

APPLICATION OF SINGLE-LINKAGE CLUSTERING METHOD IN THE ANALYSIS OF GROWTH RATE OF GROSS DOMESTIC PRODUCT (GDP) AT 1990 CONSTANT BASIC PRICES (MILLION NAIRA)

C. E. ONWUKWE AND J. N. EZEORAH

(Received 7, August 2009; Revision Accepted 18, December 2009)

ABSTRACT

Single-linkage is one of the methods in cluster analysis, which is used, for determining natural groupings in multi-variate data. Given a data set with one or more characteristics, single-linkage system classifies the data into clusters so that they are as similar as possible within each cluster and as different as possible between clusters. The objective is to show the closeness or similarity in the growth rate of GDP. Using the MINITAB software the similarity of the growth rate of GDP and the similarity in the years of production were shown.

KEYWORDS: Dendrogram, tree-diagram, nearest neighbour, single-linkage clustering, hierarchical clustering

1.0 INTRODUCTION

The importance of clustering is to reduce the amount of data by categorizing or grouping similar data items together. Clustering can be used to reduce the amount of data and to induce a categorization. The aim is to establish a set of clusters such that data within a cluster are more similar to each other than they are to cases in other cluster. Such grouping is pervasive in the way humans process information, and one of the motivations for using clustering algorithms is to provide automated tools to help in constructing categories or taxonomies Jardine and Sibson(1971) and Sneath and Sokal(1973). The methods may also be used to minimize the effects of human factors in the process. Clustering methods Anderberg(1973), Hartigan(1975), Jain and Dubes(1988), and Tryon and Bailey(1973) can be divided into two basic types: hierarchical and partitional clustering. Within each of the types there exists a wealth of subtypes and different algorithms for finding the clusters. In this paper, Single-Linkage hierarchical clustering is used in the data analysis. Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters. This is a procedure for transforming a proximity matrix into a sequence of nested partitions. The clustering methods differ in the rule by which it is decided which two small clusters are merged or which large cluster is split. The end result of the algorithm is a tree of clusters called a dendrogram, which shows how the clusters are related. By cutting the dendrogram at a desired level a clustering of the data items into disjoint groups is obtained. Single-linkage clustering is a hierarchical clustering which considers the distance between one cluster and another cluster to be equal to the shortest distance from any member of one cluster to any member of the other cluster. If the data consist of similarities we consider the similarity between one cluster and another cluster to be equal to the greatest similarity from one member to any member of the other cluster. (Stephen P. Borgatti "How to Explain Hierarchical Clustering" <http://www.Analytictech.Com/networks/hiclus.htm>).

The underlying mathematics of most of these methods are relatively simple but large numbers of calculations are needed which can put a heavy demand on the computer. The relationship between objects is represented in a proximity matrix in which rows and columns correspond to objects Maria (1999). Single-linkage method otherwise known as Minimum or Nearest Neighbour method, is employed in the analysis of growth rate of Gross Domestic Product (GDP) for the year (1994-2003) at 1990 basic prices (Million Naira); the aim is to maximize the minimum distance between clusters. The issues that often need to be considered when using Clustering in practice include how to scale the variables before calculating the distance matrix, which particular method of cluster analysis to be used and how to decide on the appropriate number of groups in the data Everitt *et al*(2001).

2.0 METHODOLOGY

The calculation of distance and similarity coefficients for pairs of items

Pairs of items are often compared on the basis of the presence or absence of some characteristics. Similar items have more characteristics in common than dissimilar items. The presence or absence of a characteristic can be described mathematically by introducing a binary variable, which assumes value 1 if the characteristic is present and value 0 if the characteristic is absent. For example, for 5 binary variables, the variable scores for two items *i* and *k* might be arranged as follows

	Variables				
	1	2	3	4	5
Item <i>i</i>	1	0	0	1	1
Item <i>k</i>	1	1	0	1	0

In this case there are two 1-1 matches, one 0-0 match, and two mismatches.

Let x_{ij} be the score (1 or 0) of the *j*th binary variable on the *i*th item and x_{kj} be the score (again, 1 or 0) of the *j*th variable on the *k*th item, $j = 1, 2, \dots, p$. consequently

$$(x_{ij} - x_{kj})^2 = \begin{cases} 0 & \text{if } x_{ij} = x_{kj} = 1 \text{ or } x_{ij} = x_{kj} = 0 \\ 1 & \text{if } x_{ij} \neq x_{kj} \end{cases}$$

and the squared Euclidean distance is given by

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

where x_{ik} and x_{jk} are the values of the *k*th variable for observations *i* and *j* provides a count of the number of mismatches. A large distance corresponds to many mismatches; that is dissimilar items Richard and Dean(1982). The different members of the class of hierarchical clustering techniques arise because of the variety of ways in which the distance between a cluster containing several observations and a single observation, or between two clusters, can be defined.

The frequencies of matches and mismatches for item *i* and *k* are arranged in the form of contingency table as in Table 2.1

Table 2.1

		Item k		
		1	0	Totals
Item i	1	a	b	a + b
	0	c	d	c + d
Totals		a + c	b + d	p = a + b + c + d

where “a” represents frequency of 1 – 1 matches, “b” is frequency of 1 – 0 matches, “c” is frequency of 0 – 1 matches and “d” is frequency of 0 – 0 matches.

Table of similarity coefficients for clustering items defined in the form of frequencies in table 2.1 above

Coefficient	Rationale
$\frac{a+d}{p}$	Equal weights for 1-1 matches and 0-0 matches
$\frac{2(a+d)}{2(a+b)+b+c}$	Double weight for 1-1 matches and 0-0 matches
$\frac{a}{p}$	No 0-0 matches numerator
$\frac{a}{a+b+c}$	No 0-0 matches in numerator or denominator. (The 0-0 matches treated as irrelevant)
$\frac{2a}{2a+b+c}$	No 0-0 matches in the numerator or denominator. Double weight for 1-1 matches
$\frac{a}{a+2(b+c)}$	No 0-0 matches in numerator or denominator. Double weight for unmatched pair
$\frac{a}{b+c}$	Rational of matches to mismatches with 0-0 matches excluded

(p binary variables, see table 2.1 above)

As earlier stated, the data employed in this study was obtained from the Central Bank of Nigeria Statistical Bulletin and is given in Table 2.3, and using the data the Dendrogram for the study was produced as in Figure 2.1.

2.1 The Algorithm

Given a set of N items to be clustered, and an N*N distance (or similarity) matrix, the basic process of hierarchical clustering defined by Johnson(1967) is this:

- 1 Start by assigning each item to a cluster, so that if you have N items, you now have N clusters, each containing just one item. Let the distances (similarities) between the clusters be the same as the distances (similarities) between the items they contain.
- 2 Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one cluster less.
- 3 Compute distances (similarities) between the new cluster and each of the old clusters.
- 4 Repeat steps 2 and 3 until all items are clustered into a single cluster of size N. (*)

Single-Linkage Clustering: The Algorithm

Let's now take a deeper look at how Johnson's algorithm works in the case of single-linkage clustering.

The algorithm is an agglomerative scheme that erases rows and columns in the proximity matrix as old clusters are merged into new ones.

The $N \times N$ proximity matrix is $D = [d(i,j)]$. The clusterings are assigned sequence numbers $0, 1, \dots, (n-1)$ and $L(k)$ is the level of the k th clustering. A cluster with sequence number m is denoted (m) and the proximity between clusters (r) and (s) is denoted $d[(r),(s)]$.

The algorithm is composed of the following steps:

- 1 Begin with the disjoint clustering having level $L(0) = 0$ and sequence number $m = 0$.
- 2 Find the least dissimilar pair of clusters in the current clustering, say pair $(r), (s)$, according to $d[(r),(s)] = \min d[(i),(j)]$
Where the minimum is over all pairs of clusters in the current clustering.
- 3 Increment the sequence number : $m = m + 1$. Merge clusters (r) and (s) into a single cluster to form the next clustering m . Set the level of this clustering to $L(m) = d[(r),(s)]$
- 4 Update the proximity matrix, D , by deleting the rows and columns corresponding to clusters (r) and (s) and adding a row and column corresponding to the newly formed cluster. The proximity between the new cluster, denoted (r,s) and old cluster (k) is defined in this way:
 $d[(k), (r,s)] = \min d[(k),(r)], d[(k),(s)]$
- 5 If all objects are in one cluster, stop. Else, go to step 2.

3.0 DATA

As earlier stated, the data employed in this study was obtained from the Central Bank of Nigeria Statistical Bulletin and is given in table 2.3, and using the data the Dendrogram for the study was produced as in Figure 2.1 and figure 2.2

Table 2.3: Data for Gross Domestic Product (1994-2003)

	YR1994	YR1995	YR1996	YR1997	YR1998	YR1999	YR2000	YR2001	YR2002	YR2003
agri(cps)	3	3.4	3.8	4.3	3.9	5.2	3	3.8	4.2	7
livstoc	1	4.2	2.9	2.5	2.7	2.8	1.3	3	4.8	4.2
forestry	2.6	2.2	0.5	0.9	1.2	1.3	1.5	2	0.7	1.5
fishing	-6.5	10	20.8	11.3	14.1	14.2	4	8	6.3	4.1
coalmin	-61	-18.8	7.7	7.1	0	6.7	15.5	-11.1	-37.5	0
crpna	-2.6	2.4	7.2	1.5	2.2	-7.5	11.1	5.2	-5.7	23.9
metor	17.6	6.5	8.9	-8.5	0.6	2.5	3	-6.2	0.5	-3.8
minqua	3	2.7	1.2	6.4	6	3.8	3.8	10	4.3	5.5
oilref	-1.5	6.3	11.4	-1.6	-10.3	4.6	-1	191.1	-7.9	9.5
cem	-8.7	-1.9	-4.7	3	-61.5	-0.3	-0.4	15.6	1.1	4
otherman	-0.9	-5.5	1	0.3	-3.9	3.5	3.6	4.2	11	5.6
electricy	7.4	-1.6	2.2	-0.6	-6.5	1.3	2	13.7	28.6	3.4
water	2.3	1	0.9	1.5	1	3.1	3.5	2.4	7.5	8.2
buildcon	3	2.7	1.2	6.4	6	3.8	4	12	4.3	8.8
who&ret	0	0.1	0.8	1.5	3	2.5	1.6	2.5	6.5	5.8
hot&rse	0.5	1	2.3	3	3.2	3.5	3.7	4.5	6.8	4.6
rdtrans	1.5	1	2	3	3.5	3	3.1	4.3	18.6	1.2
ritrans	-95.4	-25.6	9.7	10.8	4.4	5.1	4	5.4	3.7	5
watrans	-10	6	3	1	2	2.5	4	4	19	7.4
aitrans	-10	1.5	3	1	2	2.5	4	4	19	7.4
transsev	13.7	16	15	15	20	20	8	10	15	1.3
telecom	1	5	5	6	5	5.2	6.1	30	46.6	43
post	0.7	2	4	5.5	7	7.4	7.5	8	63.1	-8.2
fininst	3	4.2	4	4.2	5	3.5	3.9	4.5	29.4	-8.2
insur	2.5	2	3	3.5	3.6	3.8	5.8	4	29.4	-19.3
reales	3	3.2	1	6.4	6	3.8	3.9	5	3	3
busser	3.5	3.7	3.6	3.5	5	4.4	4.3	5	6.5	4.4
pubadmin	1.4	-0.1	2.9	1.7	1.7	1.7	1.6	1.6	18.1	0
edu	1.4	-0.1	2.9	1.7	1.7	1.7	1.6	1.6	7.9	7
health	1.4	-0.1	2.9	1.7	1.7	1.7	1.6	1.6	0.8	2

Source: Central Bank of Nigeria Statistical Bulletin 200

4.0 RESULTS

Table 4.1. This shows the similarity levels of the growth rate of GDP

/No	Products	Similarity level	Distance level
1	Minqua and reales	96.9478	0.30553
2	Minqua and buildcon	96.6518	0.33517
3	Who& ret and edu	94.9685	0.50367
4	Hot & rse and busser	94.1932	0.58128
5	Agric(cps) and hot & rse	94.0010	0.60052
6	Agric (cps) and water	93.8200	0.61864
7	Forest and health	93.8904	0.62161
8	Forest and health	93.7046	0.63019
9	Agric(cps) and livstoc	93.0289	0.69783
10	Rdtrans and pubadmin	92.6675	0.73401
11	Agric(cps) and forestry	92.5635	0.74442
12	Agric(cps) and rdtrans	92.2097	0.77983
13	Agric(cps) and minqua	91.1468	0.88623
14	Agric(cps) and aitrans	89.9701	1.00403
15	Agric(cps) and watrans	89.4490	1.05619
16	Agric(cps) and otherman	89.4377	1.05732
17	Agric(cps) and electricy	88.0325	1.19799
18	Fininst and insur	87.4574	1.25556
19	Agric(cps) and and fininst	86.0986	1.39158
20	Agric(cps) and post	75.3800	2.46455
21	Fishing and transsev	72.1254	2.79035
22	Agric(cps) and metor	69.7207	3.03106
23	Agric(cps) and crpena	59.5867	4.04551
24	Agric(cps) and telecom	57.5698	4.24741
25	Coalmin and ritrans	56.5620	4.34830
26	Agric(cps) and fishing	54.6595	4.53874
27	Agric(cps) and cem	50.5436	4.95076
28	Agric(cps) and coalmin	42.3536	5.77061
29	Agric(cps) and oilref	39.3856	6.06771

Fig.2 appendix 2, shows the similarity (dissimilarity) in years of the product. If the economy of a country is dependent on the agricultural produce in the table, this single-linkage method also shows the similarity in the economy of the country within the stipulated years. The similarities in year of the Gross domestic products are shown in the table below. This starts with the year of highest similarity to the least.

Table 4.2: This shows the similarity levels in the years of production

S/No	Years	Similarity level	Distance level
1	YR 1994 AND YR 1995	86.3372	0.136628
2	YR 1997 AND YR 1999	67.5844	0.324156
3	YR 1996 AND YR 1997	63.6029	0.363971
4	YR 1996 AND YR 1998	47.3209	0.526791
5	YR 1996 AND YR 2000	42.0105	0.579895
6	YR 1994 AND YR 1996	36.9177	0.630823
7	YR 1994 AND YR 2002	34.4057	0.655943
8	YR 1994 AND YR 2001	32.9965	0.670035
9	YR 1994 AND YR 2003	20.6225	0.793775

5.0 DISCUSSION

From fig.1 in appendix1, it can be seen that between 1994 and 2003, the growth rate of the GDP of items 8 and 26 were most similar. This forms a new cluster 8. It can be seen that the nearest item similar to new cluster 8 now formed is that of item 14, therefore a new cluster of 8 is formed. At a higher linkage distance items 15 and 29 are seen to be more similar so they form a new cluster 15. At higher distance 0.57544, items 1 and 27 are similar, so they form a new cluster of item 1. This new cluster now compared with the nearest in similarity in the growth rate of items around. It can be seen that the most similar is item 16. Then item 1 is cluster to item 16 to form a new item of cluster of item 1. This item 1 is clustered to the nearest item which is item 13 to form a new cluster of item 1. At a higher distance of 0.62161, the growth rate of item 3 is similar to that of item 30, so they are joined to produce a new cluster of item 3. This item is joined to item 15 because of their similarity to form a new cluster of item 3. At a distance level of 0.69794, item 1 is similar to item 2. They are joined to form a new cluster of item 1. At a higher distance 0.73413, item 17 and 28 are joined together because of their similarity to form a new item 17. At a distance of 0.74461, item 1 is joined to the nearest item, which is item 3 to form a new cluster of item 1. Item 1 is joined to item 17 at a linkage distance of 0.77985, to form a new cluster of item 1. The new item 1 is joined to item 8 at a linkage distance of 0.87618, to form a new cluster of item 1. The new cluster 1 is joined with item 20 to form a new cluster of item 1 with 15 observations in the new cluster. The new cluster 1 is joined to 19 to form a new cluster of item 1 at the linkage distance of 1.05619 with 16 observations in the new item. The new item 1, is joined with the most similar item which is item 11 to form a new cluster of item 1. This new item 1 is joined with item 12 to form a new item of item 1 with 18 observations in the new cluster; this is because of their similarity. Items 24 and 25 at the linkage distance of 1.25567 are similar, so they join to produce a new cluster of item 24. At a linkage distance of 1.39160 items 1 is joined to item 24 to form a new item of item 1. The new item is joined to item 23 because of their similarity to form a new item 1. Item 4 is joined to item 21 to form a new item 4. The new item 1 is a cluster joined to item 7 to form new item 1. The new item 1 is joined to 6 with the linkage distance of 4.04600 to form a new item of 1. This item 1 is clustered to item 22 to form new item 1 with 24 observations. Item 5 is clustered with item 18 at the linkage distance of 4.34896 to form a new item 5. Item 1 is a cluster to item 4 to form a new cluster item 1. This new item 1 is clustered to item 10 in similarity to form a new cluster of 1. This new cluster of item 1 is joined with item 5 to form a new cluster of 1 with 29 observations. The new cluster item 1 is joined to item 9 to form a new cluster of item 1 with 30 observations. These clusters are done in terms of their similarities.

From Fig. 2 in appendix2, Growth rate of GDP of years 1994 and 1995 are more similar than any other year in the table. So years 1994 and 1995 form a new cluster of 1, which is now 1994. After these years 1997 and 1999 followed in similarity. These years 1997 and 1999 formed a new cluster of 1997. It is now discovered in similarity that year 1996 is similar to the new year 1997. Therefore, a cluster of new year 1996 is formed. This cluster of new year 1996 is now clustered with year 1998 in similarity to produce a cluster of a new 1996. The new 1996 is clustered with year 2000 to form a new cluster of 1996. The new cluster of 1 that is 1994 formed above is now clustered with year 2002 to form a new cluster of 1994. This new cluster of 1994 is clustered to the New Year 1996 to form a new 1994. This new 1994 is clustered to year 2001 to form a new cluster of 1994. This is clustered to year 2003 to form a one cluster of a new cluster of 1994.

6.0 CONCLUSION

Similarity (dissimilarity) in the products within the years. This can be extended to the similarity (dissimilarity) in the economy within the years. The similarity of the growth rate of the Gross domestic products can also be used to judge the similarity (dissimilarity) in the economy within the corresponding years.

REFERENCES

Anderberg, M. R., 1973. Cluster Analysis for Applications. Academic Press, New York.

Central Bank of Nigeria Statistical Bulletin. Vol.14, Dec 2003. Pages 242-244

Everitt, B.S., Landau, S., and Leese, M., 2001. Cluster Analysis (4th edition), Edward Arnold, London.

Hartigan, J., 1975. Clustering Algorithms. Wiley, New York.

Jain, A. K., and Dubes, R. C., 1988. Algorithms for Clustering Data. Prentice Hall, Englewood Cliffs, NJ.

Jardine, N., and Sibson, R., 1971. Mathematical Taxonomy. Wiley, London
 Maria Irene Miranda: Clustering methods and algorithms.
<http://www.cse.iitb.ac.in/dbms/Data/Courses/CS632/1999/clustering/dbms.htm>

Richard A. J., and Dean W. W., 1982. Applied Multivariate Statistical Analysis 3rd Ed. Pages 573-589.

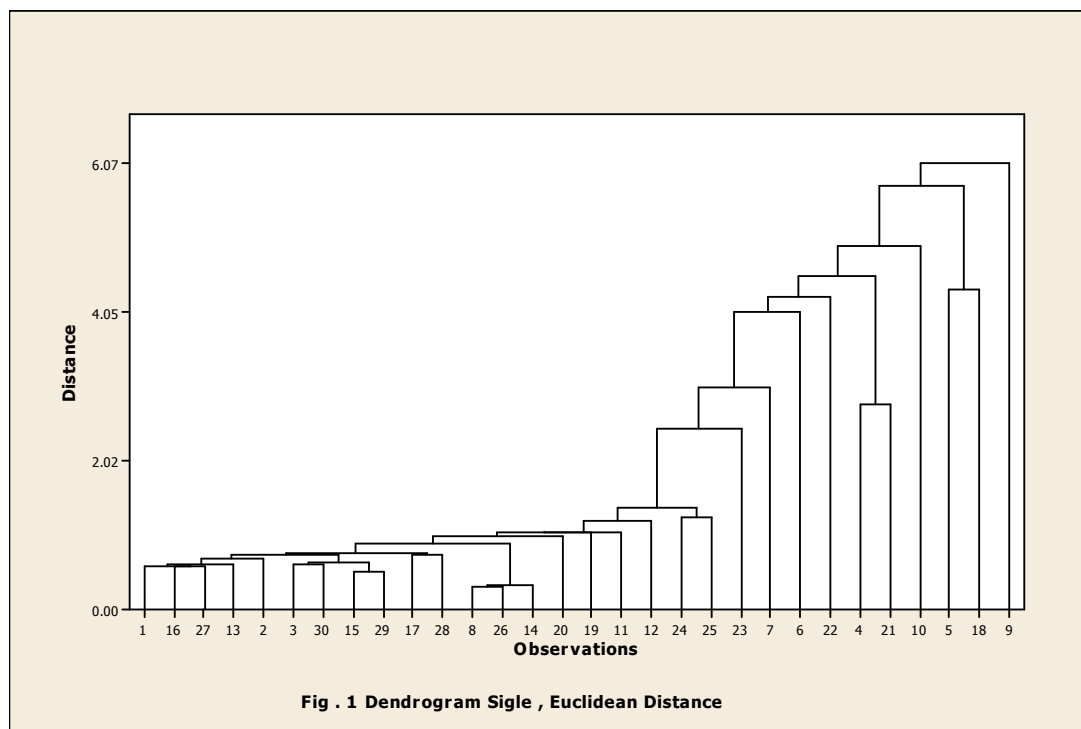
Johnson S. C., 1967. Hierarchical Clustering Schemes. Psychometrika, 2:241-254

Sneath, P. H. A., and Sokal, R. R., 1973. Numerical Taxonomy. Freeman, San Francisco, CA.

Stephen P. B., How to explain hierarchical clustering.
<http://www.analytictech.com/networks/hiclus.htm>

Tryon, R. C., and Bailey, D. E., 1973. Cluster Analysis. McGraw-Hill, New York, NY.

APPENDIX 1



Cluster analysis of observations: YR1994, YR1995, YR1996, YR1997, YR1998, ...

Standardized Variables, Euclidean Distance, Single Linkage Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		Number of obs.	
						New cluster	in new cluster
1	29	96.9478	0.30553	8	26	8	2
2	28	96.6518	0.33517	8	14	8	3
3	27	94.9685	0.50367	15	29	15	2
4	26	94.1932	0.58128	16	27	16	2
5	25	94.0010	0.60052	1	16	1	3
6	24	93.8200	0.61864	1	13	1	4
7	23	93.7904	0.62161	3	30	3	2
8	22	93.7046	0.63019	3	15	3	4
9	21	93.0289	0.69783	1	2	1	5
10	20	92.6675	0.73401	17	28	17	2
11	19	92.5635	0.74442	1	3	1	9
12	18	92.2097	0.77983	1	17	1	11
13	17	91.1468	0.88623	1	8	1	14
14	16	89.9701	1.00403	1	20	1	15
15	15	89.4490	1.05619	1	19	1	16
16	14	89.4377	1.05732	1	11	1	17
17	13	88.0325	1.19799	1	12	1	18
18	12	87.4574	1.25556	24	25	24	2
19	11	86.0986	1.39158	1	24	1	20
20	10	75.3800	2.46455	1	23	1	21
21	9	72.1254	2.79035	4	21	4	2
22	8	69.7207	3.03106	1	7	1	22
23	7	59.5867	4.04551	1	6	1	23
24	6	57.5698	4.24741	1	22	1	24
25	5	56.5620	4.34830	5	18	5	2
26	4	54.6595	4.53874	1	4	1	26
27	3	50.5436	4.95076	1	10	1	27
28	2	42.3536	5.77061	1	5	1	29
29	1	39.3856	6.06771	1	9	1	30

Final Partition

Number of clusters: 9

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	21	40.9023	1.21760	3.54885
Cluster2	2	3.8930	1.39517	1.39517
Cluster3	1	0.0000	0.00000	0.00000
Cluster4	1	0.0000	0.00000	0.00000
Cluster5	1	0.0000	0.00000	0.00000
Cluster6	1	0.0000	0.00000	0.00000
Cluster7	1	0.0000	0.00000	0.00000
Cluster8	1	0.0000	0.00000	0.00000
Cluster9	1	0.0000	0.00000	0.00000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
YR1994	0.236893	0.35993	-2.65846	0.07024	1.01408	0.12164	-0.21478
YR1995	0.064552	1.59224	-2.66743	0.17235	0.72156	0.69476	-0.40364
YR1996	-0.408054	2.76926	0.68670	0.58461	0.93171	1.44214	-1.84504
YR1997	-0.144183	2.25237	0.83782	-0.47152	-2.80962	-1.19633	-0.12080
YR1998	0.126739	1.24085	-0.07676	0.09325	-0.03040	-0.87274	-4.82946
YR1999	-0.165507	2.95411	0.62956	-2.54433	-0.30919	0.16019	-0.93503
YR2000	-0.205327	0.73528	2.97208	2.49030	-0.29709	-1.67359	-1.46711
YR2001	-0.197606	-0.07525	-0.65621	-0.18508	-0.51459	5.18815	0.11552
YR2002	0.171706	0.05781	-2.63606	-0.85693	-0.51006	-0.98001	-0.47649
YR2003	-0.224349	-0.15819	-0.42237	1.91617	-0.79419	0.50717	-0.03098

Variable	Cluster8	Cluster9	Grand centroid
YR1994	-4.26578	0.23845	0.000000
YR1995	-3.57831	0.52063	-0.000000
YR1996	1.09505	0.13543	0.000000
YR1997	1.70292	0.58063	-0.000000
YR1998	0.26327	0.30963	0.000000
YR1999	0.27194	0.29429	-0.000000
YR2000	0.04703	0.76969	0.000000
YR2001	-0.17930	0.53173	0.000000
YR2002	-0.33102	2.06913	-0.000000
YR2003	0.06686	3.78505	-0.000000

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Cluster1	0.00000	5.48435	6.07554	4.43829	3.29832	6.25055	5.45825
Cluster2	5.48435	0.00000	7.30758	7.35351	6.66614	7.86232	9.40214
Cluster3	6.07554	7.30758	0.00000	6.03537	7.38822	9.16193	8.28131
Cluster4	4.43829	7.35351	6.03537	0.00000	5.20916	7.61537	7.27666
Cluster5	3.29832	6.66614	7.38822	5.20916	0.00000	6.39692	6.59458
Cluster6	6.25055	7.86232	9.16193	7.61537	6.39692	0.00000	7.51362
Cluster7	5.45825	9.40214	8.28131	7.27666	6.59458	7.51362	0.00000
Cluster8	6.31088	7.75066	4.34830	7.44874	8.25734	8.92876	8.26209
Cluster9	4.73621	6.22903	8.08026	5.00918	6.56537	7.34973	7.72103

	Cluster8	Cluster9
Cluster1	6.31088	4.73621
Cluster2	7.75066	6.22903
Cluster3	4.34830	8.08026
Cluster4	7.44874	5.00918
Cluster5	8.25734	6.56537
Cluster6	8.92876	7.34973
Cluster7	8.26209	7.72103
Cluster8	0.00000	7.73860
Cluster9	7.73860	0.00000

APPENDIX 2

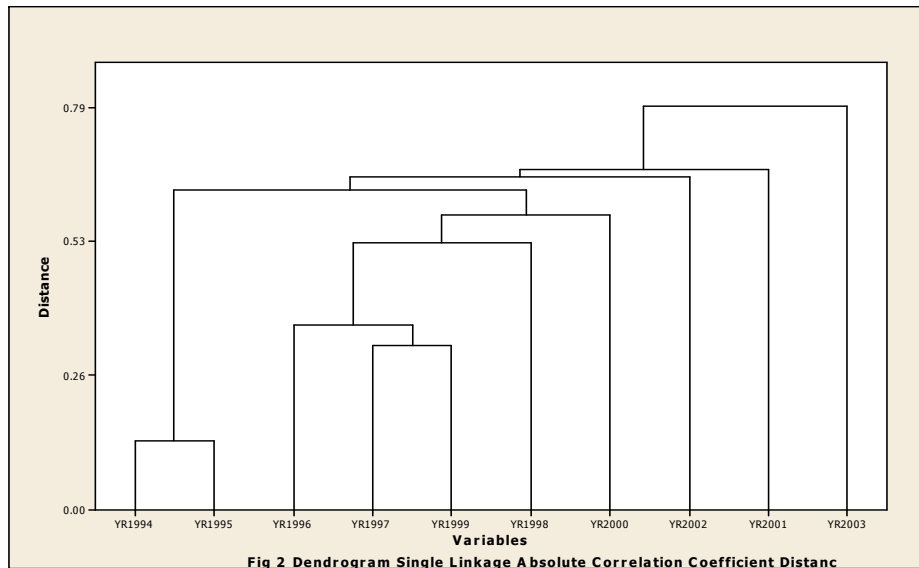


Fig 2 Dendrogram Single Linkage Absolute Correlation Coefficient Distanc
Cluster analysis of variables: YR1994, YR1995, YR1996, YR1997, YR1998, ...

Absolute Correlation Coefficient Distance, Single Linkage Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	9	86.3372	0.136628	1 2	1	2
2	8	67.5844	0.324156	4 6	4	2
3	7	63.6029	0.363971	3 4	3	3
4	6	47.3209	0.526791	3 5	3	4
5	5	42.0105	0.579895	3 7	3	5
6	4	36.9177	0.630823	1 3	1	7
7	3	34.4057	0.655943	1 9	1	8
8	2	32.9965	0.670035	1 8	1	9
9	1	20.6225	0.793775	1 10	1	10

Final Partition

- Cluster 1
YR1994 YR1995
- Cluster 2
YR1996
- Cluster 3
YR1997
- Cluster 4
YR1998
- Cluster 5
YR1999
- Cluster 6
YR2000
- Cluster 7
YR2001
- Cluster 8
YR2002
- Cluster 9
YR2003

