraceCov}W$
The minimum value of the second differences to the cluster at the left side was taken as the proposed number of clusters.

TraceW (see Edwards & Cavalli-Sforza, 1965; Friedman & Rubin, 1967; Orloci, 1967; Fukunaga & Koontz, 1970)
$\text{Trace}W$
The maximum value of the second differences is taken as the proposed number of clusters.

TraceW$^{(-1)}$B (see Friedman & Rubin, 1967)
$\text{Trace}W^{(-1)}B$, where B is the scatter matrix of the cluster centers.
The maximum difference to the cluster at the left side is taken as the proposed number of clusters.

|T|/|W| (see Friedman & Rubin, 1967)
$|T|/|W|$
The minimum value of the second differences is taken as the proposed number of clusters.

The third group consists of four algorithms not belonging to the previous ones and not having anything in common.

Hubert and Levin (C-Index) (see Hubert & Levin, 1976)
The C-Index is a cluster similarity measure expressed as:
$[d_w - \min(d_w)]/[\max(d_w) - \min(d_w)]$, where d_w is the sum of all n_d within cluster distances, $\min(d_w)$ is the sum of the n_d smallest pairwise distances in the data set, and $\max(d_w)$ is the sum of the n_d biggest pairwise distances. In order to compute the C-Index all pairwise distances in the data set have to be computed and stored. In this case of binary data, the storage of the distances is creating no problems since there are only a few possible distances. However, the computation of all distances can make this index prohibitive for large data sets.
The maximum value of the second differences is taken as the proposed number of clusters.

Davies and Bouldin (DB-Index) (see Davies & Bouldin, 1979)
$R = (1/n)\sum_{i=1}^{n}(R_i)$ where R_i stands for the maximum value of R_{ij} for $i \neq j$, and R_{ij} for $R_{ij} = (SSW_i + SSW_j)/DC_{ij}$, where DC_{ij} is the distance between the centers of two clusters i, j.
The minimum value is taken as the proposed number of clusters.

Likelihood (see Wedel & Kamakura, 1998)
Under the assumption of independence of the variables within a cluster, a cluster solution can be regarded as a mixture model for the data, where the cluster centers give the probabilities for each variable to be 1. Therefore, the negative Log-likelihood can be computed and used as a quantity

measure for a cluster solution. Note that the assumptions for applying special penalty terms, like in AIC or BIC, are not fulfilled in this model, and also they show no effect for these data sets.
The maximum value of the second differences is taken as the proposed number of clusters.

SSI (see Dolnicar et al., 1999)
This "Simple Structure Index" combines three elements which influence the interpretability of a solution, i.e., the maximum difference of each variable between the clusters, the sizes of the most contrasting clusters and the deviation of a variable in the cluster centers compared to its overall mean. These three elements are multiplicatively combined and normalized to give a value between 0 and 1.
The maximum value is taken as the proposed number of clusters.

5 Results

5.1 Description

Tables 3–18 describe the results on the artificial data sets, Table 19 provides an aggregated evaluation of the indexes. The bold printed column 0 shows the number of times the index finds the right number of clusters. The columns -1 and $+1$ give the number of results deviating from the correct solution by 1. The column -2 ($+2$) shows the number of times where the index chooses a number of clusters which is at least 2 less (more) than the right number.

To give a measure of how clear the decision of the index is, the absolute difference between the winning number and the second number is computed and scaled by the range of the numbers for all clusters. The mean value and standard deviation of this new measure in 100 runs (both multiplied by 100) is given in the columns *meansure* and *sdsure*. If the index finds the right number of clusters (almost) every time, it is optimal for the *meansure* to be high, if the index is wrong, it is favorable to be small to indicate an unsure decision. In all tables the indexes are ranked by performance with the best indexes appearing first.

5.2 Comments on Results

For the scenarios with 6 clusters, the basic scenario 1a is, as expected, the simplest one. 7 indexes (see Table 15) perform perfectly on this data set, another 2 find the right number in a majority of all runs. The second simplest scenario is the one where the number of indicators variables are not equal (Scenario 2a, Table 17). 4 indexes still find the right number in more than 90% of all runs, 3 of them (Calinski, Ratkowsky, and DB-Index) performing well in scenario 1a. C-Index is following having a performance of about 85%, all the others render unsatisfactory results. Changing the sizes of the clusters (Scenario 1b and 2b, Tables 16 & 18) makes it very difficult for the indexes to find the right number, no index finds the right number in more than 15% of all replications, except for C-Index with a success rate of 63% in Scenario 1b. Most indexes perform similar for the 4 and 5 clusters scenarios. Since the clusters are fewer, and the difficulty of clustering is diminishing, the overall performance is better. Furthermore, certain indexes demonstrate superior results for specific data situations. More specifically, Trace$W^{-1}B$ and TraceCovW win the first places in the tables for the scenarios 1b and 2b, whereas Calinski, Ratkowsky, TraceW and Likelihood lead to right results only for scenarios 1a and 2a. C-Index performs most satisfactory in both cases.

The results for the 3 cluster data sets are generally better, most of the indexes find the right number of clusters in at least 80% of all runs. This might be due to the fact that their structure is less complex. However, some indexes (Hartigan, Ball) seem to systematically propose a small number of clusters (less or equal than 4), thus performing perfectly for 3 clusters just for random reasons (Ball, Hartigan, Likelihood indexes), whereas others function in the opposite way ($|T|/|W|$, TraceCovW indexes). Moreover, for the common indexes described and analyzed in Milligan & Cooper (1985), it is clear that they do not perform equally, due to the fact that the data in this paper are binary, overlapping, and the difficulty level appears to be higher (e.g. Calinski index does not appear to be good but only in scenarios 1a and 2a, that are considered to be the easiest ones, where in Milligan & Cooper (1985) it was proved to be the best performing index). Finally, in Table 19 the overall results of the performance of the indexes are presented with the C-Index performing best, followed by Ratkowsky and TraceW indexes.

6 Summary

In this paper the performance of various indexes for determining the number of clusters in a binary data set is analyzed. To ensure that the right number of clusters is known, only artificial scenarios, designed to simulate the difficulties of real-world data, are used. Three design factors, namely the size of clusters, the number of duplicators and the number of clusters, are crossed yielding 16 different artificial data sets. For the evaluation of the performance of the indexes, *k*means and hard competitive learning methods are applied in 100 runs for every scenario so as to overcome instabilities imposed by the clustering algorithms. The selection of the number of clusters based on the indexes values for the different number of clusters is done in an automatic way.The artificial data sets in this paper are being generated to resemble real-world data. Thus, the analysis of the indexes performance helps a researcher to choose an index which is fitted to his problem, concerning the expected number or size of the clusters.

Acknowledgments

This piece of research was supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modeling in Economics and Management Science').

References

Aldenderfer, M. S. & Blashfield, R. K. (1996). *Cluster Analysis*, chap. Determining the number of clusters. Sage.

Andrews, D. F. (1972). Plots of high-dimensional data. *Biometrics*, **28**, 125–136.

Arratia, R. & Lander, E. S. (1990). The distribution of clusters in random graphs. *Advances in Applied Mathematics*, **11**, 36–48.

Baker, F. B. & Hubert, L. J. (1975). Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association*, **70**, 31–38.

Ball, G. H. & Hall, D. J. (1965). *ISODATA, A novel method of data analysis and pattern classification*. Tech. Rep. NTIS No.AD 699616, Stanford Research Institute, Menlo Park.

Calinski, R. B. & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, **3**, 1–27.

Davies, D. L. & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**, 224–227.

Dolnicar, S., Grabler, K., & Mazanec, J. (1999). A tale of three cities: Perceptual charting for analyzing destination images. In Woodside (ed.), *Consumer Psychology of Tourism, Hospitality and Leisure*, London. CAB International. to appear.

Dolnicar, S., Leisch, F., Weingessel, A., Buchta, C., & Dimitriadou, E. (1998). *A Comparison of Several Cluster Algorithms on Artificial Binary Data Scenarios from Tourism Marketing*. Working Paper 7, SFB “Adaptive Information Systems and Modeling in Economics and Management Science”, http://www.wu-wien.ac.at/am.

Edwards, A. W. F. & Cavalli-Sforza, L. (1965). A method for cluster analysis. *Biometrics*, **21**, 362–375.

Friedman, H. P. & Rubin, J. (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association*, **62**, 1159–1178.

Fukunaga, K. & Koontz, W. L. G. (1970). A criterion and an algorithm for grouping data. *IEEE Transactions on Computers*, **C–19**, 917–923.

Gengerelli, J. A. (1963). A method for detecting subgroups in a population and specifying their membership list. *Journal of Psychology*, **5**, 457–468.

Hall, D. J., Duda, R. O., Huffman, D. A., & Wolf, E. E. (1973). *Development of new pattern recognotion methods*. Tech. Rep. NTIS No. AD 7726141, Aerospace Research Laboratories, Los Angeles.

Hartigan, J. A. (1975). *Clustering algorithms*. New York: Willey.

Hubert, L. J. & Levin, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Phycological Bulletin*, **83**, 1072–1080.

Linde, Y., Buzo, A., & Gray, R. M. (1980). An algorithm for vector quantizer design. *IEEE Transactions on Communications*, **COM-28**(1), 84–95.

Marriot, F. H. C. (1971). Practical problems in a method of cluster analysis. *Biometrics*, **27**, 501–514.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clutering algorithms. *Psychometrika*, **45**(3), 325–342.

Milligan, G. W. (1981). A monte carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika*, **46**(2), 187–199.

Milligan, G. W. & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, **50**(2), 159–179.

Orloci, L. (1967). An agglomerative method of classification of plant communities. *Journal of Ecology*, **55**, 193–206.

Ratkowsky, D. A. & Lance, G. N. (1978). A criterion for determining the number of groups in a classification. *Australian Computer Journal*, **10**, 115–117.

Sarle, W. S. (1983). *Cubic clustering criterion.* Tech. Rep. A-108, SAS Institute, Cary, N.C.

Scott, A. J. & Symons, M. J. (1971). Clustering methods based on likelihood ratio criteria. *Biometrics*, **27**, 387–397.

Wedel, M. & Kamakura, W. A. (1998). *Marketing Segmentation. Conceptual and Methodological Foundations*, chap. Mixture Models, pp. 89–92. Boston Dordrecht London: Kluwer Academic.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, **5**, 329–350.

Xu, L. (1997). Bayesian ying-yang machine, clustering and number of clusters. *Pattern Recognition Letters*, **18**, 1167–1178.

Xu, L., Krzyzak, A., & Oja, E. (1993). Rival penalized competitive learning for clustering analysis RBF net and curve detection. *IEEE Transactions on Neural Networks*, **4**(4), 636–649.

Yang, M.-S. & Yu, K. F. (1990). On stochastic convergence theorems for the fuzzy c-means clustering procedure. *International Journal of General Systems*, **16**, 397–411.

Methods	-1	**0**	$+1$	$+2$	meansure	sdsure
Hartigan	0	**200**	0	0	79.66	2.38
TraceW	0	**200**	0	0	76.70	3.06
C-Index	0	**200**	0	0	67.95	12.61
Likelihood	0	**200**	0	0	64.28	5.85
Calinski	0	**200**	0	0	54.46	1.97
Ball	0	**200**	0	0	49.60	0.47
Ratkowsky	0	**200**	0	0	22.47	1.99
DB-Index	0	**200**	0	0	9.87	1.43
SSI	0	**185**	1	14	22.17	9.85
Scott	0	**177**	0	23	36.03	16.42
Marriot	0	**131**	0	69	22.54	17.78
TraceW^{-1}B	0	**110**	0	90	22.42	14.24
TraceCovW	0	**0**	0	200	2.85	2.55
$\|T\|/\|W\|$	0	**0**	0	200	45.64	32.93

Table 3: Scenario 1a 3 Cluster

Methods	-1	**0**	$+1$	$+2$	meansure	sdsure
Hartigan	0	**200**	0	0	63.01	3.38
C-Index	0	**200**	0	0	58.13	16.29
TraceW	0	**200**	0	0	57.91	4.45
Ball	0	**200**	0	0	43.01	0.52
Likelihood	0	**200**	0	0	36.57	7.04
Ratkowsky	0	**200**	0	0	16.34	5.19
Scott	0	**143**	0	57	14.50	8.54
DB-Index	0	**96**	0	104	10.12	6.77
TraceW^{-1}B	0	**7**	4	189	22.05	14.34
Marriot	0	**0**	3	197	36.34	22.28
Calinski	0	**0**	0	200	2.23	1.98
TraceCovW	0	**0**	0	200	3.85	3.15
SSI	0	**0**	0	200	20.48	15.27
$\|T\|/\|W\|$	0	**0**	0	200	48.85	30.22

Table 4: Scenario 1b 3 Cluster

Methods	-1	**0**	$+1$	$+2$	meansure	sdsure
Hartigan	0	**200**	0	0	79.40	3.12
TraceW	0	**200**	0	0	76.92	3.84
Likelihood	0	**200**	0	0	62.82	5.67
Ball	0	**200**	0	0	49.64	0.33
Calinski	0	**200**	0	0	33.29	1.91
Ratkowsky	0	**200**	0	0	26.54	1.87
SSI	0	**192**	0	8	22.34	10.18
C-Index	0	**185**	0	15	34.39	19.82
Scott	0	**181**	0	19	37.50	17.35
Marriot	0	**146**	0	54	28.80	20.95
TraceW^{-1}B	0	**102**	0	98	20.93	14.90
TraceCovW	0	**0**	1	199	3.09	2.67
DB-Index	200	**0**	0	0	2.38	0.26
$\|T\|/\|W\|$	0	**0**	0	200	48.60	31.32

Table 5: Scenario 2a 3 Cluster

Methods	-1	**0**	$+1$	$+2$	meansure	sdsure
C-Index	0	**200**	0	0	61.57	17.53
Hartigan	0	**200**	0	0	55.46	3.10
TraceW	0	**200**	0	0	50.87	4.26
Ball	0	**200**	0	0	40.47	0.37
Likelihood	0	**200**	0	0	23.78	7.44
Ratkowsky	0	**196**	4	0	15.27	7.37
DB-Index	0	**167**	0	33	4.01	1.53
Scott	0	**118**	0	82	12.08	7.13
TraceW^{-1}B	0	**1**	15	184	21.28	14.67
Marriot	0	**0**	7	193	35.04	22.95
Calinski	0	**0**	0	200	1.49	1.16
TraceCovW	0	**0**	0	200	6.49	5.51
SSI	0	**0**	0	200	17.12	12.36
$\|T\|/\|W\|$	0	**0**	0	200	51.32	31.51

Table 6: Scenario 2b 3 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
Calinski	0	0	**200**	0	0	74.59	4.21
TraceW	0	0	**200**	0	0	30.47	3.13
Likelihood	0	0	**200**	0	0	19.57	1.96
Ratkowsky	0	0	**200**	0	0	6.23	0.68
Hartigan	0	0	**200**	0	0	5.81	2.40
Scott	0	0	**195**	0	5	48.47	13.82
$\text{TraceW}^{-1}\text{B}$	0	0	**185**	0	15	43.53	18.33
Marriot	0	0	**185**	0	15	34.77	12.93
DB-Index	0	25	**175**	0	0	2.03	1.38
C-Index	0	79	**119**	0	2	5.06	4.41
SSI	0	0	**74**	19	107	9.41	6.64
Ball	0	200	**0**	0	0	29.72	0.48
TraceCovW	0	0	**0**	3	197	1.37	1.23
$\|T\|/\|W\|$	0	0	**0**	0	200	48.04	30.72

Table 7: Scenario 1a 4 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
$\text{TraceW}^{-1}\text{B}$	0		**150**	4	46	17.99	12.36
Scott	0	0	**51**	107	42	13.73	11.06
Marriot	0	0	**42**	45	113	22.98	16.66
TraceCovW	0	0	**14**	13	173	3.35	3.10
C-Index	0	137	**11**	0	52	19.97	13.93
Ratkowsky	0	192	**8**	0	0	11.01	6.90
Ball	0	200	**0**	0	0	39.31	0.40
TraceW	0	200	**0**	0	0	41.14	4.77
Hartigan	0	200	**0**	0	0	48.97	3.06
Likelihood	0	196	**0**	0	4	23.78	12.50
DB-Index	0	3	**0**	0	197	17.11	11.29
Calinski	0	0	**0**	0	200	2.38	1.90
SSI	0	0	**0**	0	200	17.71	13.01
$\|T\|/\|W\|$	0	0	**0**	0	200	44.82	31.20

Table 8: Scenario 1b 4 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
TraceW	0	0	**200**	0	0	59.13	2.01
Likelihood	0	0	**200**	0	0	45.56	3.84
Calinski	0	0	**200**	0	0	43.53	2.13
Hartigan	0	0	**200**	0	0	35.43	1.14
Ratkowsky	0	0	**200**	0	0	7.89	0.61
Scott	0	0	**196**	0	4	52.70	15.21
C-Index	0	0	**194**	0	6	36.53	16.22
Marriot	0	0	**185**	0	15	37.91	14.23
TraceW^{-1}B	0	0	**182**	0	18	45.33	18.97
SSI	0	0	**179**	0	21	13.70	8.78
DB-Index	0	197	**3**	0	0	4.19	0.60
$\|T\|/\|W\|$	0	0	**3**	0	197	44.86	32.36
Ball	0	200	**0**	0	0	25.72	0.16
TraceCovW	0	0	**0**	0	200	2.04	1.85

Table 9: Scenario 2a 4 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
Marriot	0	0	**161**	8	31	30.17	19.35
TraceW^{-1}B	0	1	**115**	12	72	10.88	8.46
Scott	0	94	**104**	0	2	5.48	4.09
TraceCovW	0	0	**10**	3	187	3.56	2.96
Ratkowsky	0	200	**0**	0	0	24.83	3.85
Likelihood	0	200	**0**	0	0	25.60	5.63
Ball	0	200	**0**	0	0	42.75	0.48
TraceW	0	200	**0**	0	0	53.11	4.87
Hartigan	0	200	**0**	0	0	60.68	3.41
Calinski	0	197	**0**	0	3	9.88	3.97
C-Index	0	196	**0**	0	4	49.76	17.72
DB-Index	0	11	**0**	0	189	13.03	10.17
SSI	0	0	**0**	0	200	19.27	13.55
$\|T\|/\|W\|$	0	0	**0**	0	200	43.33	30.95

Table 10: Scenario 2b 4 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
Calinski	0	0	**200**	0	0	96.70	4.77
C-Index	0	0	**200**	0	0	53.62	7.93
TraceW	0	0	**200**	0	0	40.73	2.76
Likelihood	0	0	**200**	0	0	28.94	3.14
DB-Index	0	0	**200**	0	0	13.35	4.03
Ratkowsky	0	0	**200**	0	0	4.73	0.58
Marriot	0	0	**197**	0	3	38.67	12.69
Hartigan	3	0	**197**	0	0	3.63	2.71
$\mathrm{TraceW}^{-1}\mathrm{B}$	0	0	**195**	0	5	46.19	14.81
Scott	0	18	**181**	0	1	14.05	8.83
SSI	0	0	**69**	28	103	8.11	6.47
$\lvert T\rvert/\lvert W\rvert$	0	0	**19**	0	181	30.26	27.35
TraceCovW	0	0	**0**	2	198	1.44	1.32
Ball	200	0	**0**	0	0	32.24	0.82

Table 11: Scenario 1a 5 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
C-Index	31	0	**126**	4	39	7.21	9.02
Marriot	0	16	**16**	71	97	16.36	13.38
$\mathrm{TraceW}^{-1}\mathrm{B}$	0	162	**15**	1	22	22.29	12.26
Scott	66	93	**15**	19	7	2.15	3.26
Ratkowsky	194	2	**4**	0	0	11.00	3.62
TraceCovW	0	0	**3**	9	188	3.85	3.15
Likelihood	199	1	**0**	0	0	20.21	7.49
Calinski	0	0	**0**	0	200	2.34	1.79
DB-Index	0	0	**0**	0	200	15.83	11.13
SSI	0	0	**0**	0	200	19.25	13.09
TraceW	200	0	**0**	0	0	27.45	2.44
Hartigan	200	0	**0**	0	0	36.22	1.22
Ball	200	0	**0**	0	0	36.35	0.13
$\lvert T\rvert/\lvert W\rvert$	0	0	**0**	0	200	42.52	31.52

Table 12: Scenario 1b 5 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
Calinski	0	0	**200**	0	0	18.10	2.31
DB-Index	0	0	**200**	0	0	16.37	2.62
Ratkowsky	0	0	**200**	0	0	6.87	0.98
TraceW	0	0	**200**	0	0	3.99	0.80
C-Index	0	0	**198**	0	2	36.81	13.03
SSI	0	0	**156**	4	40	7.77	5.03
Marriot	0	10	**120**	0	70	8.53	10.97
|T|/|W|	0	0	**2**	0	198	38.81	30.15
Scott	0	199	**0**	0	1	26.28	3.17
TraceW^{-1}B	0	192	**0**	0	8	44.90	16.11
TraceCovW	0	0	**0**	3	197	1.86	1.76
Likelihood	197	0	**0**	0	3	7.22	3.35
Hartigan	200	0	**0**	0	0	16.17	0.61
Ball	200	0	**0**	0	0	34.18	0.04

Table 13: Scenario 2a 5 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
TraceCovW	0	66	**28**	19	87	4.82	4.01
TraceW^{-1}B	67	58	**15**	5	55	7.20	7.45
Marriot	3	136	**14**	4	43	31.59	20.16
C-Index	151	0	**9**	12	28	23.08	13.77
Scott	198	1	**1**	0	0	12.58	7.63
|T|/|W|	0	0	**0**	1	199	41.66	28.57
Calinski	0	0	**0**	0	200	2.89	2.30
DB-Index	1	0	**0**	0	199	11.02	7.46
SSI	0	0	**0**	0	200	17.90	13.30
Likelihood	198	0	**0**	0	2	26.32	7.06
Ball	200	0	**0**	0	0	42.04	0.39
Ratkowsky	200	0	**0**	0	0	48.54	9.39
TraceW	200	0	**0**	0	0	52.63	5.41
Hartigan	200	0	**0**	0	0	59.73	3.93

Table 14: Scenario 2b 5 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
Calinski	0	0	**200**	0	0	71.96	4.56
C-Index	0	0	**200**	0	0	57.78	8.99
TraceW	0	0	**200**	0	0	33.99	3.83
Marriot	0	0	**200**	0	0	31.20	10.40
Likelihood	0	0	**200**	0	0	22.64	4.34
DB-Index	0	0	**200**	0	0	16.64	2.92
Ratkowsky	0	0	**200**	0	0	6.72	1.13
TraceW^{-1}B	0	65	**134**	0	1	11.28	9.24
Scott	0	71	**128**	0	1	9.50	7.26
SSI	4	0	**92**	29	75	7.26	5.76
$\|T\|/\|W\|$	0	0	**76**	0	124	27.52	25.44
Hartigan	191	0	**9**	0	0	5.46	3.17
TraceCovW	0	0	**0**	9	191	1.41	1.30
Ball	200	0	**0**	0	0	32.39	0.84

Table 15: Scenario 1a 6 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
TraceCovW	0	1	**29**	8	162	3.77	3.00
C-Index	185	0	**12**	0	3	19.70	9.33
Marriot	13	34	**10**	26	117	12.30	12.09
Ratkowsky	5	190	**4**	1	0	8.29	5.05
SSI	0	0	**3**	1	196	22.96	16.93
TraceW^{-1}B	0	196	**0**	0	4	30.29	11.99
Calinski	0	5	**0**	2	193	2.40	1.92
Scott	199	1	**0**	0	0	21.63	3.19
DB-Index	0	0	**0**	0	200	9.43	6.69
TraceW	200	0	**0**	0	0	13.84	2.11
Likelihood	196	0	**0**	0	4	15.53	5.05
Hartigan	200	0	**0**	0	0	28.58	1.51
Ball	200	0	**0**	0	0	35.10	0.29
$\|T\|/\|W\|$	2	0	**0**	0	198	36.67	26.69

Table 16: Scenario 1b 6 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
Calinski	0	0	**199**	1	0	13.21	1.15
DB-Index	0	1	**199**	0	0	11.68	3.73
Ratkowsky	0	1	**199**	0	0	9.27	1.65
SSI	0	0	**175**	5	20	11.53	7.03
C-Index	31	0	**168**	1	0	8.30	6.95
$\|T\|/\|W\|$	0	0	**31**	0	169	34.06	27.95
TraceW	189	0	**11**	0	0	1.69	0.83
TraceCovW	14	0	**1**	19	166	1.98	2.36
Likelihood	198	0	**1**	0	1	7.63	1.67
Hartigan	200	0	**0**	0	0	3.40	0.99
Ball	200	0	**0**	0	0	31.07	0.21
Marriot	199	0	**0**	0	1	31.17	4.11
Scott	199	0	**0**	0	1	32.65	3.70
TraceW^{-1}B	199	0	**0**	0	1	49.24	7.57

Table 17: Scenario 2a 6 Cluster

Methods	-2	-1	**0**	$+1$	$+2$	meansure	sdsure
TraceCovW	0	43	**29**	23	105	6.67	5.59
C-Index	109	16	**18**	9	48	18.86	14.34
Marriot	122	20	**7**	13	38	24.40	17.39
TraceW^{-1}B	149	17	**6**	7	21	14.66	9.03
Likelihood	181	4	**1**	2	12	12.87	7.65
Scott	199	0	**1**	0	0	13.23	6.35
Calinski	0	0	**0**	7	193	2.68	2.13
SSI	0	2	**0**	2	196	22.10	15.98
Ratkowsky	197	3	**0**	0	0	13.48	11.51
$\|T\|/\|W\|$	0	0	**0**	3	197	40.13	29.56
DB-Index	0	0	**0**	0	200	11.82	9.04
TraceW	200	0	**0**	0	0	22.59	4.08
Hartigan	200	0	**0**	0	0	32.82	3.53
Ball	200	0	**0**	0	0	36.29	0.51

Table 18: Scenario 2b 6 Cluster

Methods	-2	-1	**0**	$+1$	$+2$
C-Index	507	428	**2040**	26	199
Ratkowsky	596	588	**2011**	5	0
TraceW	989	400	**1811**	0	0
Likelihood	1169	401	**1602**	2	26
Calinski	0	202	**1599**	10	1389
Scott	861	477	**1491**	126	245
DB-Index	1	437	**1440**	0	1322
Marriot	337	216	**1414**	177	1056
Hartigan	1394	400	**1406**	0	0
TraceW^{-1}B	415	691	**1217**	48	829
SSI	4	2	**1125**	89	1980
Ball	1600	800	**800**	0	0
$\|T\|/\|W\|$	2	0	**131**	4	3063
TraceCovW	14	110	**114**	112	2850

Table 19: All Results