



Evaluation of Hierarchical Clustering Algorithms as Classifiers of Hypertensive Blood Pressure and Heart Rate Recordings

A.N. Kastania, M.P. Bekakos

Department of Informatics, Athens University of Economics and Business, Athens, Greece

E-Mail: ank@dias.aueb.gr and mpb@dias.aueb.gr

Abstract

One of the biggest problems in hierarchical cluster analysis is validity of the classification results since various methods are available. This study evaluates hierarchical clustering algorithms as classifiers of 24h hypertensive blood pressure and heart rate recordings based on an algorithm designed and developed for this purpose. It also attempts to provide a comparison of the derived clusters with the results of a pattern recognition algorithm for biomedical waveform analysis in order to clarify the limitations on the use of hierarchical clustering methods for the classification of biomedical waveforms.

1 Introduction

Cluster analysis is a tool of exploratory data analysis that tries to find the intrinsic structure of data by organizing patterns into groups of clusters.¹ It is important for clustering how to define a cluster and how to evaluate clustering results. The clustering problems can be divided into attributive and non-attributive according to presence of absence of any properties of the objects considered. In a attributive case, selected properties of the objects are utilized in the clustering process.² In non-attributive case, non properties are known except the distance between the objects.

In hierarchical clustering,³ objects are linked to each other in a hierarchical way. A similarity or distance measure has to be defined for this technique. When two objects are linked to each other, the distances of this pair of objects to all other objects, not yet linked to this pair by previous links, are modified. Define the distance between two objects i and j as being $d(i, j)$. When the objects i and j are linked (denoted by object ij), the distance of an object k to this pair ij is in general given by:

$$d(k, ij) = a.d(k, i) + b.d(k, j) + c.|d(k, i) - d(k, j)|$$

250 Simulations in Biomedicine IV

For various values of a , b and c one can derive several clustering algorithms. Final clusters are defined by placing a threshold on the values of $d(i, j)$, at which (groups of) objects are linked, taking into account the structure of the dendrogram.⁴ Cluster validation refers to the procedures that evaluate the result of the clustering algorithms.

In this study an algorithm designed and developed in order to evaluate the validity^{5,6} of hierarchical clustering for the following algorithms:^{7,8} between groups linkage, within groups linkage, nearest neighbor, furthest neighbor, centroid clustering, median clustering, Ward's method. Statistical criteria are used to test the cluster validity. A comparison between the resulted valid clusters and the results of a pattern recognition algorithm⁹ is carried out to exploit the abilities and the limitations of cluster analysis in the classification of blood pressure and heart rate recordings.

2 Patients and clinical methods

A number of 230 subjects with untreated essential hypertension were included in the study. Mean age $\pm$ SE of these patients was 52.82 ± 12.05 years (range 19-79 years). All subjects were running a normal active life and had an eight to twelve year history of arterial hypertension.

The 230 subjects underwent 24-hour ambulatory blood pressure (BP) monitoring on a usual working day. They were instructed to act and work as usual. The Spacelabs 90202 ABP monitor (Spacelabs Inc., Redmond, Wash.) was used. The system was set to obtain BP and heart rate measurements every 15 min. Some values were rejected but no subject was included in this study having less than 3 readings per hour, in any of the 24-hours. Namely, 12 subjects out of an initial number of 242 were excluded because of unsatisfactory results of non-invasive BP recordings (having less than 3 readings per hour). Thus, a number of 80 to 96 systolic BP values per subject per 24-hour and their corresponding time of measurement, were fed into an IBM PC/PS 2-30 computer (IBM Corp., Purchase, N.Y.).

The accuracy of the Spacelabs monitoring system used, was tested against readings obtained by a mercury sphygmomanometer in all 230 cases. This testing was performed twice: Before each subject left the hospital, and 24 hours later, when he returned. If the recorded BP value obtained by the Spacelabs system differed by more than 9 mm Hg from the corresponding sphygmomanometer readings, the 24-hour recording was not included in this study. In 148 out of the 230 subjects included in the study, the above difference did not exceed 5 mm Hg.

Pattern recognition of 24h blood pressure and heart rate recordings was conducted via a pattern recognition algorithm for biomedical waveform analysis.⁹ The waveform recognition scheme carries out a classification in time domain based on the direction of change and on the relevant change of height among two consecutive time units during the 24h period.

3 Formulation of the hierarchical cluster evaluation problem

3.1 Agglomerative algorithm abstraction

The input for a clustering algorithm is the proximity (or the pattern) matrix. The output is a description of grouping the patterns into clusters. Partition methods produce a single partition of the entities to be clustered. Hierarchical algorithms develop tree structures representing a nested sequence of partitions.

The problem of clustering can be regarded as a state problem. Let S be a non-negative function measuring the sizes of subsets of a set W we want to partition and D be a distance function measuring the dissimilarity of subsets of W . Also let $P_q = \{X_1, X_2, \dots, X_q\}$ be a q partition of the set W .

The subsets $X_1, X_2, \dots, X_q$ are called the clusters of the partition P_q .

The pair $\langle P_q, R(P_q) \rangle$ where $R(P_q) = \langle S(P_q), D(P_q \times P_q) \rangle$, is called a q state.

Given an initial state $\langle P_q, R(P_q) \rangle$, a final state $\langle P_f, R(P_f) \rangle$, $f \leq q$, satisfying a certain condition $\text{expr}(R(P_f))$ by repeatedly applying permitted operations. If initially $q = n$, then every object forms a singleton cluster. In agglomerative clustering, the only operation applicable in any state is to merge two clusters into one, thus yielding a new state.

Let D be a dissimilarity function, $2 \leq q \leq n$, P_q a partition and $X, Y \in P_q$, $X \neq Y$, such that $D(X, Y) \leq D(U, V)$ for all $U, V \in P_q$, $U \neq V$. Then we denote the set $\{X, Y\}$ by $\text{mpair}(P_q, D)$, and the distance $D(X, Y) = D(Y, X)$ by $D^{\text{mpair}}(P_q)$. The dissimilarity function measures the distance or difference between clusters; the closer the clusters, the smaller the dissimilarity.

The general agglomerative algorithm follows.

```
procedure clustering(state)
  while not final(state) do
    let P be the partition of current state
    state := merge(state, mpair(P, D))
  od
end clustering
function merge ( $\langle P_q, R(P_q) \rangle$ ,  $\{X, Y\}$ ) returns state
  return  $\langle P_{q-1}, R(P_{q-1}) \rangle$ 
  where  $P_{q-1}$  is derived from  $P_q$ 
    by merging the clusters  $X$  and  $Y$ 
end merge
function final( $\langle P_q, R(P_q) \rangle$ ) returns boolean
  return ( $q=1$  or  $\text{expr}(R(P_q))$ )
end final.
```

Here, the search for $\text{mpair}(P, D)$ is done by any deterministic procedure. The condition expr is some case dependent function involving D and S functions. The merge function returns a partition with the number of clusters one smaller



than the original, formed by merging two given clusters into one; a new state. A common expr in *final* is $q=k$ for some fixed k .

The algorithm is specialized by defining the distances between clusters in several ways, yielding methods known as furthest neighbor, group average and centroid. A mechanism for controlling the termination of the agglomerative process has to be introduced since the number of clusters is not known in advance.

3.2 A strategy for assessing hierarchical cluster validity

When clustering algorithms are applied, the goal is to solve a classification problem. However, these algorithms are susceptible to two problems: (i) the criterion they optimize may not be a good estimator of "true" classification quality, and (ii) they often admit many (sub optimal) solutions. Cluster validation refers to the procedures that evaluate the results of the clustering algorithms. An algorithm to assess validity assurance is necessary at each clustering state to migrate problems (i) and (ii).

In order to have internal validation, it is necessary to use some kind of null hypothesis as the base for a statistic.⁵ It is very difficult to suggest a reasonable hypothesis for data of several dimensions. In order to use 1-D analysis the multi-dimensional data have to be projected on a line per dimension. In this case the natural strategy is to use a projection pursuit algorithm for finding the best directions of clustering data. A projection index that estimates the quality of clustering is associated with each projection. The goal is to find the projection that maximizes the index.¹⁰

A second level validation is needed to ensure the stability of the algorithm and to improve its performance. It is based on the use of a statistical score to choose the best clustering candidate.

4 The structure of the cluster evaluation algorithm for the classification of blood pressure and heart rate recordings

A major problem with hierarchical clustering methods is the tendency for classification errors to occur when the empirical data departs from the ideal conditions of compact isolated clusters. Many empirical data sets have structural imperfections that confound the identification of clusters. Clustering of BP and HR recordings is an attributive clustering problem. It is possible to locate every object in a multi-dimensional feature space thus threat them locally with a divide and conquer method. The method is applied to a given matrix of feature objects yielding to pseudo metric distance measures.

Agglomerative clustering methods are applied even though the number of clusters is not known a priori. A special stopping mechanism for deciding the termination of the agglomerative process is proposed in this study, based on a validation strategy applied at each state (Table I).

The algorithm has two levels of validation: (i) Performance of a projection pursuit using the clustering-indicating function as an index until a requested

statistical bound is achieved. Repetition of this process, several times, to obtain a list of clustering candidates. (ii) Use of a statistical score to choose the best clustering candidate (second level of validation). If the highest score passes a validation bound then the data are split into two clusters according to the best candidate, otherwise, it indicates the absence of (additional) clustering.

Step 1:	Begin with N clusters each containing one case. Denote the clusters 1 through n .
Step 2:	Choose a clustering-indicating function and a validation bound.
Step 3:	Apply validation strategy to find the most similar pair of clusters p and q ($p > q$). Denote this similarity S_{pq} .
Step 4:	Reduce number of clusters by one through merger of clusters p and q . Label the new cluster $t (= q)$ and update similarity matrix (by the method specified) to reflect revised similarities between cluster t and all other clusters. Delete the row and column of S corresponding to cluster p .
Step 4:	Perform Steps 3 and 4 until all entities are in one cluster or absence of additional clustering is detected.

Table 1: Notation and general procedure; S : a matrix of similarity or dissimilarity measures, S_{ij} : similarity or dissimilarity measure between cluster i and cluster j .

The classification validity of seven hierarchical clustering methods (between groups linkage, within groups linkage, nearest neighbor, furthest neighbor, centroid clustering, median clustering and Ward's method) is examined in terms of the hierarchical cluster evaluation algorithm. At each state the agglomeration schedule is computed as a clustering-indicating function. It must be noted however that for Ward's method, the coefficient given in the agglomeration schedule is really the within-cluster sum of squares at that state. For all other methods, this coefficient represents the distance at which the clusters p and q are joined. Usually, it is enough not to choose the validation bound to be near 0 (much too low) or near 1 (too strict). If it is too low, too many clusters will be found. If it is too high some clusters may not be found. In practice bounds in $[0.5, 0.9]$ have been very useful.

5 Experimental results

The performance of the proposed algorithm for the evaluation of hierarchical clustering algorithms used as classifiers of 24h blood pressure and heart rate recordings is demonstrated on real data. The important points to be checked are: (i) the ability of the algorithm to find the correct number of clusters (ii) the quality of the results.

Using a high bound for second level validation along with a suitable statistic for the first level of validation leads to the fact that the results of each run of the algorithm are almost the same. The evaluation algorithm resulted to

254 Simulations in Biomedicine IV

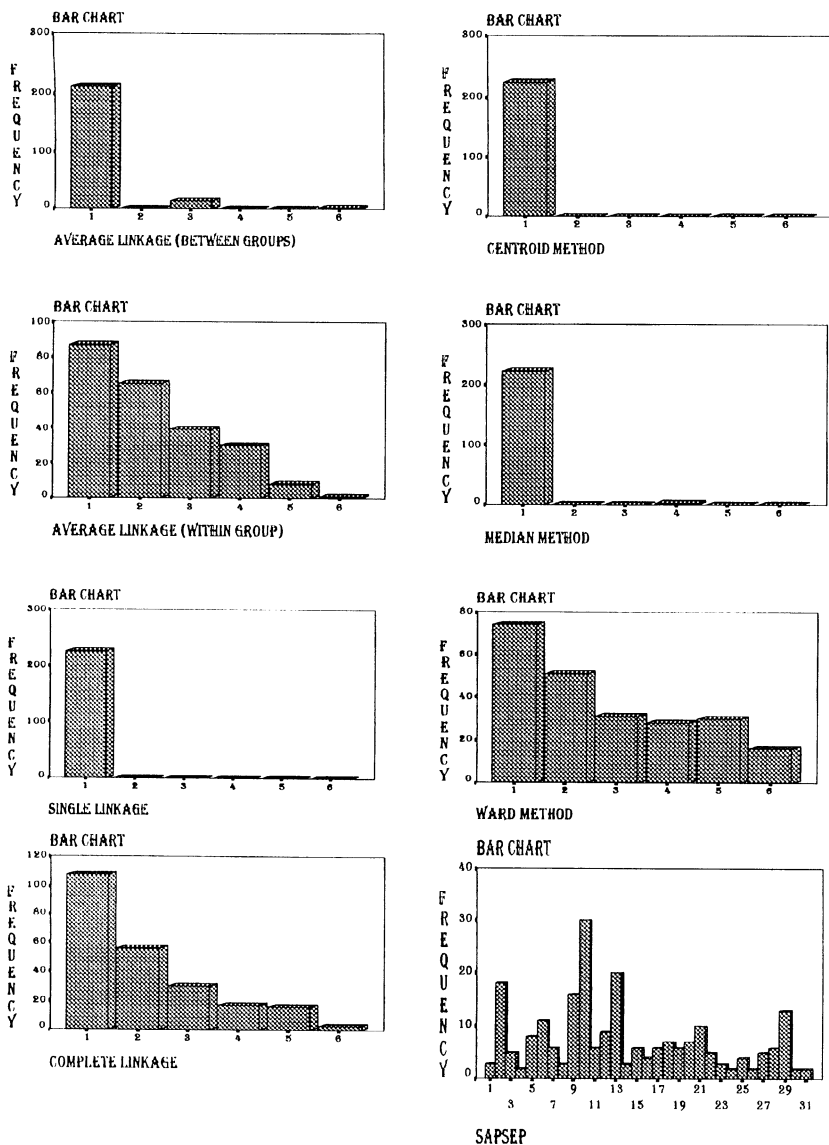
terminate at the average number of six clusters for all methods and for all parameters. This finding indicates the capability of the algorithm to detect the correct number of clusters.

There is a vast body in literature concerning the comparison of clustering methods. The comparisons are based on different comparison criteria, the most popular of which is the study of misclassifications. The results of a discriminant analysis performed on the resulted clusters are presented in Table 2.

Hierarchical Clustering Method	Percentage of correct classifications for SAP	Percentage of correct classifications for DAP	Percentage of correct classifications for HR
Between groups linkage	1.0000	1.0000	0.9913
Within groups linkage	0.9478	0.9783	0.9616
Nearest neighbor	1.0000	1.0000	1.0000
Furthest neighbor	0.9870	0.9739	0.9870
Centroid clustering	1.0000	1.0000	1.0000
Median clustering	1.0000	1.0000	1.0000
Ward's method	0.9957	0.9870	0.9783

Table 2: Performance evaluation of hierarchical cluster analysis methods used as classifiers of 24h blood pressure and 24h heart rate recordings; SAP: Systolic arterial pressure, DAP: Diastolic arterial pressure, HR: Heart rate.

Among the agglomerative clustering methods, the furthest neighbor method is considered to be the most reliable.^{5,6} The results in Table 2 support that the best method for hierarchical classification of 24h blood pressure and 24h heart rate recordings is the Ward's method. The agglomerative clustering methods that have previously been compared against each other, are also compared with the classification results of a pattern recognition algorithm for biomedical waveform analysis (Figure 1-Figure 3). After the application of the pattern recognition algorithm in the available data set, different number of clusters was derived and the number of cases correctly classified was reduced for all parameters but remained in acceptable bounds (greater 65%). The results reveal that the quality of clustering is closely related not only to the measurements available per clustering variable but also to the inter-relationships between the clustering variables that constitute the clustering cases.



SAPSEP

Pattern Recognition results

Correct classifications: 94.35%

Figure 1: The distribution of the cardinality of the clusters for 24h systolic arterial blood pressure recordings in comparison with the distribution of groups derived via a waveform pattern recognition algorithm.



256 Simulations in Biomedicine IV

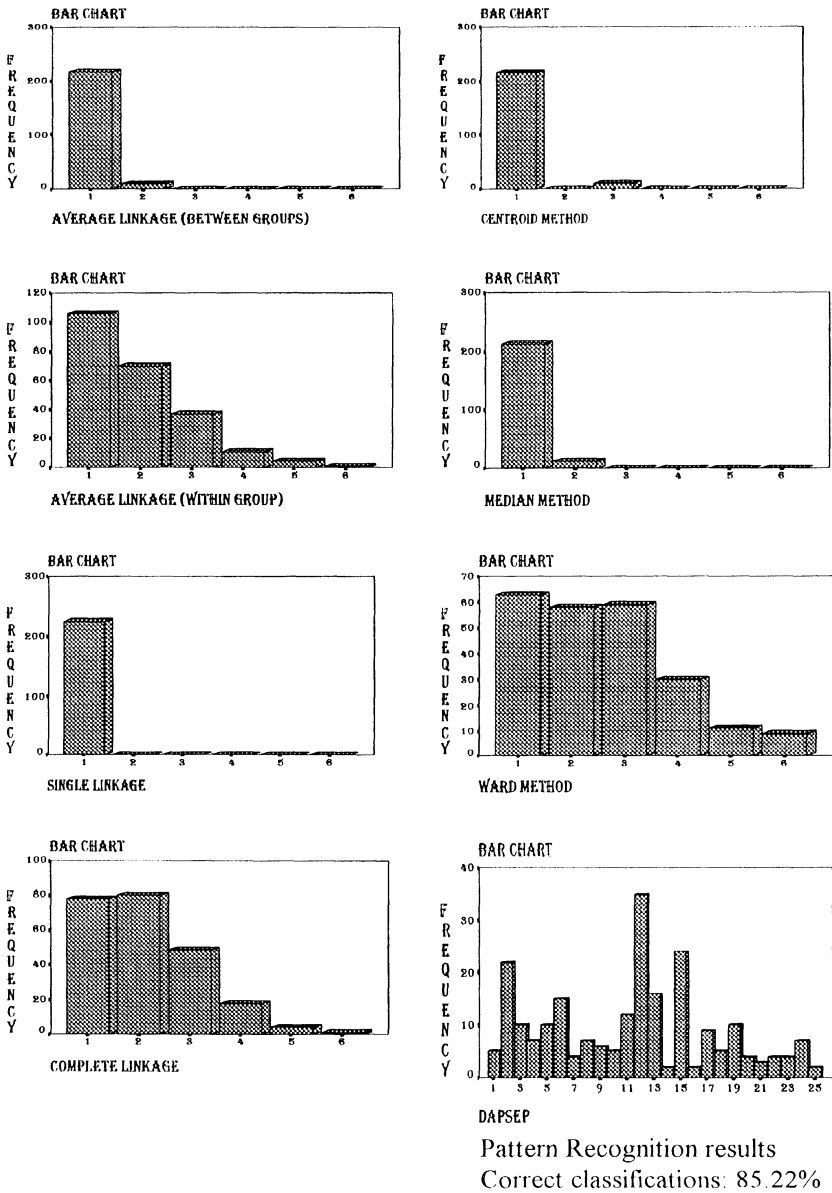
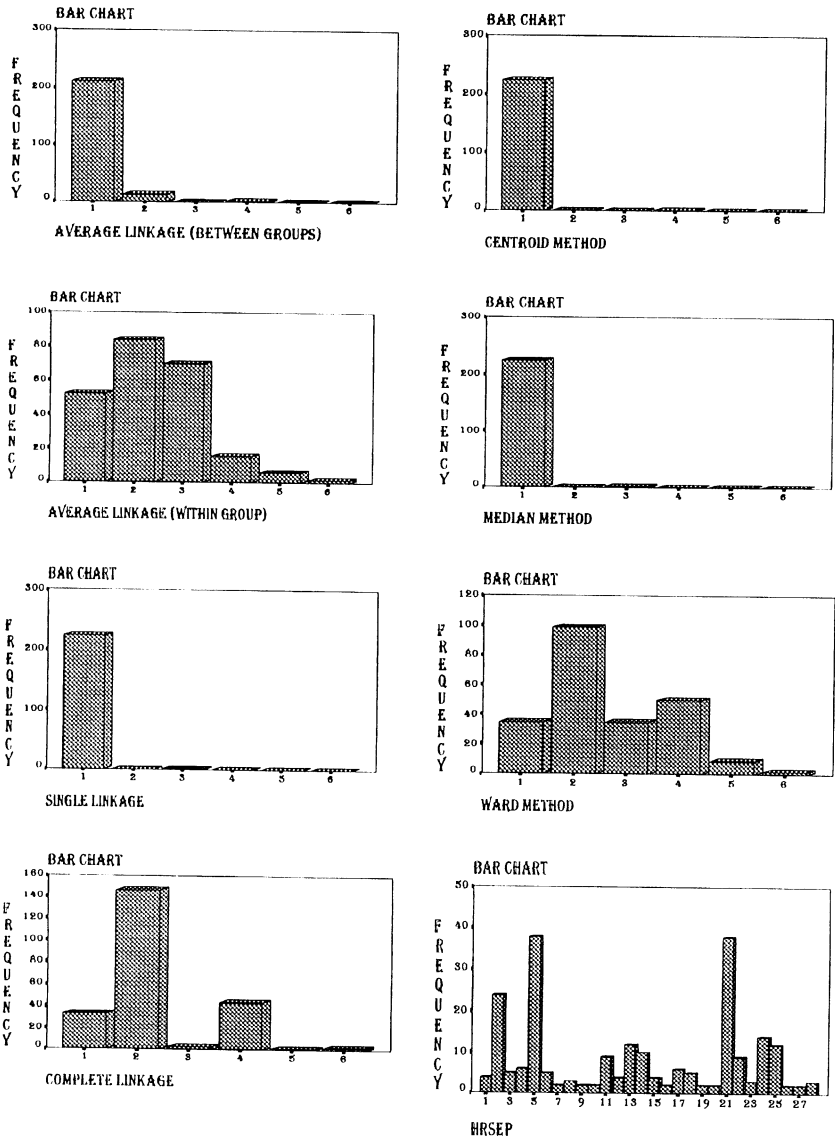


Figure 2: The distribution of the cardinality of the clusters for 24h diastolic arterial blood pressure recordings in comparison with the distribution of groups derived via a waveform pattern recognition algorithm.



HRSEP
Pattern recognition results
Correct classifications: 76.96%

Figure 3: The distribution of the cardinality of the clusters for 24h heart rate recordings in comparison with the distribution of groups derived via a waveform pattern recognition algorithm.



6 Conclusive remarks - further research

The algorithm designed and developed for the evaluation of hierarchical clustering algorithms can be used as a framework for assessing validity in hierarchical classification of 24h blood pressure and 24h heart rate recordings.

Although the Ward's method seems to be an excellent choice for the hierarchical classification of all parameters, the most suitable specialization for the classification problem of 24h blood pressure and heart rate waveforms should be determined by carefully considering the characteristics of the object space at hand.

The proposed algorithm can be optimized to a second stage consisting in using the best method on the entire sample of the multi-dimensional feature space when applying an automatic method to cut the dendrograms. This approach is currently under investigation.

References

1. Jain, A.K. and Dubes R.C. *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
2. Patrick, E.A. *Fundamentals of Pattern Recognition*, Prentice-Hall, London, 1972.
3. Hartigan J.A. *Clustering algorithms*, Wiley, New York, 1975.
4. Zahn, C.T. Graph-theoretical methods for detecting and describing gestalt clusters, *IEEE Trans. on Comp.*, 1971; C-20:68-86.
5. Dubes, R., Jain A.K. *Validity studies in clustering methodologies*. Pattern Recognition. 1979; 11:235-254.
6. Murtagh, F. Comments on Parallel algorithms for hierarchical clustering and cluster validity, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, (vol 14) 10,1056-7.
7. Lance G.N. & Williams W.T. A general theory of classificatory sorting strategies. I. Hierarchical systems, *Comput. J.*, 9, 373-380, 1967.
8. Sneath P.H. and Sokal R.R. *Numerical Taxonomy*, Freeman, San Francisco, 1973.
9. Kastania A.N. & Bekakos M.P. A parallel pattern recognition algorithm for biomedical waveform analysis, in ASE 95 (ed H. Power), pp. 269-277, *Proceedings of the 4th Int. Conf. on Applications of High Performance Computing in Engineering*, Milan, Italy, 1995, Computational Mechanics Publications, Southampton, UK, 1995.
10. Friedman, J.H. Exploratory projection pursuit. *J. Amer. Stat. Assoc.*, 82, 249-266, 1987.