SOME STATISTICAL MEASURES OF REGIONAL VARIATION IN SOCIOECONOMIC PERFORMANCE- THE ETHIOPIAN SETTING

A THESIS SUBMITTED TO
THE SCHOOL OF GRADUATE STUDIES,
ADDIS ABABA UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE IN
STATISTICS

BY
SILESHI FANTA
JUNE, 1994.

ACKNOWLEDGEMENTS

It is a pleasant duty to express my sincere gratitude and appreciation to professor Asmerom Kidane, Head of the department of Statistics, who took a personal interest in my work as a research advisor, generously gave his time, energy, valuable advice and guidance at all stages of this project.

I would like to thank my employer, Ministry of Planning and Economic Development of Ethiopia who granted me sponsorship for M.Sc. program. My thanks also go to Ato Temesgen Zeweter, Ato Grimay Medhin and Ato Fentaw Abegaz, for their unforgettable supply of some materials which are essential throughout my study.

I am heavily indebted to Ato Hailae Mekonen and Ato Tesfaye Admassie, Colleagues in the Ministry of Planning and Economic Development, the work would not have been completed at all had it not been for their encouragement and support that I received at a crucial stage of editing and printing the paper.

ABSTRACT

This study focused on the variation in the regional distribution of some socioeconomic services in Ethiopia. The paper also attempts to compare different natures of regional variation which were obtained on the basis of socioeconomic data for the year 1979 with the natures of regional variation obtained on the data of the same sort for the year 1989 .

Some Uni-variate descriptive measures and a multivariate method called cluster analysis were used to study the variation among regions. The six clustering procedures that are generated by a common algorithm [**Wishart (1969)**] were applied to investigate the natures of regional disparities in their socioeconomic status for the year 1979 and the year 1989. Different natures of regional variation for changes in socioeconomic characteristics of regions were also studied by applying six clustering procedures on the ratios the characteristics of regions in the year 1989 to that of the year 1979 .

Some conclusions were drawn on the basis of regional characteristics considered. Recommendations were also made to indicate some possibilities for future work in the area

INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

In many countries of the world there are regions which are at different levels of economic or socioeconomic development. Due to variations in area, population size, fertility of soil, the availability of water resources and climatic conditions, some regions may be in a better status than others. There may also be regional variations in the distribution of some socioeconomic services such as primary schools, junior secondary schools, senior secondary schools, health centres, health stations, hospitals, industrial establishments, transport and communication services, etc.

This problem has a long existence almost all over the world. For example, in 1959, the Brazilian northeast contained 25 percent of the country's population but only 10 percent of its income. The southern states, on the other hand, accounted for 35 percent of the population but 50 percent of the national income. In Canada the Atlantic region, consisted of 11 percent of the national population, while the region produced less than 6 percent of the gross national product during 1958 [Chaudhry (1971)]. France, Italy and America have the problem of regional disparity which they call 'North-south' problem.

The question of regional disparity is also a considerable problem in Ethiopia. There are regions which may have quite different socioeconomic status. The differences may be in educational, health, transport and communication services, etc. If these services and other economic variables are believed to gauge the relative level of development of regions, then equitable regional distributions of these characteristics is really essential to keep the balanced status of regions.

1.2 OBJECTIVES OF THE STUDY

The socioeconomic status of a given region at one particular time may be similar to the status of another. This similarity in socioeconomic status may change from time to time. Regions which were similar at one particular time may not be similar at another period of time due to the differences in the change of socioeconomic characteristics. These differences in the change of socioeconomic characteristics must receive special attention, otherwise the disparities among regions may in fact become much greater than the existing disparities. Therefore one has to study the extent of regional variation in socioeconomic performance.

The main purpose of this study is to indicate the degree of similarity among regions of Ethiopia by applying some statistical measures on socioeconomic data of the years 1979 and 1989 then to compare the structures of similarity in the socioeconomic performance of regions for the two specified years. Moreover we shall classify the regions according to similarity in change of some socioeconomic variables. Therefore the specific objectives of this particular project are to

(I). assess the similarity among regions on the socioeconomic characteristics in 1979 .

(II). assess the similarity among regions based on the socioeconomic characteristics in 1989 .

(III). see if there is a change in the structure of similarity among regions based on (I) and (II).

(IV). investigate the extent of regional similarity in the changes of socioeconomic characteristics in the ten years time 1979-1989 .

The central government should keep balanced regional growth by providing closer assistance and support to the relatively less developed regions. While it can be argued that backward regions on account of inequitable distribution

of health stations, schools, hospitals, communication facilities, etc. deserve a relatively more central government support, it is equally important to consider that better off regions should not be punished as such regions contribute the lion's share to the revenue. A strong bias, may thus create not only political problems, but also economic inefficiency . To overcome such problem planners may participate in regional allocation of development grant. Therefore, it is expected that the final result of this paper may serve those planners who have a participation in the study of regional allotment of development grant. It is believed that it may give them a highlight of the past and present status of the Ethiopian regions.

Chapter two deals with sources and an overview of the data used in this thesis. Chapter three is devoted to the description of procedures, methods and underlying concepts that could be applied in the study of regional similarity. The results of the study are given in Chapters four and five. Chapter six is confined to discussions, conclusions and recommendations.

CHAPTER II

SOURCES AND AN OVERVIEW OF THE DATA

2.1 SOURCES OF THE DATA

The socioeconomic data that are utilized in this paper are obtained from various statistical abstracts and statistical bulletins which are produced by the Central Statistical Authority, the Ministry of Education, the Ministry of Health especially Planning and Programming Bureau of the Ministry and the Ministry of Planning and Economic Development. The data obtained from the above sources are attached to the paper as Appendices I and II .

2.2 OVERVIEW OF THE DATA

The main problem in obtaining the regional data is the classification of administrative regions. Different governments put the size and structure of administrative regions in different ways. Therefore, the regional data which were obtained before ten years may not be comparable with the present regional data. The classification of regions used in this study is as follows:

1. Arsi	5. Gonder	9. Sidamo
2. Bale	6. Harerghe	10. Tigray
3. Gamogofa	7. Illubabor & Kefa	11. Welega
4. Gojam	8. Shewa	12. Welo

The number given in front of each region will be used as a code for the region throughout this paper. Illubabor and Kefa were taken as one category since the data for the two regions were inseparably mixed in 1989 .

It was originally planned to consider more socioeconomic variables in this study. For example, total employment in the public (state-owned) and in private sector, population supplied with potable water, population supplied with electricity, educational structure of government and public sector employees and the number of professionals (by type of qualifications) and employment in manufacturing industries. However, dictated by data constraints we will use only nineteen socioeconomic variables.

CHAPTER III

PROCEDURES AND METHODS OF DATA ANALYSIS

3.1 PROCEDURES

The variables were originally in the following form for each region.

- Hospital beds - the number of hospital beds in each region.
- Telephone lines - total number of telephone lines in each region.
- Length of roads - total length of roads in kilometres in each region and so on .

We will use these variables to cluster regions into groups that are relatively homogeneous with respect to these variables. However, regions differ on the basis of these variables simply as functions of their populations, areas or part of populations to which particular service is rendered. We, therefore, re-scaled the variables in the following way:

y_1 = Hospital beds per 10,000 population

y_2 = Length of roads in kilometres per 1,000 square kilometres.

y_3 = Percentage of urban population

y_4 = Number of telephone lines per 10,000 population

y_5 = Number of post offices per 10,000 population

y_6 = Number of health stations per 10,000 population

y_7 = Number of health centres per 100,000 population

y_8 = Number of hospitals per 1,000,000 population

y_9 = Number of medical doctors per 100,000 population

y_{10} = Number of nurses per 10,000 population

y_{11} = Number of health assistants per 10,000 population

y_{12} = Primary school teachers per 1,000 pupils

y_{13} = Primary schools per 1,000 Pupils

y_{14} = Junior secondary school teachers per 1,000 pupils

y_{15} = Junior secondary schools per 1,000 pupils

y_{16} = Senior secondary school teachers per 1,000 pupils

y_{17} = Senior secondary schools per 1,000 pupils.

y_{18} = Number of agricultural extension workers per 1,000 rural population.

y_{19} = Per capita food crop production in quintals

The above list of socioeconomic variables are believed to gauge the relative level of development of regions. We will try to compare regions at an aggregate level, the sum total of all indicators, indicated above and at disaggregate level (i.e educational status, health facilities and other socioeconomic indicators). An attempt will be made to see the similarity of regions on the basis of changes made

between the years 1979 and 1989, by making use of the following ratios of indicators.

R_1 = The ratio of the 1989 urban population to the urban population of the 1979.

R_2 = The ratio of total population in 1989 to the total population of 1979.

R_3 = The ratio of length of roads in 1989 to the length of roads in 1979.

R_4 = The ratio of the number of hospital beds in 1989 to the number of hospital beds in 1979.

R_5 = The ratio of the number of health centres in 1989 to the number of health centres in 1979.

R_6 = The ratio of the number of health stations in 1989 to that of 1979.

R_7 = The ratio of the number of hospitals in 1989 to the number hospitals in 1979.

R_8 = The ratio of the number of medical doctors in the year 1989 to that of 1979.

R_9 = The ratio of the number of nurses in the year 1989 to the number of nurses in 1979.

R_{10} = The ratio of number of health assistants in the year 1989 to the number of health assistants in the year 1979.

R_{11} = The ratio of number of primary schools in the year 1989 to the number of primary schools in the year 1979.

R_{12} = The ratio of the number of primary school teachers in the year 1989 to that of the year 1979.

R_{13} = The ratio of the number primary school pupils in the year 1989 to that of 1979.

R_{14} = The ratio of the number of junior secondary schools in 1989 to that of 1979.

R_{15} = The ratio of the number of junior secondary school teachers in 1989 to that of the year 1979.

R_{16} = The ratio of the number of junior secondary school pupils in the year 1989 to that of 1979.

R_{17} = The ratio of the number of senior secondary schools in the year 1989 to senior secondary schools in 1979.

R_{18} = The ratio of the number of senior secondary school teachers in the year 1989 to that of the year 1979.

R_{19} = The ratio of the number of senior secondary school pupils in the year 1989 to that of the year 1979.

R_{20} = The ratio of area under major crops in the year 1989 to that of the year 1979.

R_{21} = The ratio of the quantity production of major crops in the year 1989 to that of the year 1979.

R_{22} = The ratio of number of post offices in the year 1989 to that of the year 1979.

R_{23} = The ratio of number of telephone lines in the year 1989 to that of 1979.

We shall also see the similarity in the changes at an aggregate level (the sum total of all ratios) and at disaggregate level (i.e, educational changes, changes in health facilities and changes in other variables). The name of the variables used in the description of the above procedures will be used in Chapter 4 and following Chapters of this paper.

3.2 METHODS OF DATA ANALYSIS

In this study Cluster Analysis will be utilized to achieve the required objectives. It is natural and procedural to give the definition of cluster before we embark on the actual technique. There is no universally accepted single definition of cluster. Some definitions include statements such as 'A cluster is a set of entities which are alike, and entities from different clusters are not alike.' Others propose that entities within a cluster are more similar to each other than the entities from other clusters [Everitt (1980)]. In the Oxford Dictionary of Current English cluster is defined as number of things of

the same kind growing closely together. However, in Norusis (1986), cluster analysis is defined as a statistical technique that attempts to identify similar groups of objects or cases based on a variety of attributes. In this paper, we will use cluster analysis to identify regions with similar socioeconomic characteristics.

The next step is to develop a quantitative scale on which to measure the association (similarity) between regions. The concept of distance and similarity are basic at this step. Distance is a measure of how far apart two cases are while similarity is a measure of their closeness.

As mentioned in Everitt (1980, P. 15) 'Blakith and Reyment (1971) conclude that the choice between correlations and distances measured in clustering is one which is difficult to make. Nevertheless the balance of evidence would seem to favour distances. ' However, in Johnson and Wichern (1992, P. 575) there is a subjective suggestion which proposes the following :

> **'When items (units or cases) are clustered, proximity is usually indicated by some sort of distance. On the other hand, variables are usually grouped on the basis of correlation coefficients or like measures of association. '**

Using the above suggestions we will favour distance measures to classify regions (cases) in this paper.

3.2.1 SOME SIMPLE MEASURES OF DISTANCE

There are many methods for estimating the distance between two cases. Some of them are discussed below.

Squared Euclidian distance : Probably this is the most commonly used distance measure. It is defined in Johnson and Wichern (1992) as follows :

The squared Euclidean distance between two p- dimensional observations (items) $X=[x_1, x_2, x_3, x_4 \ldots x_p]'$ and $Y=[y_1, y_2, y_3, y_4 \ldots y_p]'$ is

$$d^2(X, Y) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_p - y_p)^2$$
$$= (X - Y)'(X - Y) \qquad (1)$$

Euclidian distance

The second measure is Euclidian distance which is the square root of squared Euclidean distance. It is defined in Johnson and Wichern (1992) as :

$$d(X, Y) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_p-y_p)^2} \qquad (2)$$

City block or Manhattan distance

The distance between two cases is the sum of the absolute differences in values for each variable Norusis (1992).

$$d(X, Y) = \Sigma_i |X_i - Y_i| \qquad (3)$$

3.2.2 CLUSTERING TECHNIQUES

Clustering methods may be categorized broadly as being hierarchical or non-hierarchical. The hierarchical class is one in which every cluster obtained at any stage is a merger of clusters at the previous stages. In this case, therefore, it is possible to visualize not only the two extremes of clustering viz., n clusters with one unit per cluster (weak clustering) and a single cluster with all n units (strong clustering), but also a monotonically increasing strength of clustering as one goes from one level to another. In the non-hierarchical procedures, on the other hand, new clusters are obtained by both lumping and splitting of old clusters, and , although the two extremes of clustering are still the same, the intermediary stages of clustering do not have this natural monotone character of strength of clustering [Gnanadesikan, (1977)].

Our purpose is to classify regions according to their socioeconomic status. Moreover, we want to see whether the status of each region has undergone a change or not in the

time interval 1979 -1989. Therefore, our choice of clustering method will be hierarchical clustering since we need to compare the similarity structures in the socioeconomic performance of regions for the two specified years.

3.2.3 HIERARCHICAL CLUSTERING TECHNIQUES

Hierarchical clustering techniques proceed by either a series of successive mergers or a series of successive divisions.

Agglomerative hierarchical methods : These methods proceed by a series of successive fusions of the N entities into groups. In this method there are initially as many clusters as entities. Most similar entities are first grouped, and these initial groups are merged according to their similarities. Finally, all subgroups are fused into a single cluster.

Divisive hierarchical methods: These methods work in the opposite direction. Here an initial single group of entities is divided into two subgroups such that the entities in one subgroup are far from the entities in the other. These subgroups are then further divided into

dissimilar subgroups. The process continues until there are as many subgroups as entities.

The results of both methods can be displayed in the form of a two-dimensional diagram known as a **dendrogram.** The dendrogram illustrates the mergers or divisions which have been made at successive levels. It constitutes a summary of information in the distance measure.

Both types of hierarchical technique may be seen as attempts to find the most efficient step in clustering entities. In our case we choose agglomerative hierarchical procedure, because divisive methods are only possible for very small sets of data.

3.2.4 AGGLOMERATIVE HIERARCHICAL CLUSTERING

In the literature available there are six agglomerative hierarchical clustering methods, namely :

1. Nearest Neighbour
2. Furthest Neighbour
3. Group Average
4. Centroid
5. Median
6. Ward's Sum of Squares Method

In general, the above six clustering procedures are divided into three groups known as linkage methods, error sum of squares or variance methods, and centroid methods. All of them are based on either a matrix of distances or a matrix of similarities between pairs of cases [Norusis (1986)].

Nearest Neighbour (Single Linkage)

This is one of the simplest methods for joining clusters. The first two cases combined are those with the smallest distance between them. Smallest distance refers to largest similarity.

First we should find the smallest distance in D, D = $\{d_{ik}\}$ and merge the corresponding entities, say A and B to get cluster (AB). The distances between (AB) and any other cluster C are given in Johnson and Wichern (1992)

$$d_{(AB)C} = \min\{d_{AC}, d_{BC}\} \qquad (1)$$

where d_{AC} and d_{BC} are the distances between the nearest neighbours of clusters A and C and clusters B and C, respectively.

Complete Linkage (Furthest Neighbour)

In this method the distance between two clusters is calculated as the distance between their two furthest points. That means, at each stage the distance between clusters is determined by the distance between the two elements, one from each cluster, that are most distant. Therefore, complete linkage ensures that all items in a cluster are within some maximum distance (or minimum similarity) of each other.

The algorithm starts by finding the minimum entry in D, $D= \{d_{ik}\}$ and merging the corresponding entities, such as A and B, to get cluster (AB).

The distances between (AB) and any other cluster C are given in Johnson and Wichern (1992) as

$$d_{(AB)C} = \max\{d_{AC}, d_{BC}\} \qquad (2)$$

where d_{AC} and d_{BC} are the distance between the most distant members of clusters A and C and clusters B and C, respectively.

Average Linkage

The average linkage between groups method defines the distance between two clusters as the average of the distances between all pairs of items where one member of a pair belongs to each cluster.

The average linkage algorithm proceeds in the following manner. We begin by searching the distance matrix $D = \{d_{ik}\}$

to find the nearest (most similar) entities, for example, A and B. These entities are merged to form cluster (AB). The distances between (AB) and any other cluster C are indicated in Johnson and Wichern (1992) as

$$d_{(AB)C} = (\Sigma_i \Sigma_k d_{ik}) / (N_{(AB)} N_C) \qquad (3)$$

where d_{ik} is the distance between entity i in cluster (AB) and entity k in the cluster C, and $N_{(AB)}$ and N_C are the number of items in clusters (AB) and C, respectively.

Centroid Method

This is a method in which the distance between groups is defined as the distance between the group centroid. The term centroid can be described in the following way [Green and Carroll (1978)] :

Let X be a pxp matrix, represented by

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{bmatrix}$$

where $x_i=[x_{i1}\ x_{i2}\ldots\ x_{ip}]$, then the row vector of means, $\bar{x}$, called the centroid is given by

$$\bar{X} = 1/p\ 1'X \qquad (4)$$

Where 1′ denotes a 1xp unit row vector.
In this method groups are depicted to lie in the Euclidian space and are replaced on formation by the coordinates of their centroid. The procedure depends on fusing groups according to the distance between their centroids, the group with the smallest distance being fused first.

The undesirable property of the centroid method is that the distance at which clusters are combined can actually decrease from one step to the next while clusters merged at later stages are more dissimilar than those merged at early stages.

Median Method

In the centroid method, the centroid of a merged cluster is a weighted combination of the centroids of the two individual clusters, where the weights are proportional to the sizes of the clusters. In the median method, the two clusters being combined are weighted equally in the computation of the centroid, regardless of the number of cases in each.

That means, for calculating the new distances d^2_{in} between any group g_i and the group g_r which is formed by fusing groups g_p and g_q is independent of the relative sizes of g_p and g_q.

For this method Hawkins (1982) gave

$$d^2_{in} = 1/2(d^2_{ip} + d^2_{iq}) - 1/4d^2_{pq} \qquad (5)$$

Ward's method

Ward (1963) suggested, a method in which the loss of information which results from grouping entities into clusters is quantified in terms of some appropriate functional relationship. According to Ward, any functional relation that an investigator selects to reflect the relative desirability of groupings is called an objective function. Ward also uses error sum of squares as an objective function.

For the univariate case the error sum of squares (E.S.S) is given by Everitt (1980) as

$$E.S.S = \Sigma^n_{i=1} x^2_i - 1/n(\Sigma x_i)^2 \qquad (6)$$

where x_i is the value of the i^{th} individual.

3.2.5 AN ALGORITHM FOR AGGLOMERATIVE HIERARCHICAL CLASSIFICATIONS

A common algorithm for the above six clustering procedures was developed by Wishart (1969). According to Wishart the method of analysis cited by Ward (1963), using an 'error sum of squares' as objective function, is compatible with a general classification scheme, developed by Lance and Williams (1967), by which five other classification procedures can be computed for a variable transformation of the triangular squared Euclidean distance matrix. The work of Wishart (1969) is summarized below.

In general, if X_{ijt} is the value of the jth variable for the ith point of cluster S_t, containing K_t points, and $\bar{X}_{jt}$ is the mean of the jth variable for S_t, then the error sum of squares for St is defined as

$$E_t=\sum_{i=1}^{k_t}\sum_{j=1}^{m}(X_{ijt}-\bar{X}_{jt})^2$$

which when expanded, becomes

$$E_t=\sum_{i=1}^{k_t}\sum_{j=1}^{m} X_{ijt}^2-k_t U_t^2 \qquad (1)$$

Note: U^2 denotes the inner product $U'U$, etc.

where $U'_t=(\bar{X}_{1t}, \bar{X}_{2t}, \ldots, \bar{X}_{mt})$ is the position vector of the mean for cluster S_t.

The value of the objective function E is the sum of the error sum of squares for each of the T clusters,

$$E=\sum_{t=1}^{T} E_t$$

and at the suggested fusion of clusters S_p and S_q, the increase in E is given by

$$I_{pq}=E_r- E_p -E_q ,$$

where E_r is the error sum of squares for the union set

$S_r=S_p \cup S_q$.

Thus, from Equation (1),

$$I_{pq}=\sum_{i=1}^{k_r}\sum_{j=1}^{m} X_{ijr}^2-k_r U_r^2-\sum_{i=1}^{k_p}\sum_{j=1}^{m} X_{ijp}^2+k_p U_p^2-\sum_{i=1}^{k_q}\sum_{j=1}^{m} X_{ijq}^2+K_q U_q^2$$

The sum of the squares X^2_{ij} cancel, and hence

$$I_{pq}=K_pU^2_p + K_qU^2_q-K_rU^2_r \qquad (2)$$

but $K^2_rU^2_r=(K_pU_p + K_qU_q)^2$

$=K^2_pU^2_p + k^2qU^2_q + 2K_pK_qU'_pU_q$

$=K^2pU^2_p+K^2_qU^2_q+K_pK_q(U^2_p+U^2_q-(U_p-U_q))^2$

which reduces to

$$U^2r=(K_p/K_r)U^2_p + (K_q/k_r)U^2_q- K_pK_q/k^2_r(U_p-U_q)^2 \qquad (3)$$

On substitution for U^2_r, Equation (2) becomes

$$I_{pq}=(K_pK_q/K_r)(U_p-U_q)^2.$$

But since $(U_p-U_q)^2=d^2pq$, the distance between the means of clusters S_p and S_q is

$$I_{pq}=(K_pK_q/K_r)d^2_{pq} \qquad (4)$$

and fusion occurs when I_{pq} is a minimum.

At the fusion $S_r=S_p \cup S_q$, the suggested fusion of any other cluster S_i with the new union S_r will result in an increase in the objective function of

$$I_{ir}=[K_iK_r/(K_i+K_r)]\ d^2_{ir} \qquad (5)$$

The distance between the means of S_i and S_r is given by

$$d^2_{ir}=(U_i-U_r)^2$$

which reduces, in the same manner as Equation (3) to

$$d^2_{ir}=(K_p/K_r)\,d^2_{ip}+(K_p/K_r)\,d^2_{iq}-(K_pK_q/K^2_r)\,d^2_{pq}\ . \qquad (6)$$

On substitution for d^2ij in terms of K_i, K_j and I_{ij} from Equation (5), Equation (6) after manipulation becomes,

$$I_{ir}=1/(K_r+K_i)\,[(K_i+K_p)\,I_{ip}+(K_i+K_q)\,I_{iq}-K_iI_{pq}]\ . \qquad (7)$$

The algorithm

If the triangular matrix of all inter-point squared Euclidean distances $D=\{d^2_{ij};\ i=1,\ j-1;\ j=2,\ n\}$ is calculated and stored, then the increase in the objective function which results from the fusion of any two single- element clusters S_p, S_q is

from Equation (5)

$$I_{pq}=1/2\ d^2_{pq}$$

The first fusion therefore concerns those two points S_p and S_q for which d^2_{pq} is a minimum.

If, at the union $S_r=S_p \cup S_q$, the elements

$\{ d^2_{ip};\ i=1,\ n;\ i \neq p,\ q \}$, of the matrix D, are replaced by

$$d^2_{ip}=2I_{ip}. \qquad (8)$$

then these new values are consistent with the original distances in D; that is, Equation (8) holds for all $\{ d^2_{ij};\ i,\ j \neq q \}$

By replacing the cluster S_p with the union $S_p \cup S_q$, cluster S_q becomes inactive, and therefore the elements of the qth column and row of D are redundant.

Equation (8) becomes, after substitution for I_{ip} from Equation (7),

$$d^2_{ip}=2/(K_i+K_r)\,[\,(K_i+K_p)\,I_{ip}+(k_i+K_q)\,I_{iq}-k_i I_{pq}]$$

$$= 1/(k_i+K_r)\,[\,(K_i+K_q)\,d^2_{ip}+(K_i+K_q)\,d^2_{iq}-K_i d^2_{pq}]\ . \qquad (9)$$

Transformations similar to (9) have been derived by Lance and Williams (1967) for clustering processes: Nearest Neighbour, Furthest Neighbour, Median, Group Average and Centroid. The general form of the transformation is

$$d^2_{ir}=\alpha_p d^2_{ip}+\alpha_q d^2_{iq}+\beta d^2_{pq}+\gamma\,|d^2_{ip}-d^2_{iq}|$$

from which the six methods are obtained using the following parameter values.

Nearest Neighbour: $\alpha_p=\alpha_q=1/2$; $\beta=0$; $\gamma=-1/2$

Furthest Neighbour: $\alpha_p=\alpha_q=1/2$; $\beta=0$; $\gamma=1/2$

Median: $\alpha_p=\alpha_q=1/2$; $\beta=-1/4$; $\gamma=0$

Group Average: $\alpha_p=K_p/K_r$; $\alpha_q=K_q/K_r$; $\beta=\gamma=0$

Centroid: $\alpha_p=K_p/K_r$; $\alpha_q=K_q/K_r$; $\beta=-\alpha_p\alpha_q$; $\gamma=0$

Ward's Method : $\alpha_p=(K_i+K_p)/(K_i+K_r)$; $\alpha_q=(K_i+K_q)/(K_i+K_r)$;

$\beta=-K_i/(K_i+K_r)$; $\gamma=0$.

Thus all six methods are obtained from the same algorithm by varying the transformation parameters α_p, α_q, β and γ ,
thereby making the computer implementation of agglomerative methods relatively easy.

3.3 CHOICE OF PARTICULAR METHOD FOR A GIVEN PROBLEM.

The major difficulty of agglomerative hierarchical techniques lies in the choice of one particular method from

the many available. Jardine and Sibson (1971) provide a strong argument that Single Linkage is the only agglomerative hierarchical technique to satisfy the various mathematical criteria proposed by them. Although their presentation was accepted to be mathematically superior, several other authors have criticized the Single Link method from the practical point of view. Most of these authors have made empirical investigations which give useful indications of which methods are likely to be useful in practice. For instance, Kuiper and Fisher (1975) consider the above six hierarchical techniques and find that for equal number of points from multivariate normal distributions, Ward's method classifies almost as well as linear discriminant function knowing all the parameters. With unequal sample size they found that Centroid, Group Average and Furthest Neighbour are more successful. According to their investigation Nearest Neighbour does not behave well under a 'nice' situation considered. Baker (1974) produces evidence that Complete Link clustering is better than Single Linkage in being less sensitive to data errors. According to him such an advantage would appear to have practical significance to a person selecting a grouping method. Hubert (1974) also suggest that Complete Link clustering is superior to Single Linkage for practical purpose.

The results of some empirical investigations which were summarized by Everitt (1979) are given below.

(1) No single method is best in every situation.

(2) The mathematically respectable Single Linkage is, in most cases, the least successful for the data used.

(3) Group Average clustering and a method due to Ward (1963) do fairly well overall.

Such empirical studies can, of course, never afford a complete evaluation of clustering methods. Therefore, from the above summary we can see that the question as to which clustering method should be used is one to which there can be no absolute answer. As a remedy for this, we will use all agglomerative hierarchical methods on the same data and clusters that will be produced by all or by the majority of the methods may be accepted. We will also use squared Euclidean distance because this measure should be used with Centroid, Median and Ward's methods [Norusis (1986)]. This distance has the disadvantage that it depends on unit of measurement for the variables. The second disadvantage of this distance is that when variables are measured on different scales, variables that are measured in larger numbers will contribute more to the distance than the variables that are recorded in smaller numbers. To circumvent these problems we will express all variables in the standardized form and use these in the process of clustering.

CHAPTER IV

RESULTS AND ANALYSES OF THE 1979 SOCIOECONOMIC DATA.

4.1. ANALYSES OF THE 1979 DATA AT AN AGGREGATE LEVEL.

Table 4.1 presents the means, standard deviations, coefficient of variations, minimums and maximums of the 19 socioeconomic variables considered in the analysis. From the Table we can easily observe that the distribution of socioeconomic characteristics shows a large gap between the minimum and maximum values of the variables. This could be an indication of variation in the regional distribution of variables considered in the analysis.

Table 4.1:- SOME DESCRIPTIVE MEASURES OF SOCIOECONOMIC VARIABLES IN 1979.

Variable	Mean	Std.Dev	Coeff.-of variation	Minimum	Maximum
Y_1	2.14	1.33	62.14	1.25	6.07
Y_2	10.57	6.49	61.40	1.75	28.36
Y_3	8.96	5.98	66.74	4.53	27.40
Y_4	11.2	17.21	153.66	2.54	64.81
Y_5	0.14	0.03	21.43	0.08	0.18
Y_6	0.50	0.17	34.00	0.28	0.87
Y_7	0.05	0.02	40.00	0.03	0.08
Y_8	2.17	0.81	37.33	1.17	3.87
Y_9	0.92	1.04	113.04	0.33	4.16
Y_{10}	0.34	0.27	79.41	0.21	1.18
Y_{11}	1.46	0.55	37.67	0.88	3.02
Y_{12}	25.74	8.62	33.49	12.93	38.93
Y_{13}	4.34	1.78	41.01	2.27	8.03
Y_{14}	27.53	6.85	24.88	18.89	42.14
Y_{15}	4.35	1.37	31.49	2.48	6.89
Y_{16}	23.72	9.28	39.12	12.94	50.39
Y_{17}	1.17	0.48	41.03	0.71	2.24
Y_{18}	101.65	44.50	43.78	0.00	155.96
Y_{19}	2.49	1.13	45.38	0.76	4.11

Table 4.2. shows the squared Euclidean dissimilarity coefficient matrix for all possible pairs of the 12 regions, based on the standardized socioeconomic variables for the year 1979. The first entry in Table 4.2 is the distance between case 1 and case 2, Arsi and Bale.

Table 4.2:- SQUARED EUCLIDEAN DISSIMILARITY COEFFICIENT MATRIX (1979)

Case	1	2	3	4
2	17.1577			
3	47.5360	27.3350		
4	12.3150	15.0369	42.0432	
5	24.4969	13.2818	42.5680	12.3053
6	29.7069	21.9072	61.7300	28.8327
7	30.9358	21.4239	24.7491	27.0487
8	100.6979	116.9716	134.0784	95.9933
9	14.7038	15.8014	42.8064	17.0959
10	25.2318	10.9233	45.8625	13.9721
11	14.0702	20.3635	45.8532	17.3289
12	13.2426	11.7166	45.5847	12.1520

Case	5	6	7	8
6	26.0679			
7	27.7612	41.5478		
8	107.0810	99.6184	98.5261	
9	27.8661	32.6365	35.9715	108.0594
10	11.7177	18.3472	35.7065	110.3016
11	33.7660	26.6901	15.3330	89.2633
12	10.1481	18.2280	20.0955	87.7186

Case	9	10	11
10	18.4880		
11	22.2327	29.8116	
12	15.5828	11.9194	12.6535

The smallest distance in the matrix is 10.1481. This shows that the most similar regions for that particular year were Gonder and Welo. The results of six clustering procedures which are based on dissimilarity matrix in Table 4.2 are presented below.

4.1.1. RESULTS OF SINGLE LINKAGE METHOD

We usually stop the agglomeration as soon as the increase between two adjacent steps becomes too large. Taking the step which occurs before the increase in distance between two adjacent steps becomes too large, the Single Linkage method gave four distinct groups of regions each of which is relatively more homogeneous in its socioeconomic characteristics. These groups were:

Group 1: Arsi, Bale, Gojam, Gonder,
Illubabor and Kefa, Sidamo,
Tigray, Welega and Welo

Group 2 : Gamogofa

Group 3 : Harerghe

Group 4 : Shewa

According to this method Shewa is on the way not to join the group of other regions being at a large distance from them. The dendrogram and the agglomeration schedule which illustrate the results produced by single linkage method are given in Fig. 4.1 and Table 4.3, respectively.

Table 4.3:- AGGLOMERATION SCHEDULE USING SINGLE LINKAGE

Clusters Combined				Stage 1st Appears	Cluster
Stage	Clus.1	Clus.2	Coefficient	Clus.1	Clus.2
1	5	12	10.148131	0	0
2	2	10	10.923328	0	0
3	2	5	11.716643	2	1
4	2	4	12.151998	3	0
5	1	2	12.315047	0	4
6	1	11	12.653512	5	0
7	1	9	14.703808	6	0
8	1	7	15.333023	7	0
9	1	6	18.227989	8	0
10	1	3	24.749121	9	0
11	1	8	87.718590	10	0

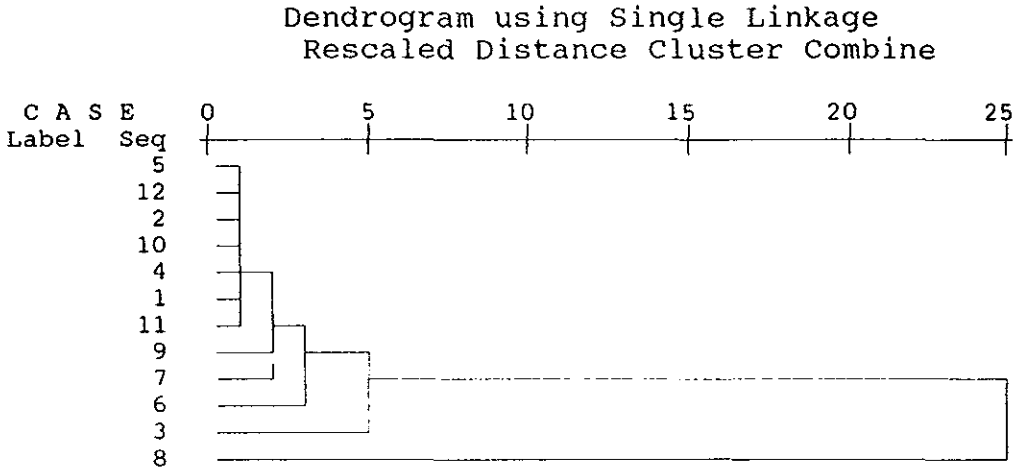

FIGURE 4.1

4.1.2 RESULTS OF COMPLETE LINKAGE METHOD

It appears that the seven cluster solution may be appropriate since it occurs before the distances at which clusters are combined become too large. The seven groups of regions were :

Group 1: Arsi and Welega

Group 2: Bale, Gojam, Gonder, Tigray and Welo

Group 3: Gamogofa

Group 4: Harerghe

Group 5: Illubabor and Kefa

Group 6: Shewa

Group 7: Sidamo

The agglomeration schedule and the dendrogram are given in Table 4.4 and Fig. 4.2, respectively. The results of this method suggest that Shewa was at a far distance from other regions in the year 1979.

TABLE 4.4 AGGLOMERATION SCHEDULE USING COMPLETE LINKAGE

	Clusters Combined			Stage Cluster 1st Appears	
Stage	Clus 1	Clus 2	Coefficient	Clus 1	Clus 2
1	5	12	10.148131	0	0
2	2	10	10.923328	0	0
3	4	5	12.305278	0	1
4	1	11	14.070157	0	0
5	2	4	15.036949	2	3
6	1	9	22.232718	4	0
7	3	7	24.749121	0	0
8	2	6	28.832743	5	0
9	1	2	33.765965	6	8
10	1	3	61.730022	9	7
11	1	8	134.078354	10	0

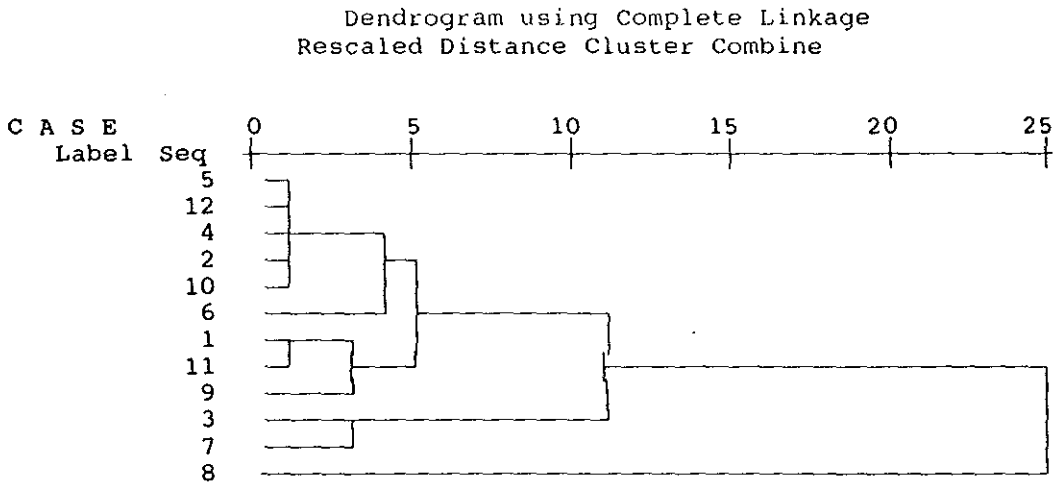

FIGURE 4.2

4.1.3 RESULTS OF GROUP AVERAGE METHOD

The agglomeration schedule and dendrogram show the existence of five groups of regions each of which is relatively homogeneous in its socioeconomic characteristics considered in this particular paper. These groups were :

Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray and Welo

Group 2 : Gamogofa

Group 3 : Harerghe

Group 4 : Illubabor and Kefa and Welega

Group 5 : Shewa

These results suggest that Shewa was far from other regions in its socioeconomic characteristics for that particular year. The agglomeration schedule and dendrogram illustrating the above results are given in Table 4.5 and Fig. 4.3, respectively.

Table 4.5 AGGLOMERATION SCHEDULE USING AVERAGE LINKAGE (BETWEEN GROUPS)

Stage	Clusters Combined			Stage Cluster 1st Appears	
	Clus.1	Clus.2	Coefficient	Clus.1	Clus.2
1	5	12	10.148131	0	0
2	2	10	10.923328	0	0
3	2	5	12.158905	2	1
4	1	4	12.315047	0	0
5	7	11	15.333023	0	0
6	1	9	15.899858	4	0
7	1	2	17.611139	6	3
8	1	7	24.940674	7	5
9	1	6	27.107138	8	0
10	1	3	42.606815	9	0
11	1	8	104.391792	10	0

Dendrogram using Average linkage (Between Groups)

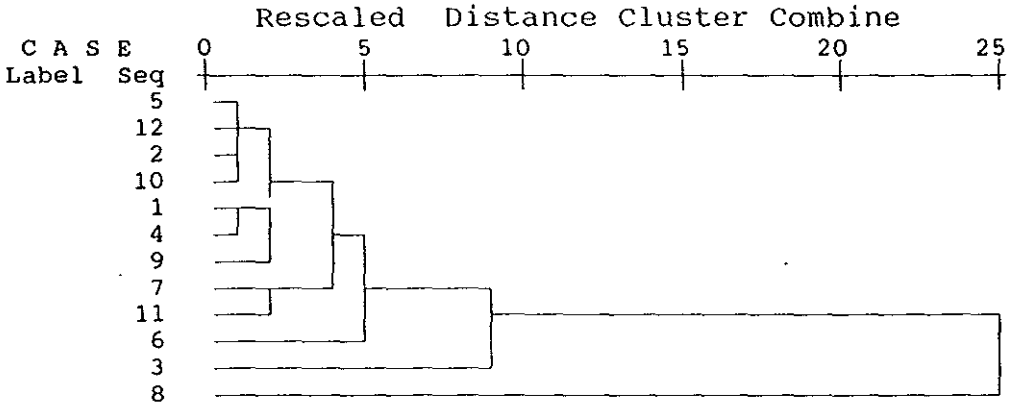

FIGURE 4.3

4.1.4 RESULTS OF CENTROID METHOD

The agglomeration schedule and the dendrogram indicate that there were basically five distinct groups of regions in 1979. The groups were :

Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray, Welega and Welo

Group 2 : Gamogofa

Group 3 : Harerghe

Group 4 : Illubabor and Kefa

Group 5 : Shewa

The results of this method show that Shewa was far from other regions being followed by Gamogofa. Table 4.6 indicates the agglomeration schedule for this method. The dendrogram is presented in Fig. 4.4. The results obtained

by this method are roughly the same as the results obtained by Group Average method.

Table 4.6 AGGLOMERATION SCHEDULE USING CENTROID METHOD

Stage	Clusters Combined Clus. 1	Clus. 2	Coefficient	Stage Cluster 1st Appears Clus.1	Clus.2
1	5	12	10.148131	0	0
2	5	10	9.281541	1	0
3	2	5	8.220016	0	2
4	2	4	9.009892	3	0
5	1	2	13.561861	0	4
6	1	9	12.266979	5	0
7	1	11	14.825083	6	0
8	1	6	17.874216	7	0
9	1	7	20.056999	8	0
10	1	3	33.270592	9	0
11	1	8	93.154678	10	0

Dendrogram using Centroid Method

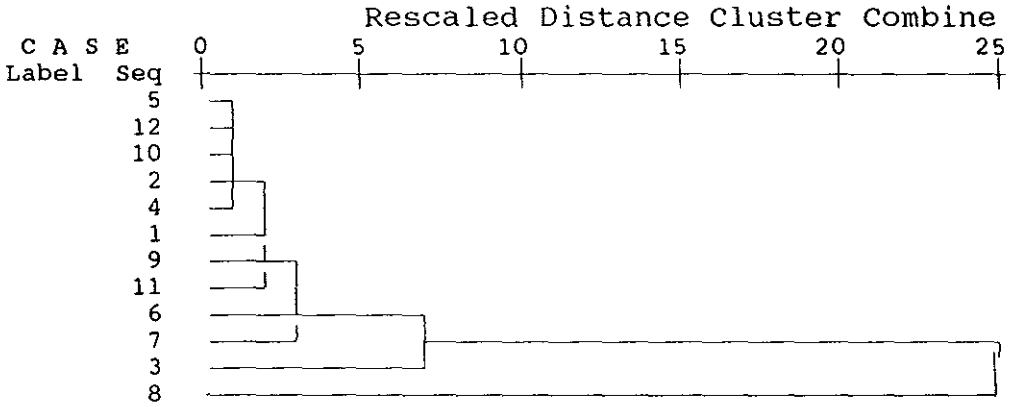

FIGURE 4.4

4.1.5 RESULTS OF MEDIAN METHOD

The Median method was applied to the squared Euclidean distances between the twelve regions and produced the dendrogram in Fig. 4.5 and the agglomeration schedule in Table 4.7. The schedule and the dendrogram show the

existence of six distinct groups of regions. The groups were:

Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray and Welo

Group 2 : Gamogofa

Group 3 : Harerghe

Group 4 : Illubabor and Kefa

Group 5 : Shewa

Group 6 : Welega

The results of this method are roughly the same as the results obtained by the Centroid and Group Average method. The only difference lies on group membership of Welega.

Table 4.7 AGGLOMERATION SCHEDULE USING MEDIAN METHOD

Clusters combined				Stage Cluster 1st appears	
Stage	Clus.1	Clus.2	Coefficient	Clus.1	Clus.2
1	5	12	10.148131	0	0
2	5	10	9.281541	1	0
3	2	5	8.122381	0	2
4	2	4	10.243616	3	0
5	1	2	11.486215	0	4
6	1	9	11.006073	5	0
7	1	11	14.423992	6	0
8	1	7	17.559364	7	0
9	1	3	27.084023	8	0
10	1	6	37.805550	9	0
11	1	8	93.069595	10	0

Dendrogram using Median Method

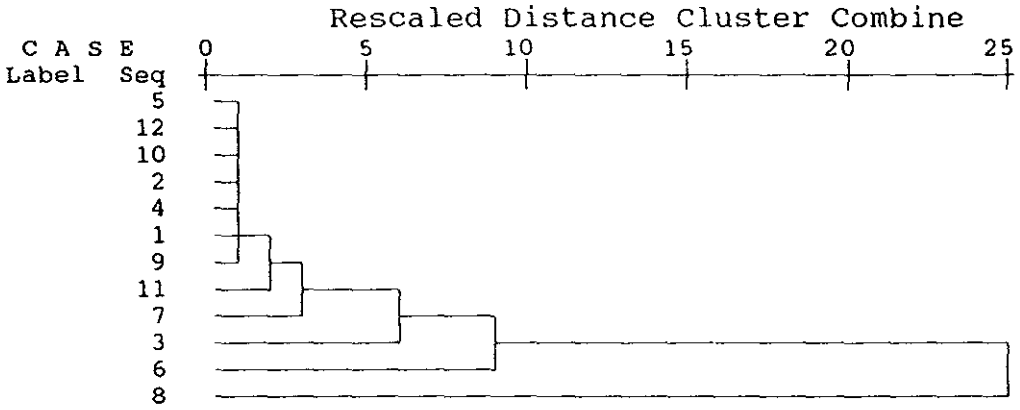

FIGURE 4.5

4.1.6 RESULTS OF WARD'S METHOD

Table 4.8 shows the agglomeration schedule using Ward's method. The dendrogram obtained by this method is displayed in Fig. 4.6 . Ward's method gave eight distinct groups of regions each of which presumably is relatively more homogeneous in its socioeconomic status. These groups were

Group 1: Arsi, Gojam

Group 2: Bale, Gonder, Tigray and Welo

Group 3: Gamogofa

Group 4: Harerghe

Group 5: Illubabor and Kefa

Group 6: Shewa

Group 7: Sidamo

Group 8: Welega

According to this method Shewa was at a large distance from other regions in 1979.

Table 4.8:- AGGLOMERATION SCHEDULE USING WARD'S METHOD

Stage	Clusters Combined Clus.1	Clus. 2	Coefficient	Stage Cluster 1st Appears Clus.1	Clus.2
1	5	12	5.074066	0	0
2	2	10	10.535730	0	0
3	1	4	16.693254	0	0
4	2	5	23.584295	2	1
5	7	11	31.250807	0	0
6	1	9	39.798203	3	0
7	2	6	53.222900	4	0
8	1	2	70.276749	6	7
9	3	7	91.255363	0	5
10	1	3	123.608215	8	9
11	1	8	209.000000	10	0

Dendrogram using Ward's Method

Rescaled Distance Cluster Combine

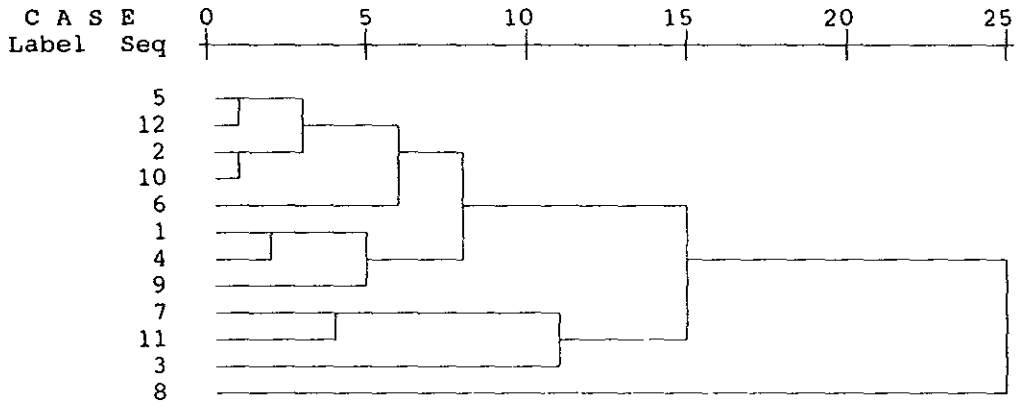

FIGURE 4.6

Table 4.9:- SUMMARY OF THE RESULTS FOR SIX CLUSTERING PROCEDURES (1979).

CLUSTERING METHOD	NUMBER OF CLUSTERS	GROUP MEMBERSHIP
Single-Linkage	4	Group 1 : Arsi, Bale, Gojam , Gonder, Illubabor & Kefa, Sidamo, Tigray, Welega and Welo Group 2 : Gamogofa Group 3 : Harerghe Group 4 : Shewa
Complete-Linkage	7	Group 1 : Arsi and Welega Group 2 : Bale, Gojam, Gonder, Tigray and Welo Group 3 : Gamogofa Group 4 : Harerghe Group 5 : Illubabor & Kefa Group 6 : Shewa Group 7 : Sidamo
Group Average	5	Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray and Welo Group 2 : Gamogofa Group 3 : Harerghe Group 4 : Illubabor and Kefa and Welega Group 5 : Shewa
Centroid	5	Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray and Welo Group 2 : Gamogofa Group 3 : Harerghe Group 4 : Illubabor and Kefa and Welega Group 5 : Shewa
Median	6	Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray and Welo Group 2 : Gamogofa Group 3 : Harerghe Group 4 : Illubabor and Kefa Group 5 : Shewa Group 6 : Welega

CLUSTERING METHOD	NUMBER OF CLUSTERS	GROUP MEMBERSHIP
Ward's	8	Group 1 : Arsi and Gojam Group 2 : Bale, Gonder, Tigray and Welo Group 3 : Gamogofa Group 4 : Harerghe Group 5 : Illubabor and Kefa Group 6 : Shewa Group 7 : Sidamo Group 8 : Welega

As Table 4.9 shows, three of the six methods gave nearly the same results. These three methods are the Centroid, Median and Group Average methods. The results of these three methods are roughly the same as the result obtained by the Single Linkage method. The only difference lies in the placement of Welega and Illubabor and Kefa, in that Single Linkage method includes Illubabor and Kefa in group one, while two of the three methods put Welega and Illubabor and Kefa in a separate group. The results obtained by Ward's method resemble the results of Complete Linkage method. The two methods gave the same results for groups 3 up to 7 and differ in placing Gojam and Welega for other groups.

Table 4.9 reveals that there was a considerable regional disparity based on the socioeconomic characteristics in 1979. It would also give an interesting conclusion that, on the basis of the 19 variables considered to reflect the socioeconomic status, the 12 regions are grouped neither strictly along ethnic or

linguistic lines nor on the basis of entirely geographical proximity.

In this section we have observed that the existence of regional disparity was quite obvious on the basis of the 19 socioeconomic variables in 1979. It is sensible to consider similarity at disaggregate level to see whether the existing disparity is more severe for educational variables, health variables or for variables other than educational or health services. In succeeding sections of this Chapter we will apply the six clustering procedures to each of the following groups of variables.

I. EDUCATION : Y_{12} TO Y_{17}

II. HEALTH : Y_1, Y_6 TO Y_{11}

III. OTHERS : Y_2 TO Y_5, Y_{18} and Y_{19}

4.2 SIMILARITY OF REGIONS ON THE BASIS OF EDUCATIONAL VARIABLES IN 1979.

The six clustering procedures were used to classify regions according to their homogeneity based on educational services. The results are given in Table 4.10.

Table 4.10 SUMMARY RESULTS OF SIX CLUSTERING METHODS BASED ON EDUCATIONAL VARIABLES (1979)

CLUSTERING METHOD	NUMBER OF CLUSTERS	GROUP MEMBERSHIP
Single-Linkage	4	Group 1 : Arsi, Bale, Gojam, Gonder, Shewa, Sidamo, Tigray, Welega and Welo Group 2 : Gamogofa Group 3 : Harerghe Group 4 : Illubabor and Kefa
Complete-Linkage	5	Group 1 : Arsi, Shewa and Sidamo Group 2 : Bale, Gojam, Gonder, Tigray and Welo Group 3 : Gamogofa Group 4 : Harerghe Group 5 : Illubabor and Kefa
Group Average	6	Group 1:Arsi, Bale, Gojam, Shewa, Sidamo and Welo Group 2 : Gamogofa Group 3 : Gonder and Tigray Group 4 : Harerghe Group 5 : Illubabor and Kefa Group 6 : Welega
Centroid	6	Group 1 : Arsi, Bale, Gojam, Shewa, Sidamo and Welo Group 2 : Gamogofa Group 3 : Gonder and Tigray Group 4 : Harerghe Group 5 : Illubabor and Kefa Group 6 : Welega

CLUSTERING METHOD	NUMBER OF CLUSTERS	GROUP MEMBERSHIP
Median	6	Group 1 : Arsi, Bale, Gojam, Shewa, Sidamo and Welo Group 2 : Gamogofa Group 3 : Gonder and Tigray Group 4 : Harerghe Group 5 : Illubabor & Kefa Group 6 : Welega
Ward's	7	Group 1 : Arsi, Shewa and Sidamo Group 2 : Bale, Gojam and Welo Group 3 : Gamogofa Group 4 : Gonder and Tigray Group 5 : Harerghe Group 6 : Illubabor and Kefa Group 7 : Welega

Glancing at Table 4.10 , the Centroid, Median and Group Average methods gave the same results. The outcomes from the six clustering methods are roughly consistent with one another.

4.3 SIMILARITY OF REGIONS ON THE BASIS OF VARIABLES REPRESENTING HEALTH SERVICES IN 1979.

The pattern of homogeneity of regions based on health service indicators was examined by applying the six clustering procedures. The results are presented in Table 4.11 below. Looking at this Table, we can see that the results of Complete Linkage method are the same as that of Median method. Group Average and Centroid methods gave the same results.

The results of all but Single Linkage methods are roughly the same. This is an indication of the existence of natural grouping.

Table 4.11: SUMMARY RESULTS OF SIX CLUSTERING PROCEDURES BASED ON VARIABLES REPRESENTING HEALTH SERVICES (1979)

CLUSTERING METHOD	NUMBER OF CLUSTERS	GROUP MEMBERSHIP
Single-Linkage	2	Group 1 : Arsi, Bale, Gamogofa, Gojam, Gonder, Harerghe, Illubabor and Kefa, Sidamo, Tigray Welega and Welo Group 2 : Shewa
Complete-Linkage	4	Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray, Welega and Welo Group 2 : Gamogofa, Illubabor & Kefa Group 3 : Harerghe Group 4 : Shewa

CLUSTERING METHOD	NUMBER OF CLUSTERS	GROUP MEMBERSHIP
Group-Average	3	Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Harerghe, Tigray, Welega and Welo Group 2 : Gamogofa, Illubabor and Kefa Group 3 : Shewa
Centroid	3	Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Harerghe, Tigray, Welega and Welo Group 2 : Gamogofa, Illubabor and Kefa Group 3 : Shewa
Median	4	Group 1 : Arsi, Bale, Gojam, Gonder, Sidamo, Tigray, Welega and Welo Group 2 : Gamogofa and Illubabor and Kefa Group 3 : Harerghe Group 4 : Shewa
Ward's	5	Group 1 : Arsi, Gojam, Sidamo, Tigray, Welo and Welega Group 2 : Bale and Gonder Group 3 : Gamogofa and Illubabor and Kefa Group 4 : Harerghe Group 5 : Shewa

4.4 SIMILARITY OF REGIONS ON THE BASIS OF OTHER SOCIOECONOMIC VARIABLES IN 1979

The variables other than educational and health service indicators were also considered. Based on these variables we have tried to classify the Ethiopian regions. Again the six clustering procedures were employed to see the structure of similarity among regions. The results of six clustering procedures are presented in Table 4.12. From this we can see that Single-Linkage ,Centroid, Group Average and Median methods gave the same results. Ward's and Complete-Linkage methods gave results which are different from the other four methods.

To see the sector for which disparity was more severe in 1979, the distances at which clusters combined were compared at each stage of cluster formation for three groups of variables using the six clustering procedures. The computed results are presented in Tables 4.13- 4.15.

TABLE 4.12 SUMMARY RESULTS OF SIX CLUSTERING PROCEDURES BASED ON OTHER VARIABLES (1979)

CLUSTERING METHOD	NUMBER OF CLUSTERS	CLUSTER MEMBERSHIP
Single-Linkage	3	Group 1 :Arsi Group 2 :Bale, Gamogofa,Gojam, Gonder, Welo,Harerghe, Illubabor and Kefa, Tigray, Welega and Sidamo Group 3 :Shewa

CLUSTERING METHOD	NUMBER OF CLUSTERS	CLUSTER MEMBERSHIP
Complete-Linkage	6	Group 1 : Arsi, Gojam, Sidamo, Tigray, Welega and Welo Group 2 : Bale Group 3 : Gamogofa and Illubabor & Kefa Group 4 : Gonder Group 5 : Harerghe group 6 : Shewa
Group Average	3	Group 1 : Arsi Group 2 : Bale, Gamogofa, Gojam, Gonder, Welo,Harerghe, Illubabor and Kefa, Sidamo Tigray and Welega Group 3 : Shewa
Centroid	3	Group 1 : Arsi Group 2 : Bale, Gamogofa, Gojam, Gonder, Welo,Harerghe, Illubabor and Kefa, Sidamo, Tigray and Welega Group 3 : Shewa
Median	3	Group 1 : Arsi Group 2 : Bale, Gamogofa, Gojam, Gonder, Welo, Harerghe, Illubabor and Kefa, Sidamo, Tigray and Welega Group 3 : Shewa
Ward's	5	Group 1 : Arsi Group 2 : Bale, Gamogofa, Tigray and Sidamo Group 3 : Gojam Group 4 : Gonder, Harerghe, Illubabor and Kefa, Welega and Welo group 5 : Shewa

TABLE 4.13: THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF VARIABLES IN SINGLE-LINKAGE AND COMPLETE- LINKAGE METHODS (1979).

	Single -Linkage			Complete-Linkage		
Stage	I	II	III	I	II	III
1	1.119	.418038	.714842	1.119	.41804	.7148
2	1.184	.60425	1.35703	1.649	.63435	1.4877
3	1.649	.681838	1.45543	2.713	.68184	1.5337
4	2.204	.719349	1.53375	2.799	.95445	1.9674
5	2.430	.852487	1.73802	3.576	2.7791	3.1054
6	3.576	2.08443	1.99018	6.608	2.9782	3.1589
7	4.385	2.81198	2.12064	8.725	4.3894	6.0728
8	4.566	2.97818	3.56565	13.57	6.6538	8.6827
9	8.381	3.30761	3.68719	15.33	10.669	15.433
10	9.116	3.96137	7.2171	24.07	22.145	19.433
11	11.36	40.3023	27.2556	34.91	63.106	52.5491

TABLE 4.14 THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF VARIABLES IN GROUP-AVERAGE AND CENTROID METHODS (1979)

Stage	Group-Average method			Centroid method		
	I	II	III	I	II	III
1	1.119	0.418	0.7148	1.1193	0.418	.7148
2	1.649	0.6193	1.4877	1.6490	0.5148	1.4877
3	1.992	0.6818	1.5337	1.7122	0.6818	1.3122
4	2.459	0.8369	1.8527	2.0467	0.6664	1.5337
5	3.576	1.6285	2.2778	3.497	1.1827	1.8769
6	4.793	2.9782	2.5478	3.5763	2.9781	2.1644
7	8.725	3.9174	4.7525	7.7233	3.3988	4.0158
8	10.23	4.1359	6.4531	9.1554	3.2753	5.19663
9	13.95	6.3679	7.6958	9.8572	5.2566	5.6199
10	17.16	10.9043	11.025	13.0355	8.6528	8.4508
11	18.82	54.7886	38.575	13.9889	52.133	35.5361

TABLE 4.15 THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF VARIABLES IN MEDIAN AND WARD'S METHODS (1979)

Stage	Median method			Ward's method		
	I	II	III	I	II	III
1	1.119	.418	0.7148	0.559	0.2090	0.3574
2	1.649	.5148	1.4877	1.384	0.5499	1.1013
3	1.712	.6818	1.3122	2.525	0.8931	1.8681
4	2.046	.6664	1.5337	3.890	1.3374	2.9841
5	2.871	1.4464	1.8769	5.678	2.8265	4.4270
6	3.576	2.9782	2.1644	10.04	4.6006	6.1248
7	7.367	3.4431	3.9863	14.45	6.7953	9.1367
8	7.636	3.1262	5.7345	22.01	10.379	13.478
9	12.53	6.7363	5.8578	31.00	15.052	21.147
10	12.61	8.5191	9.6452	42.48	29.211	33.425
11	16.79	46.3844	39.5966	66	77	66

The values given in Tables 4.13 - 4.15 are the distances at which clusters are combined at each stage of clustering. By examining these values we can get an idea of

how unlike the clusters being combined are. Small values indicate that fairly homogeneous clusters are being merged. Large values indicate that clusters containing quite dissimilar members are being combined. Therefore, the group of variables which has large values for all stages or for the majority of stages will be considered as a group for which disparity is severe. The results given in the specified Tables indicate that the distances for group I variables are larger than those for the remaining two groups in almost all of the cases. This is an indication that the disparity was more severe in the distribution of educational services as compared to health and other group of variables in 1979 .

CHAPTER V

RESULTS AND ANALYSES OF THE 1989 SOCIOECONOMIC DATA

Analyses of the 1979 data represent an examination of socioeconomic characteristics of regions in a fairly remote past. Analyses of the same sort for the year 1989 are expected to provide some interesting comparisons. It may help to answer the question as to whether the similarity of regions in 1979 is consistent with the similarity after ten years or not. It may also help us to know whether the patterns of regional disparity in 1979 are quite different from that of 1989. The results obtained from the 1989 socioeconomic data are discussed in the following sections.

5.1. ANALYSES OF THE 1989 DATA AT AN AGGREGATE LEVEL

The descriptive measures, mean, standard deviation, coefficient of variation, minimum and maximum values for each of the 19 socioeconomic variables in the year 1989 are indicated in Table 5.1. From this we can see that there is a wide difference between the minimum and maximum values of the socioeconomic variables. This could be an indication of variation in the regional distribution of socioeconomic services in the year 1989.

TABLE 5.1 SOME DESCRIPTIVE MEASURES OF SOCIOECONOMIC VARIABLES IN THE YEAR 1989

Variable	Mean	Std Dev	Coeff- of variation	Minimum	maximum
Y_1	1.62	0.79	48.77	0.81	3.73
Y_2	12.05	7.72	64.07	1.81	32.73
Y_3	7.95	4.68	58.87	4.43	22.2
Y_4	14.63	24.57	167.94	4.65	92.45
Y_5	0.17	0.13	76.47	0.08	0.57
Y_6	0.49	0.20	40.82	0.31	0.86
Y_7	0.04	0.01	25.00	0.02	0.07
Y_8	1.58	0.75	47.47	0.86	3.48
Y_9	1.32	0.94	71.21	0.69	4.14
Y_{10}	0.48	0.19	39.58	0.31	1.01
Y_{11}	0.03	0.01	33.33	0.02	0.06
Y_{12}	25.03	5.34	21.33	18.19	38.98
Y_{13}	4.12	2.11	51.21	2.01	9.7
Y_{14}	24.29	2.83	11.65	19.96	27.55
Y_{15}	2.38	0.67	28.15	1.7	3.82
Y_{16}	25.86	2.41	9.32	22.62	29.42
Y_{17}	0.78	0.20	25.64	0.48	1.1
Y_{18}	63.50	26.90	42.36	29.86	118.83
Y_{19}	1.54	0.79	51.30	0.56	3.67

Table 5.2 indicates the squared Euclidean dissimilarity coefficient matrix for all possible pairs of the 12 regions

based on the standardized socioeconomic variables for the year 1989.

TABLE 5.2: SQUARED EUCLIDEAN DISSIMILARITY COEFFICIENT MATRIX (1989)

Case	1	2	3	4
2	19.0764			
3	41.6439	34.1058		
4	9.2207	19.8312	49.8556	
5	47.2713	18.9272	50.8331	36.9034
6	49.6397	27.7058	37.9638	36.9430
7	43.4801	22.6240	47.5564	36.3308
8	85.1933	88.3247	98.2615	69.3145
9	38.5901	16.5533	49.5915	21.3989
10	51.9931	25.3272	49.3779	38.0658
11	27.0959	14.9864	24.9324	24.5399
12	23.6922	13.9319	37.4007	14.9219
Case	5	6	7	8
6	23.2381			
7	33.8942	22.9825		
8	92.2067	81.3544	78.3479	
9	20.9611	22.5902	37.3237	88.5099
10	12.1997	23.5621	40.9894	88.7553
11	26.8985	18.6223	18.8776	68.9982
12	23.7808	14.1460	20.8062	81.4746
Case	9	10	11	
10	20.6842			
11	23.8792	30.1788		
12	8.7492	23.0056	17.5781	

The smallest distance in the matrix is 8.749234. This shows that the most similar regions for the year 1989 were Welo and Sidamo. The results of six hierarchical clustering procedures based on the above dissimilarity matrix are summarized in the following sections.

5.1.1 RESULTS OF SINGLE LINKAGE METHOD

The Single Linkage method gave five distinct groups of regions each of which is relatively more homogeneous in its socioeconomic status in 1989. These groups were:-

Group 1 : Arsi, Bale, Gojam, Harerghe, Sidamo, Welega and Welo

Group 2 : Gamogofa

Group 3 : Gonder and Tigray

Group 4 : Illubabor & Kefa

Group 5 : Shewa

The results of this method suggest that Shewa was far from other regions in its socioeconomic characteristics in 1989.

5.1.2 RESULTS OF COMPLETE LINKAGE METHOD

It appears that the eight distinct clusters solution may be appropriate. The eight groups of regions were:

Group 1 : Arsi and Gojam

Group 2 : Bale and Welega

Group 3 : Gamogofa

Group 4 : Gonder and Tigray

Group 5 : Harerghe

Group 6 : Illubabor and Kefa

Group 7 : Shewa

Group 8 : Sidamo and Welo

According to this method Shewa was far from other group of regions in its socioeconomic services in 1989.

5.1.3 RESULTS OF GROUP AVERAGE METHOD

The results of this method show the existence of five groups of regions each of which is more homogeneous in its socioeconomic characteristics in 1989. The five groups were:

Group 1 : Arsi and Gojam

Group 2 : Bale, Gonder, Harerghe, Sidamo, Tigray, Welega and Welo

Group 3 : Gamogofa

Group 4 : Illubabor and Kefa

Group 5 : Shewa

The results of this method confirm that Shewa was at large distance from other regions in its socioeconomic performance in 1989.

5.1.4 RESULTS OF CENTROID METHOD

The results of Centroid method indicate that there were basically five distinct groups of regions in 1989. The groups were:

Group 1 : Arsi and Gojam

Group 2 : Bale, Gonder, Harerghe, Sidamo, Tigray, Welega and Welo

Group 3 : Gamogofa

Group 4 : Illubabor and kefa

Group 5 : Shewa

According to this method it was found that Shewa was at a far distance from the group of other regions.

5.1.5 RESULTS OF MEDIAN METHOD

The results of Median method showed the existence of five distinct groups of regions each of which is more homogeneous in its socioeconomic characteristics. The groups were:

Group 1 : Arsi and Gojam

Group 2 : Bale, Gonder, Harerghe, Sidamo, Tigray, Welega and Welo

Group 3 : Gamogofa

Group 4 : Illubabor and kefa

Group 5 : Shewa

The results of this method also suggest that Shewa was at far distance from other regions in its socioeconomic status in 1989.

5.1.6 RESULTS OF WARD'S METHOD

Ward's method was also applied to the squared Euclidean dissimilarity coefficient matrix in Table 5.2. The method gave 8 distinct groups of regions each of which are relatively more homogeneous in their socioeconomic characteristics in 1989.

The groups were

Group 1 : Arsi and Gojam

Group 2 : Bale and Welega

Group 3 : Gamogofa

Group 4 : Gonder and Tigray

Group 5 : Harerghe

Group 6 : Illubabor and kefa

Group 7 : Shewa

Group 8 : Sidamo and Welo

From the above results we can see that four of the six methods gave five clusters solution. Three of these four

methods gave identical results. Ward's method and the Complete Linkage method gave 8 clusters solution. The group membership obtained for both methods is the same.

The above results reveal that there was a considerable regional disparity on the socioeconomic characteristics in 1989. These results also showed the existence of at least five groups of regions on the basis of variables considered in this study. We can also see that, on the basis of the 19 variables considered to reflect the socioeconomic status of regions, the 12 regions are grouped neither strictly along ethnic or linguistic lines nor on the basis of entirely geographical proximity.

To search for an area where disparity is relatively serious, the six clustering procedures were also applied to each of the following groups of variables for the 1989 data.

I. EDUCATIONAL : Y_{12} TO Y_{17}

II. HEALTH : Y_1, Y_6 TO Y_{11}

III. OTHERS : Y_2 TO Y_5 Y_{18} AND Y_{19}

5.2 SIMILARITY OF REGIONS ON THE BASIS OF EDUCATIONAL VARIABLES IN 1989.

The results of six clustering procedures on the basis educational variables are given below.

The Single Linkage method gave three groups of regions. The groups were:

Group 1 : Arsi, Bale, Gamogofa, Gojam, Harerghe, Illubabor and kefa, Shewa, Sidamo, Welega and Welo

Group 2 : Gonder

Group 3 : Tigray

The Complete Linkage method gave seven distinct groups of regions. The groups were:

Group 1 : Arsi, Gojam and Shewa

Group 2 : Bale

Group 3 : Gamogofa, Sidamo and Welo

Group 4 : Gonder

Group 5 : Harerghe and Illubabor and kefa

Group 6 : Tigray

Group 7 : Welega

The Group Average method resulted in three distinct groups of regions. The groups were:

Group 1 : Arsi, Gojam and Shewa

Group 2 : Bale, Gamogofa, Harerghe, Illubabor and Kefa, Sidamo, Welega and Welo

Group 3 : Tigray and Gonder

The Centroid method resulted in four distinct groups of regions. The groups were :

Group 1 : Arsi, Bale, Gamogofa, Gojam, Shewa, Sidamo, Welega and Welo

Group 2 : Gonder

Group 3 : Harerghe and Illubabor and Kefa

Group 4 : Tigray

The Median method gave four distinct groups of regions. The groups were :

Group 1 : Arsi, Bale, Gamogofa, Gojam, Shewa, Sidamo, Welega and Welo

Group 2 : Gonder

Group 3 : Harerghe and Illubabor and kefa

Group 4 : Tigray

Ward's method resulted in seven groups of regions. The groups were :

Group 1 : Arsi, Gojam and Shewa

Group 2 : Bale

Group 3 : Gamogofa, Sidamo and Welo

Group 4 : Gonder

Group 5 : Harerghe and Illubabor and kefa

Group 6 : Tigray

Group 7 : Welega

The results of the Median method are the same as the results of the Centroid method. The method which gave the same results as that of Ward's method is the Complete Linkage method.

5.3 SIMILARITY OF REGIONS ON THE BASIS OF VARIABLES REPRESENTING HEALTH SERVICES IN 1989.

Six clustering procedures were employed to see the structure of similarity among regions based on variables representing health services. The results are summarized below.

The Single Linkage method gave three distinct groups of regions. The groups were :

Group 1 : Arsi, Bale, Gojam, Gonder, Harerghe, Illubabor & kefa, Sidamo, Tigray, Welega and Welo.

Group 2 : Gamogofa

Group 3 : Shewa

The Complete Linkage method resulted in five distinct groups of regions. The groups were :

Group 1 : Arsi, Bale, Illubabor and kefa, and Welega

Group 2 : Gamogofa

Group 3 : Gojam, sidamo and Welo

Group 4 : Gonder, Harerghe and Tigray

Group 5 : Shewa

The Group Average method gave three groups of regions. The groups were:

Group 1 : Arsi, Bale, Gojam, Gonder, Harerghe, Illubabor and Kefa, Sidamo, Tigray, Welega and Welo.

Group 2 : Gamogofa

Group 3 : Shewa

The Centroid method resulted in three groups of regions. The groups were:

Group 1 : Arsi, Bale, Gojam ,Gonder, Harerghe, Illubabor and kefa, Sidamo, Tigray, Welega and Welo.

Group 2 : Gamogofa

Group 3 : Shewa

The Median method resulted in three groups of regions. The groups were:

Group 1 : Arsi, Bale, Gojam , Gonder, Harerghe, Illubabor and Kefa, Sidamo, Tigray, Welega and Welo.

Group 2 : Gamogofa

Group 3 : Shewa

Ward's method resulted in five groups of regions. The groups were :

Group 1 : Arsi, Bale, Illubabor and Kefa, and Welega

Group 2 : Gamogofa

Group 3 : Gojam, Sidamo and Welo

Group 4 : Gonder, Harerghe and Tigray

Group 5 : Shewa

The above results suggest that the results of Complete Linkage method are the same as that of Ward's method. The other four clustering procedures gave the same results.

5.4 SIMILARITY OF REGIONS ON THE BASIS OF OTHER VARIABLES IN 1989.

The pattern of homogeneity of regions on the basis of variables other than educational and health service indicators was studied for the year 1989. Six clustering procedures were used to classify the Ethiopian regions. The results are summarized below.

The Single Linkage method gave three distinct groups of regions. The groups were:

Group 1 : Arsi, Bale, Gamogofa,Gojam, Gonder, Harerghe, Sidamo, Tigray, Welega and welo

Group 2 : Illubabor and kefa

Group 3 : Shewa

The Complete Linkage method resulted in five distinct groups of regions each of which is relatively homogeneous on the basis of the variables considered. The groups were:

Group 1 : Arsi

Group 2 : Bale, Gonder, Harerghe, Sidamo and Tigray

Group 3 : Gamogofa, Gojam, Welega and Welo

Group 4 : Illubabor and kefa

Group 5 : Shewa

The Group Average method resulted in to four distinct groups of regions namely :

Group 1 : Arsi

Group 2 : Bale, Gamogofa, Gojam, Gonder, Harerghe, Sidamo, Tigray, Welega and Welo

Group 3 : Illubabor and Kefa

Group 4 : Shewa

When Centroid method is applied to the above groups of variables four distinct groups of regions were obtained. The groups were :

Group 1 : Arsi

Group 2 : Bale, Gamogofa, Gojam, Gonder, Harerghe, Sidamo, Tigray, Welega and Welo

Group 3 : Illubabor and Kefa

Group 4 : Shewa

The Median method resulted in three groups of regions each of which is relatively homogeneous. The groups were:

Group 1: Arsi, Bale, Gamogofa, Gojam, Gonder, Harerghe, Sidamo, Tigray, Welega and Welo

Group 2 : Illubabor and Kefa

Group 3 : Shewa

Ward's method gave five distinct groups of regions each of which is relatively more homogeneous. The groups were :

Group 1 : Arsi

Group 2 : Bale, Gonder, Harerghe, Sidamo and Tigray, Welega and Welo

Group 3 : Gamogofa and Gojam

Group 4 : Illubabor and Kefa

Group 5 : Shewa

These results reveal that four of the six methods gave roughly the same results. These methods are Single Linkage, Group Average, Centroid and Median methods. The difference lies only on the group membership of Arsi. Ward's and the Complete Linkage method gave nearly identical results.

To find out an area of activity in which disparity was more serious in 1989, the distances at which clusters combined were compared at each stage of cluster formation for three groups of variables using the six clustering procedures . The results are indicated in Tables 5.3- 5.5, below.

TABLE 5.3 : THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF VARIABLES IN SINGLE LINKAGE AND COMPLETE LINKAGE METHODS (1989)

Stage	Single- Linkage			Complete- Linkage		
	I	II	III	I	II	III
1	.7244	1.398	.2648	.724	1.398	.2649
2	1.6153	1.4684	.5163	1.62	1.627	.5163
3	1.6237	1.627	.5119	3.476	1.7808	1.2973
4	3.2381	1.7808	.964	3.836	2.3929	1.7645
5	3.4762	2.236	1.1471	3.880	2.8975	2.5928
6	4.8376	2.3894	1.2233	7.075	4.735	3.8704
7	5.5388	2.7014	2.5316	8.257	7.034	7.5772
8	6.0801	4.1171	4.1506	9.121	13.028	11.804
9	6.3154	4.6027	4.9249	12.95	14.818	15.6306
10	9.1215	11.4235	9.2171	18.60	35.656	26.745
11	9.5297	26.1835	29.266	32.07	42.855	43.8999

TABLE 5.4 : THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF VARIABLES IN GROUP AVERAGE AND CENTROID METHODS(1989)

Stage	Group-Average method			Centroid method		
	I	II	III	I	II	III
1	.7244	1.398	.2649	.7244	1.398	.2649
2	1.6153	1.627	.5163	1.615	1.627	.5163
3	2.7299	1.7808	.9743	2.326	1.781	.779
4	3.476	2.1829	1.7645	3.378	1.8335	1.4969
5	3.5593	2.3145	1.9299	3.476	1.9077	1.6957
6	5.8193	4.2906	2.3656	5.033	3.5955	1.893
7	7.5074	4.4261	4.1506	6.150	3.9808	4.1506
8	9.1215	7.6363	6.6636	6.418	5.8144	4.8541
9	9.6482	8.2321	13.1135	9.121	4.6721	11.443
10	12.219	23.944	15.2062	9.303	20.9265	12.673
11	17.981	34.793	35.1505	11.46	30.3204	31.800

TABLE 5.5 : DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF VARIABLES IN MEDIAN AND WARD'S METHOD (1989)

Stage	Median method			Ward's method		
	I	II	III	I	II	III
1	.7244	1.398	.2649	.3622	.699	.1324
2	1.615	1.627	.5163	1.169	1.5125	.3906
3	2.326	1.7808	.779	2.720	2.4029	1.1193
4	3.378	1.8334	1.4969	4.458	3.6252	1.9651
5	3.476	1.9077	1.497	6.710	4.897	3.4916
6	4.790	3.9808	2.67	10.48	7.5509	5.4027
7	6.594	4.2963	4.1506	14.91	10.2475	7.478
8	5.120	5.1628	5.4979	19.47	18.9692	11.557
9	8.285	5.3382	8.8275	29.86	30.1822	22.003
10	9.122	18.7109	18.0106	44.97	49.2063	36.849
11	10.88	31.8484	31.5208	66	77	66

The values given in Tables 5.3-5.5 are the distances at which clusters are merged at each stage of clustering. Small values imply that fairly homogeneous clusters are

being merged. Large values show that clusters containing quite dissimilar members are being combined. From these Tables we can neither strictly find the group of variables which have relatively large distance values for all stages nor for the majority of stages. This is an indication that the relative seriousness of disparity of regions for three groups of variables was more or less similar to each other in the year 1989.

5.5 COMPARISON OF THE HOMOGENEITY PATTERN OF REGIONS FOR THE YEAR 1979 WITH THAT OF 1989.

The results obtained by six agglomerative hierarchical procedures were used to indicate the change in pattern of homogeneity of regions in ten years time interval (1979-1989). Each stage of cluster formation in 1979 was compared to its counterpart in 1989. The results are presented in Tables 5.6 -5.11.

TABLE 5.6 : HOMOGENEITY PATTERN OF REGIONS FOR SINGLE-LINKAGE METHOD IN THE YEAR 1979 AND 1989

Stage	Year 1979			Year 1989		
	clusters combined		coeffi-cient	Clusters combined		coeffi-cient
	clu1	clu2		clu1	clu2	
1	5	12	10.148	9	12	8.749
2	2	10	10.923	1	4	9.221
3	2	5	11.717	5	10	12.199
4	2	4	12.152	2	9	13.931
5	1	2	12.315	2	6	14.146
6	1	11	12.653	1	2	14.922
7	1	9	14.704	1	11	14.986
8	1	7	15.333	1	7	18.877
9	1	6	18.228	1	5	18.927
10	1	3	24.749	1	3	24.932
11	1	8	87.719	1	8	68.998

Glancing at Table 5.6 the homogeneity pattern of regions in 1979 was completely altered in 1989. The column labelled coefficient indicate the distance at which clusters combined at each stage of cluster formation. According to the Single Linkage method the distance of Shewa from other

group of regions declined from 87.719 in 1979 to 68.998 in 1989.

TABLE 5.7 : HOMOGENEITY PATTERN OF REGIONS FOR COMPLETE LINKAGE METHOD IN THE YEAR 1979 AND 1989.

Stage	YEAR 1979			YEAR 1989		
	Clusters combined		Coeffi-cient	Clusters combined		Coeffi-cient
	Clus1	Clus2		Clus1	Clus2	
1	5	12	10.148	9	12	8.749
2	2	10	10.923	1	4	9.221
3	4	5	12.305	5	10	12.2
4	1	11	14.07	2	11	14.986
5	2	4	15.036	6	9	22.59
6	1	9	22.233	2	7	22.624
7	3	7	24.749	5	6	23.781
8	2	6	28.833	2	5	40.989
9	1	2	33.766	1	3	49.856
10	1	3	61.73	1	2	51.856
11	1	8	134.08	1	8	98.261

According to the Complete Linkage method the homogeneity pattern of regions in 1979 was altered in 1989. The distance of Shewa from other group of regions declined from 134.078 in 1979 to 98.261 in 1989.

TABLE 5.8: HOMOGENEITY PATTERN OF REGIONS FOR GROUP AVERAGE METHOD IN THE YEAR 1979 AND 1989.

Stage	Year 1979			Year 1989		
	Clusters combined		Coeffi-cient	Clusters combined		Coeffi-cient
	Clu1	Clu2		Clu1	Clu2	
1	5	12	10.148	9	12	8.749
2	2	10	10.923	1	4	9.221
3	2	5	12.159	5	10	12.199
4	1	4	12.315	2	11	14.986
5	7	11	15.333	2	9	17.985
6	1	9	15.899	2	6	20.766
7	1	2	17.611	2	5	23.656
8	1	7	24.941	2	7	28.214
9	1	6	27.107	1	2	33.111
10	1	3	42.610	1	3	42.326
11	1	8	104.39	1	8	83.704

From Table 5.8, we can see that the homogeneity pattern of regions in 1979 was almost completely altered in 1989. The distance of Shewa from other group of regions declined from 104.391 in 1979 to 83.704 in 1989.

TABLE 5.9 : HOMOGENEITY PATTERN OF REGIONS FOR CENTROID METHOD IN THE YEAR 1979 AND 1989.

Stage	YEAR 1979			YEAR 1989		
	Clusters combined		Coeffi-cient	Clusters combined		Coeffi-cient
	Clus1	Clus2		Clus1	Clus2	
1	5	12	10.148	9	12	8.749
2	5	10	9.281	1	4	9.221
3	2	5	8.22	5	10	12.199
4	2	4	9.01	2	9	13.055
5	1	2	13.562	2	11	14.455
6	1	9	12.267	2	5	14.69
7	1	11	14.825	2	6	13.37
8	1	6	17.874	2	7	19.489
9	1	7	20.057	1	2	21.04
10	1	3	33.271	1	3	30.686
11	1	8	93.154	1	8	70.586

From Table 5.9, we can observe that the homogeneity pattern of regions in 1979 is quite different from that of 1989. The distance of Shewa from the group of other regions declined from 93.154 in 1979 to 70.586 in 1989.

TABLE 5.10 : HOMOGENEITY PATTERN OF REGIONS FOR MEDIAN METHOD IN THE YEARS 1979 AND 1989.

Stage	YEAR 1979			YEAR 1989		
	Clusters combined		Coeffi-cient	Clusters combined		Coeffi-cient
	Clus1	Clus2		Clus1	Clus2	
1	5	12	10.148	9	12	8.749
2	5	10	9.281	1	4	9.221
3	2	5	8.122	5	10	12.199
4	2	4	10.243	2	9	13.055
5	1	2	11.486	2	11	13.5
6	1	9	11.006	2	6	15.276
7	1	11	14.424	2	5	14.718
8	1	7	17.559	2	7	21.554
9	1	3	27.084	1	2	27.301
10	1	6	37.805	1	3	32.594
11	1	8	93.069	1	8	74.32

From Table 5.10, we can see that the homogeneity pattern of regions in 1979 was completely altered in 1989. The Table also indicates that the distance of Shewa from other group of regions declined from 93.069 in 1979 to 74.32 in 1989.

TABLE 5.11:- HOMOGENEITY PATTERN OF REGIONS FOR WARD'S METHOD IN THE YEARS 1979 AND 1989.

Stage	YEAR 1979			YEAR 1989		
	Clusters combined		Coeffi-cient	clusters combined		coeffi-cient
	Clus1	Clus2		Clus1	Clus2	
1	5	12	5.074	9	12	4.375
2	2	10	10.536	1	4	8.985
3	1	4	16.693	5	10	15.085
4	2	5	23.584	2	11	22.578
5	7	11	31.251	6	9	33.365
6	1	9	39.798	2	7	44.701
7	2	6	53.223	2	6	60.936
8	1	2	70.277	2	5	82.736
9	3	7	91.255	2	3	110.917
10	1	3	123.608	1	2	144.296
11	1	8	209.00	1	8	209.00

From Table 5.11, we can see that the homogeneity pattern of regions in 1979 was completely altered in 1989. Ward's method suggested that the distance of Shewa from other group of regions in 1989 was the same as that of 1979. The results of all six clustering procedures suggested that the structure of similarity of regions in 1979 was altered in 1989. The decline of the distance of Shewa from other group of **(for methods other than Ward's method)** regions is an indication that Shewa is on the way to join the group of other regions provided that a similar trend as that of 1979-1989 continues.

5.6 RESULTS AND ANALYSES OF THE CHANGES IN SOCIOECONOMIC CHARACTERISTICS

An attempt was made to see the disparity of regions on the basis of the changes in socioeconomic characteristics between the years 1979 and 1989. To do so the ratios of the 1989 socioeconomic variables to that of 1979 were employed. Some descriptive measures for these ratios are indicated in Table 5.12 below.

TABLE 5.12: SOME DESCRIPTIVE MEASURES FOR THE CHANGES IN SOCIOECONOMIC VARIABLES

Variable	Mean	Std dev	coeff.-of variation	Minimum	Maximum
R_1	1.44	0.15	10.42	1.21	1.75
R_2	1.59	0.17	10.69	1.32	1.88
R_3	1.13	0.10	8.85	1.02	1.36
R_4	1.30	0.40	30.77	.62	2.02
R_5	1.23	0.28	22.76	.92	1.88
R_6	1.58	0.40	25.32	1.08	2.45
R_7	1.14	0.21	18.42	1.00	1.67
R_8	2.76	0.91	32.97	1.33	4.25
R_9	2.46	0.44	17.89	1.53	3.22
R_{10}	2.02	0.57	28.22	1.32	2.93
R_{11}	2.26	0.86	38.05	1.10	4.21
R_{12}	2.47	0.97	39.27	0.78	4.87
R_{13}	2.48	0.98	39.52	1.11	3.75
R_{14}	1.72	0.61	35.47	1.00	3.21
R_{15}	2.82	1.12	39.72	1.69	5.70

Variable	Mean	Std dev	coeff.-of variation	Minimum	Maximum
R_{16}	3.23	1.25	38.70	1.81	6.15
R_{17}	1.85	0.55	29.73	1.36	3.33
R_{18}	3.02	0.57	18.87	2.09	3.82
R_{19}	2.88	1.34	46.53	1.53	6.45
R_{20}	1.11	0.34	30.63	0.72	1.95
R_{21}	1.03	0.35	33.98	0.60	1.68
R_{22}	2.03	1.36	67.00	0.88	6.08
R_{23}	2.28	0.92	40.35	0.93	4.37

From Table 5.12, we can observe that there is a large gap between the minimum and maximum values of the ratios for most of the cases. This could be an indication of the existence of regional variation in the changes of socioeconomic characteristics considered in this particular study.

Table 5.13 shows the squared Euclidean dissimilarity coefficient matrix for all possible pairs of the 12 regions based on the ratios of socioeconomic variables in the year

1989 to that of 1979. The smallest distance in the matrix is 3.2644. This indicates that the most similar regions in the changes of socioeconomic characteristics were Arsi and Gojam.

TABLE 5.13 SQUARED EUCLIDEAN DISSIMILARITY COEFFICIENT MATRIX(RATIOS)

Case	1	2	3	4
2	9.2556			
3	36.8709	41.9021		
4	3.2644	10.9648	37.2413	
5	14.0938	28.3040	75.1413	12.8284
6	12.3008	17.0988	78.8851	11.9696
7	26.9930	37.2712	76.8437	22.4296
8	15.6780	17.9962	75.0851	15.2338
9	17.8530	22.1688	79.2075	18.7730
10	27.7702	37.8690	107.1807	28.6732
11	18.6290	22.1210	84.9369	21.0782
12	12.7771	16.6095	75.1136	11.6381

Case	5	6	7	8
6	8.2522			
7	30.5822	32.0666		
8	16.0126	11.2478	27.5396	
9	12.0132	10.2730	32.8614	11.4734
10	10.6012	13.6928	44.6760	13.8592
11	14.9192	9.8834	28.7944	7.3074
12	10.1701	4.7155	29.9853	5.9451

Case	9	10	11
10	8.4592		
11	6.0084	11.1290	
12	8.0823	8.6609	8.6055

The results of the six clustering procedures which are based on dissimilarity coefficient matrix of Table 5.13 are given below.

The Single Linkage method resulted in three distinct groups of regions. The groups were :

Group 1 : Arsi, Bale, Gojam, Gonder, Harerghe, Shewa, Sidamo, Tigray, Welega and Welo

Group 2 : Gamogofa

Group 3 : Illubabor and Kefa

The Complete Linkage method gave nine distinct groups of regions. The groups were :

Group 1 : Arsi and Gojam

Group 2 : Bale

Group 3 : Gamogofa

Group 4 : Gonder

Group 5 : Harerghe and Welo

Group 6 : Illubabor and Kefa

Group 7 : Shewa

Group 8 : Sidamo and Welega

Group 9 : Tigray

The Group Average method resulted in four distinct groups of regions. The groups were:

Group 1 : Arsi, Bale and Gojam

Group 2 : Gamogofa

Group 3 : Gonder, Harerghe, Shewa, Sidamo, Tigray, Welega and Welo

Group 4 : Illubabor and Kefa

The Centroid method gave four distinct groups of regions. The groups were :

Group 1 : Arsi, Bale, and Gojam

Group 2 : Gamogofa

Group 3 : Gonder, Harerghe, Shewa, Sidamo, Tigray, Welega and Welo

Group 4 : Illubabor and Kefa

The Median method resulted in four distinct groups of regions. The groups were:

Group 1 : Arsi, Bale and Gojam

Group 2 : Gamogofa

Group 3 : Gonder, Harerghe, Shewa, Sidamo, Tigray, Welega and Welo

Group 4 : Illubabor and Kefa

Ward's method resulted in nine groups of regions. The groups were :

Group 1 : Arsi and Gojam

Group 2 : Bale

Group 3 : Gamogofa

Group 4 : Gonder

Group 5 : Harerghe and Welo

Group 6 : Illubabor and Kefa

Group 7 : Shewa

Group 8 : Sidamo and Welega

Group 9 : Tigray

Three of the six methods gave the same results. These are the Group Average , the Centroid and the Median methods. According to the Complete Linkage and Ward's method, almost all regions were dissimilar. Most of the regions were in a separate group. This could be an indication that regions were dissimilar in the changes of their socioeconomic characteristics.

5.6.1.COMPARISON OF THE EXTENT OF SIMILARITY OF REGIONS FOR THREE GROUPS OF RATIOS REPRESENTING CHANGES IN SOCIOECONOMIC VARIABLES

We have already observed that the existence of regional disparity was quite obvious on the basis of changes in socioeconomic variables. Similarity of regions at disaggregate level was also considered to search for an area of activity in which disparity was more severe.

The six clustering procedures were applied to each of the following group of ratios.

I. EDUCATION : R_{11} TO R_{19}

II. HEALTH : R_4 TO R_{10}

III. OTHERS : R_1 TO R_3, R_{20} TO R_{23}

The distances at which clusters combined at each stage of cluster formation were compared for three groups of variables using six clustering procedures. The results are indicated in Tables 5.14- 5.16.

TABLE 5.14: THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF RATIOS IN SINGLE-LINKAGE AND COMPLETE- LINKAGE METHODS.

Stage	Single Linkage			Complete Linkage		
	I	II	III	I	II	III
1	1.39	.597	.087	1.392	.597	.087
2	1.95	.634	.396	1.957	.634	.396
3	2.15	.68	.406	2.545	.68	.581
4	2.52	.777	.582	3.408	.85	.588
5	2.59	.85	.588	4.128	1.262	1.276
6	2.72	.885	.734	4.868	1.436	1.356
7	2.94	1.093	.833	5.023	2.007	1.382
8	3.40	1.494	.924	7.367	2.389	2.256
9	3.58	1.726	1.27	13.83	3.243	5.078
10	5.06	1.774	2.956	29.41	6.722	14.639
11	27.4	2.965	14.257	96.01	11.162	28.66

TABLE 5.15: THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF RATIOS IN GROUP-AVERAGE AND CENTROID METHODS

Stage	Group-Average method			Centroid method		
	I	II	III	I	II	III
1	1.39	0.597	0.087	1.392	0.597	0.087
2	1.96	0.634	0.396	1.957	0.634	0.396
3	2.34	0.680	0.582	1.856	0.680	0.505
4	3.19	0.850	0.588	2.46	0.850	0.588
5	3.41	1.019	0.973	3.248	0.870	0.802
6	3.79	1.265	1.138	3.17	1.106	0.935
7	4.49	1.607	1.724	3.448	1.145	1.248
8	5.02	1.896	1.760	3.788	1.683	1.189
9	7.55	3.095	2.219	5.761	2.457	2.276
10	13.595	3.586	7.042	10.39	2.580	6.252
11	55.5	4.774	22.285	51.58	4.145	21.05

TABLE 5.16: THE DISTANCES AT WHICH CLUSTERS COMBINED FOR THREE GROUPS OF RATIOS IN MEDIAN AND WARD'S METHODS

Stage	MEDIAN method			WARD'S method		
	I	II	III	I	II	III
1	1.392	0.597	0.087	0.696	0.298	0.044
2	1.957	0.634	0.396	1.675	0.616	0.242
3	1.856	0.680	0.505	2.912	0.955	0.532
4	2.372	0.850	0.588	4.616	1.381	0.826
5	3.448	0.870	0.743	6.461	1.961	1.490
6	3.547	1.106	1.013	8.76	2.698	2.175
7	3.53	1.170	0.977	11.456	3.82	3.131
8	4.062	1.680	1.063	14.525	5.194	4.601
9	6.597	2.275	3.444	20.584	7.252	7.902
10	16.952	3.270	8.508	41.124	12.037	13.585
11	47.658	4.069	22.846	90.558	19.717	32.881

The results of Tables 5.14- 5.16 reveal that the distances for ratios in group I are larger than those of the remaining two groups at each stage of cluster formation for all clustering methods . This is an indication that regions were more dissimilar in the changes of educational services in the time interval 1979 - 1989 .

CHAPTER VI

DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS

In the preceding chapters we tried to express some of our ideas about the results found in this study. We would like to sum up these ideas now.

Cluster analysis was used to study the degree of similarity among regions of Ethiopia on the basis of some socioeconomic characteristics. In order to see the similarity structures of regions, we applied the six agglomerative hierarchical clustering procedures on the data of the years 1979 and 1989 separately. The similarity structures of regions which were obtained by using the six clustering procedures on the 1979 data, were compared with the similarity structures of regions which were obtained by using the same six procedures on the 1989 data. Regions were also grouped on the basis of changes in the socioeconomic characteristics by applying the six agglomerative clustering procedures on the ratios of the year 1989 to that of 1979 .

The conclusions of our results are limited by the amount of variables included in the study. On the basis of the socioeconomic characteristics considered in the study the following conclusions were drawn.

From the analyses of the 1979 data we conclude that :

I. The study of the homogeneity pattern of the Ethiopian regions for the year 1979 indicated the existence of considerable variation among regions.

II. The existence of at least four and at most eight groups of regions was confirmed depending on the clustering method employed.

III. Shewa was at a far distance from other group of regions in the year 1979 .

IV. The study to search for an area of activity in which disparity was relatively more serious showed that Ethiopian regions were more dissimilar on the basis of variables representing educational services in 1979.

V. The twelve regions were grouped neither strictly along ethnic or linguistic lines nor on the basis of entirely geographical proximity.

From the analyses of the same sort for the 1989 data we conclude that :

I. The existence of regional variation was quite obvious in the year 1989.

II. The results revealed that there were at least five and at most eight groups of regions depending on the clustering method used.

III. Shewa was at a far distance from other group of regions but at a shorter distance as compared to that of the year 1979.

IV. The study to search for an area of activity in which disparity was more severe did not show a clear answer for the 1989 data. This could be an indication that the relative seriousness of disparity among regions for three groups of variables became more or less similar to each other in the year 1989 .

V. The twelve regions were grouped neither strictly along ethnic or linguistic lines nor on the basis of entirely geographical proximity for the 1989 data.

When we compare the similarity structures of regions for the two specified Calendar years, the structures in the year 1979 were completely altered in 1989. The distance of Shewa from other group of regions declined in 1989 as compared to that of the year 1979. This could be an indication that Shewa was on the way to join the group of other regions. However, this does not mean that the disparity among regions of Ethiopia shows a decline; because some regions which were similar in the year 1979 became dissimilar in 1989 and the minimum number of clusters obtained in 1989 was also larger than that of 1979.

In dealing with the changes in the socioeconomic characteristics of regions a considerable variation was observed in that regions were grouped from three (**Single-linkage method**) to nine (**Complete-Linkage and Ward's methods**) groups. According to the Complete Linkage and Ward's methods almost all regions were dissimilar in the changes of their socioeconomic characteristics.

Finally we recommend the following :

I. The variation in the changes of socioeconomic characteristics of Ethiopian regions needs special attention ; otherwise it may aggravate the existing disparity among regions.

II. Our results endorsed the view that the agglomerative hierarchical clustering procedures can serve as good statistical measures to compare the socioeconomic status of regions in a fairly remote past with that of the present.

III. The study was made at the regional level and, therefore, the results refer to regions. But it is our belief that such methods can be useful tools if one goes down the regional hierarchy and considers the Awraja and Wereda regional units.

REFERENCES

1. Baker, F.B.(1974) Stability of two Hierarchical Grouping Techniques Case 1. Sensitivity to Data errors. J.Am. Statist. Assoc., 69, 440- 445.

2. Central Statistical Authority(CSA), Statistical abstract 1988, Ethiopian Government publication Addis Ababa.

3. Central Statistical Authority(CSA), Statistical bulletin 90, Ethiopian Government publication, Addis Ababa.

4. Central Statistical Authority(CSA), Statistical bulletin 103, Ethiopian Government publication, Addis Ababa.

5. Central Statistical Organization(CSO), Statistical Abstract 1980, Ethiopian Government publication, Addis Ababa.

6. Chaudhry, M.D.(1971) Economic Distance Among Regions: A Statistical Analysis, J. Economic Development and Cultural change, 19, 527-544.

7. Cormack, R.M. (1971) A Review of Classification J.R. Stat. Soc. 134, 321-67

8. Ethiopian Telecommunication Authority, Statistical Bulletin 1989, Ethiopian Government Publication, Addis Ababa.

9. Everitt, B. S.(1980) Cluster Analysis, Heineman, London.

10. Everitt, B. S.(1978), Graphical Techniques for Multivariate Data, Heinman;London.

11. Everitt, B. S.(1979) Unresolved Problems in Cluster Analysis. Biometrics, 35 169- 181.

12. Gee, C. Problems and Approaches to Measuring Regional Development Disparities in : Regional Planning and Development in Ethiopia I, Treuner, P. etal (eds.), IDR and IREUS, Addis Ababa, 1985

13. Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. Wiley, New York.

14. Gower, J. C.(1966) A Comparison of Some Methods of Cluster Analysis. Biometrics, 23, 623-637.

15. Green, P.E. and Carroll, J.D.(1978) Analyzing Multivariate Data, Holt, Rinehart and Winston.

16. Hawkins, D.M.(1982) Topics in Applied Multivariate Analysis, Cambridge, London

17. Humbert, L.J (1974) Approximate Evaluation Techniques for Single-link and Complete-link Hierarchical Clustering Procedures. J. Am. Statist.Ass., 69, 698- 704.

18. Jardine, N. and Sibson, R.,(1971) Mathematical Taxonomy, London: John Wiley and Sons, inc.

19. Johnson, R.A. and Wichern, D.W.(1992) Applied Multivariate Statistical Analysis, Engle Wood Cliffs, N.J, Prentice Hall, Inc.

20. Kuiper, F.K. and Fisher, L.(1975) A monte Carlo Comparison of Six Clustering Procedures Biometrics, 31, 777-783.

21. Minstry of Education, Basic Education Statistics 1989 Bulletin.

22. Mulatu Wubneh, Time Series Data on Socioeconomic Characteristics of Ethiopia, Addis Ababa, Sep., 1990.

23. Negarit gazeta, no. 33 Addis Ababa, 20 th, Oct., 1992.

24. Norusis, M.J.(1986) SPSS/PC+ Advanced Statistics Guide. SpssInc. Chicago.

25. Office of the National Committee for Central Planning, Regional Planning Department, 1989, An Outline for Elaboration of a Methodology of Regional Planning in Ethiopia, page 56-66 Indicators.

26. Planning and Programming Bureau (1989), Comprehensive Health Service Directory : Ministry of Health Addis Ababa.

27. Ward, J.H.(1963). Hierarchical Grouping to Optimize an Objective Function , J. Amer.Stat. Ass., 58 , 236-244.

28. Wishart, D.(1969) An Algorithm for Hierarchical Classification. Biometrics 25, 165- 170.

APPENDICES

Appendix I The 1979 socioeconomic data

Region's Code	Urban Population in thousand	Total Population in thousand	Length of roads in Kms.
1	99.40	1119.30	197
2	47.10	856.10	225
3	44.30	977.10	213
4	179.70	1984.40	837
5	153.30	1999.60	827
6	257.60	3043.20	1727
7	176.50	2362.90	1292
8	1697.40	6195.30	2425
9	201.30	2734.70	982
10	182.90	2105.40	600
11	112.20	1966.30	692
12	172.50	2544.10	981

Region's Code	Area in 1,000 Sq. Kms.	Num. of Hospital Beds	Num.of Health Centers	Num.of Health Stations
1	24.60	168	5	51
2	128.30	189	4	59
3	40.10	131	8	85
4	64.40	248	8	74
5	73.40	336	13	82
6	254.80	930	13	118
7	103.80	373	16	172
8	85.50	3761	17	172
9	116.70	470	9	106
10	65.70	410	9	86
11	69.80	306	8	102
12	79.00	449	13	120

Region's Code	Num. of Hospitals	Num.of Primary Schools	Num. of Primary School Teachers	Num. of Primary School Pupils	Num. of Junior Secondary Schools
1	2	146	1288	52678	31
2	1	136	874	28077	13
3	3	146	971	47466	14
4	3	254	1831	58247	28
5	3	249	1574	40427	19
6	10	367	1871	59066	50
7	5	442	2052	128184	53
8	24	1214	8391	463590	236
9	5	298	2509	131094	32
10	4	436	2073	54270	26
11	4	579	2394	185135	91
12	5	410	2122	80470	35

Region's Code	Num. of Junior Secondary School Teachers	Num. of Junior Secondary School Pupils	Num. of Senior Secondary Schools	Num. of Senior Secondary School Teachers
1	155	6509	6	135
2	91	2866	3	67
3	96	3325	5	130
4	207	7676	9	203
5	274	7665	11	176
6	306	7261	9	157
7	201	9721	16	163
8	1264	66798	51	1579
9	298	11243	8	234
10	208	7036	8	165
11	262	13869	12	197
12	261	9825	11	229

Region's Code	Num. of Senior Secondary School Pupils	Area Under Major crops Per 1, 000 Hectares	Production of Major Crops per 1,000 quintals	Num. of Agricultural Extension Workers
1	6077	355.4	4462.10	0
2	2472	196.7	1441.20	75
3	2580	137.9	1158.20	136
4	8268	806.7	8163.20	222
5	7568	600.0	6152.50	185
6	12135	355.9	6535.30	248
7	7149	411.9	5922.80	341
8	71661	1464.6	20577.30	690
9	9622	161.9	2073.50	130
10	7805	269.8	2315.60	195
11	12568	504.7	6592.30	225
12	9082	552.7	6592.30	203

Region's Code	Num.of medical doctors	Num. of Nurses	Num. of Health Assistants	Num. of Telephone Lines	Num.of Post Offices
1	5	28	179	443	13
2	6	25	107	355	11
3	4	31	174	248	8
4	13	52	225	1163	20
5	14	55	246	2493	29
6	27	78	344	3177	50
7	20	63	344	2760	37
8	258	729	1871	40153	108
9	12	57	242	1069	29
10	7	48	246	1111	26
11	14	64	290	659	29
12	18	60	360	1509	45

Appendix II The 1989 socioeconomic data

Region's Code	Urban Population in thousand	Total Population in thousand	Length of roads in Kms.
1	132.73	1914.39	235
2	82.40	1159.29	230
3	63.51	1435.21	225
4	262.37	3737.31	977
5	248.80	3365.16	874
6	334.36	4792.52	1891
7	213.04	3928.68	1645
8	2452.12	11044.93	2792
9	278.03	4364.13	998
10	272.62	2779.98	670
11	146.82	2850.05	942
12	275.54	4193.58	1043

Region's Code	Area in 1,000 Sq. Kms.	Num. of Hospital Beds	Num.of Health Centres	Num.of Health Stations
1	23.70	205	7	103
2	127.10	118	5	83
3	40.30	265	10	124
4	61.20	414	9	117
5	79.60	561	13	132
6	272.60	1030	12	164
7	103.00	562	16	250
8	85.30	4117	32	422
9	119.80	353	10	137
10	64.90	570	10	102
11	70.50	444	13	209
12	82.10	433	14	130

Region's Code	Num. of Hospitals	Num.of Primary Schools	Num. of Primary School Teachers	Num. of Primary School Pupils	Num. of Junior Secondary Schools
1	2	400	3635	176614	47
2	1	321	2445	91219	21
3	5	615	4733	178110	45
4	4	776	4608	215690	64
5	4	654	3769	96683	43
6	11	796	4649	192924	77
7	5	997	6447	297750	83
8	24	1661	15066	828305	420
9	5	660	4927	173262	43
10	4	597	1607	61515	26
11	5	637	5221	206062	103
12	5	693	4082	184161	47

Region's Code	Num. of Junior Secondary School Teachers	Num. of Junior Secondary School Pupils	Num. of Senior Secondary Schools	Num. of Senior Secondary School Teachers
1	558	27713	14	447
2	292	11951	10	256
3	549	20467	10	473
4	644	32262	18	598
5	478	17815	19	438
6	545	20154	17	545
7	645	26121	22	548
8	3656	165462	87	4195
9	638	24904	12	567
10	352	12779	12	346
11	683	32926	18	702
12	573	22416	15	574

Region's Code	Num.of Senior Secondary School Pupils	Area Under Major Crops Per 1,000 Hectares	Production of Major Crops per 1,000 quintals	Num. of Agricultural Extension Workers
1	19270	469.10	7025.87	187
2	10480	175.92	1995.50	55
3	16631	268.47	1951.85	163
4	26441	897.72	7823.77	312
5	17957	479.67	4723.15	139
6	18524	331.27	5130.74	202
7	19992	354.34	4738.62	169
8	182635	1438.67	17442.57	549
9	21526	202.68	2457.04	122
10	13842	366.79	2017.41	146
11	25850	365.39	3973.46	156
12	19866	645.57	6879.94	205

Region's Code	Num.of Medical Doctors	Num. of Nurses	Num.of Health Assistants	Num. of Telephone Lines	Num. of Post Offices
1	17	70	366	890	23
2	8	60	230	1056	19
3	17	64	316	1084	16
4	40	138	400	2714	50
5	56	143	436	2327	37
6	58	175	687	5606	61
7	62	203	994	3373	225
8	457	1119	2588	102109	264
9	32	171	710	3291	37
10	23	116	343	1993	23
11	30	173	790	1603	46
12	35	130	474	2950	75