DOCUMENT RESUME

ED 047 957 SE 010 795

AUTHOR Lockley, J. Elaine
TITLE A Comparative Study of Some Cluster Analytic Techniques With Application to the Mathematics Achievement in Junior High Schools.
INSTITUTION Merritt Coll., Oakland, Calif.
PUB DATE Feb 71
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (Feb. 4-7, 1971, New York City, N.Y.)

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement, *Cluster Grouping, Junior High School Students, Mathematical Models, National Surveys, *Research Methodology, Research Tools, *Statistical Analysis

ABSTRACT

Reported are the results of a study designed to investigate and compare four cluster analytic procedures as potential methods for the analysis of educational data. A secondary objective was to determine whether or not there was some underlying multidimensional structure to a set of mathematics achievement data. The four clustering procedures (Ball and Hall's ISODATA, Johnson's HICLUS, Friedman and Rubin's iterative procedure, Singleton and Kantz's iterative procedure) were compared by applying them to a data set from the National Longitudinal Study of Mathematical Abilities of SMSG. The clustering variables were scales which described the characteristics of thirty junior high schools and their communities. The four clustering techniques produced very similar sets of clusters, and from all indications three or four clusters seem appropriate for clustering the mathematics achievement data. It was found that the students' mathematics achievement across clusters was not the same after adjustments were made for differences in aptitude and initial understanding of mathematical concepts. It was concluded that the differences in achievement were due at least in part to the effect of the particular school on the student. (Author/RS)

A COMPARATIVE STUDY OF SOME CLUSTER ANALYTIC TECHNIQUES WITH APPLICATION TO THE MATHEMATICS ACHIEVEMENT IN JUNIOR HIGH SCHOOLS

J. Elaine Lockley
Merritt College
Oakland, California

Paper read at the annual meeting of the American Educational Research Association, New York, New York, February, 1971.

INTRODUCTION

Educational researchers are often confronted with the problem of attempting to arrange objects (individuals, tests, test items, etc.) into groups by utilizing a set of measurements observed on the objects. The researcher attempts to determine a natural grouping of the data using a small number of clusters.

The method called cluster analysis takes a set of heterogeneous data and subdivides it into smaller more homogeneous groups called clusters. The purpose is to form groups of similar objects. In testing a hypothesis, the heterogeneity of the data may not permit us to detect any differences. However, by combining into homogeneous units we can detect differences more easily.

An overall description of the clusters may be obtained by listing the objects in each of the clusters or by using the center of gravity (mean) of each cluster. Hopefully, this description could be reproduced if another sample of the same size were to be chosen from the same population.

CLUSTERING TECHNIQUES

Suppose we have p variates, each observed on N objects (or individuals). We may write x_{ij} as the jth observation for the ith object. The data may be represented as a point in a p dimensional space as

$$x_i = (x_{i1}, \ldots, x_{ip}),\ i = 1, \ldots, N.$$

The point x_i represents the p measurements or observations made on the ith object or individual. These observations made on the N objects may be summarized in a matrix of observations, X, of order N x p. If we let T denote the matrix of sums of squares and cross-products of deviations about the mean, then

$$T = (X - M)'(X - M) = \sum_{i=1}^{N} (x_i - \bar{x})'(x_i - \bar{x})$$

where M is the matrix of means. Since the total sum of squares and cross-products may always be written as the sum of two terms: The sum of squares and cross-products within clusters, W, and the sum of squares and cross-products between clusters, B, we have that

$$T = B + W.$$

The between-cluster scatter matrix, B, reflects the inter-group differences, and can be used to measure the contribution made to these differences as a result of applying the different treatments to the G groups. Since objects in the same cluster will vary only in accordance with individual or chance differences and not as to treatment applied, the within-cluster scatter matrix, W, reflects intragroup differences.

A good clustering procedure for organizing data will produce clusters such that objects within clusters are more homogeneous than objects between clusters. That is, partitioning of the data into clusters is done in such a way that there is minimum variation within clusters. This may be accomplished by minimizing the matrix W, which by necessity then maximizes the matrix B. This is because the sum of W and B is constant, and is independent of the partitioning of the data points.

In each of the clustering techniques compared, the N objects are partitioned into a predetermined number of clusters, say G. Their common goal is the minimization of the amount of variation within the clusters, while at the same time producing a fixed number of clusters. Hence, either directly or indirectly the methods are designed to minimize a function of W and/or B. It is important to note that although all methods attempt to find an absolute minimum (or maximum) for the chosen criterion, the algorithm generally stops as soon as a local minimum (or maximum) is obtained. This means that two algorithms using the same criterion may yield different results when there are several extrema points.

All of the techniques used to cluster a group of objects are dependent upon four basic steps. (1) Selection of variables (measurements or observations) used to describe each of the objects, and the scaling of these variables. (2) Proper choice of a proximity parameter which will be used to measure the similarity between pairs of objects to be clustered. (3) Selection of a criterion function (algebraic function) to measure the "goodness" of the clustering technique. (4) Interpretation of the clusters formed by the technique.

The methods of cluster analysis compared in this study are: Ball and Hall's ISODATA (1965), a hierarchical clustering procedure (HICLUS) described by Johnson (1967), and two other iterative procedures, Friedman and Rubins' procedure (1967), and Singleton and Kautzs' procedure (1965). In each of the methods, the variation within the clusters is minimized in accordance with some criterion.

Singleton and Kautz (1965) devise a clustering algorithm which minimizes the sum of the squared deviations from the cluster means of the pooled within-groups scatter matrix, W. This function called the "Trace W" criterion partitions the data directly into G groups using a hill-climbing process.

Ball and Hall (1965) develop a clustering procedure called ISODATA, an acronym for Iterative Self-Organizing Data Analysis. This procedure summarizes a large data set by choosing a smaller set of cluster means called "centers" that tend to minimize the sum of squared distances of each data point from its nearest center. The process implicitly minimizes the Trace W function.

Friedman and Rubin (1967a, 1967b) develop a clustering procedure to find the "best" partition of N objects into a given number of groups, G, using a hill-climbing process. Here best partition is defined as the partition which maximizes a chosen criterion function. Friedman and Rubin discuss and use three criteria for clustering: Negative Trace W, Trace $W^{-1}B$, and det(B+W)/det(W).

Johnson (1967) describes a procedure for grouping objects in a manner that establishes a taxonomy of nonoverlapping clusters called hierarchical groups, where each larger unit is the union of the next subordinate units. The process begins by placing the N objects into N clusters and continues until all N objects are placed into one cluster. These groups of clusters are formed by using one of two criteria. One criterion forms clusters so that variation within each cluster is minimally increased at each stage of clustering. That is, its' goal is the formation of clusters that are optimally compact. The second criterion attempts to form clusters that are optimally connected. It should be noted that the restriction that the clustering be strictly hierarchical may have the consequence that some level of the clustering may not be truly optimal.

All of the above procedures have as an objective the analysis of multivariate heterogeneous data by partitioning the data set into smaller more homogeneous groups. As a result of the clustering, the groups should lend more insight into the structure of the data. These clustering procedures could then be applied to any discipline where the researcher has gathered N objects to study and has described each object by taking a set of one or more measurements on each of the N Objects.

Formal statistical theory has not been developed for clustering procedures, so that traditional sampling theory and tests of hypothesis are unavailable. However, in this study once the clusters have been determined, formal statistical analysis is used to determine the extent to which the various groups differ in terms of their students' mathematics achievement.

MATHEMATICS ACHIEVEMENT DATA SETS

The data sets analyzed were collected by the National Longitudinal Study of Mathematical Abilities (NLSMA) of the School Mathematics Study Group (SMSG). This study focuses attention on thirty junior high schools from a population of 197 junior high schools. These schools remained in the NLSMA study for the entire period of five years. There were 2995 students tested in the thirty schools.

The sets of measurements taken on each school are divided into two main groups: Student-test variables which consist of mathematical and psychological scales, and a set of non-test variables which are grouped into two classifications-- school-community and teacher. The school-community scales provide information about the individual school and the community served by the school. The teacher scales include information on the teachers' educational background and questions designed to measure the teachers' attitude toward teaching mathematics.

One of the goals of the analysis of the clusters is to identify some of the variables associated with the development of mathematical abilities. By grouping the school into smaller more homogeneous clusters, we hope to reach our goal by comparing the students' mathematics achievement across these clusters which have been made as dissimilar as possible.

CLUSTERING RESULTS

Clustering of the schools is done on the school means obtained using twelve school-community variables: Average daily attendance, residential description, parents' yearly income, teachers' starting salary, teachers salary index, innovations, mathematics supervisor, heavy use of SMSG, heavy use of other experimental mathematics programs, inservice training of teachers, mathematics class size, and other academic class size. The teacher scales are not used to cluster the schools but are used for descriptive purposes only. Seventeen teacher scales are used.

Principal Component Analysis

A principal component analysis is performed to interpret the data in fewer than twelve dimensions, in terms of the school-community variable description. The first five principal components accounted for 72 per cent of the total variance.

Each of the five factors is bipolar. The first factor is called "School Characteristics". The largest positive loadings are on variables- mathematics class size, academic class size, inservice training, and heavy use of experimental mathematics; parents' median yearly income has a large negative loading. This result is consistent with the factor interpretation inasmuch as low income is often associate with large class size. In a similar manner the other four factors were named "District Professional Expenditures", "Family Socioeconomic Status", "Innovations", and "SMSG Usage", respectively.

Number of Clusters

Three clusters are extracted using the Friedman-Rubin, Singleton-Kautz, and Johnson procedures; and four clusters are extracted using the Ball-Hall procedure. Johnsons' set of three clusters is very similar to Ball-Halls' set of four clusters; infact, they differ only in the placement of two schools. The sets of three clusters obtained under the Friedman-Rubin and Singleton-Kautz procedures using the Trace W criterion are almost identical. The only exception is the placement of one school. (The Singleton-Kautz and Friedman-Rubin procedures give identical results for four clusters.) Over all four procedures, only five schools vary in their cluster position.

Interpretation of the Clusters

The problem of deciding which is the best clustering is not well defined. Hence, the best grouping must be based on what the investigator purposes to do with the clusters. The set of clusters obtained using Johnsons' hierarchical procedure is used for further interpretation and statistical analysis. However, the other clustering procedures are suitable for analysis and produce similar results.

The three Johnson clusters are termed "lower average", "average", and "upper average", in terms of the school-community characteristics. For example, the lower average cluster is characterized by the following: Low average daily attendance, large class size, less use of innovative methods, low-cost residential areas, parents receiving the lowest yearly income, teachers receiving the lowest salaries, and over seventy-five per cent of the teachers are involved in inservice training. The teachers serving the lower average cluster as compared to those in the other two clusters have had less teaching experience; and none of these teachers holds an advanced degree. All of the teachers have a strong theoretical orientation; and they are also more involved in teaching than those in the other two clusters. The greatest percentage of female teachers is concentrated in this cluster.

THE STUDENTS' MATHEMATICS ACHIEVEMENT RESULTS

Several statistical analyses are performed in the analysis of the clusters using nine student test scales: Lorge-Thorndike Verbal, Lorge-Thorndike Nonverbal, Rationals-Computation, Rationals-Noncomputation, Whole Numbers, Geometry, Numbers-Whole, Algebra-Sentences, and Conversion. The first six variables termed covariates were administered during the fall of the first year of testing. The last three iables are used as variates and were administered during the spring of the third r of testing. The variates are used to measure the change in the students' hematics achievement over the three-year period.

Canonical Correlation Analysis

The covariates are used to measure (or predict) the change in the variates; hence, we should first determine if the differences among the variate means can actually be explained by the differences in the covariates. If this is the case then the two sets of variables are dependent and analysis of covariance methods may be used to remove the effects of variations in the covariates, insofar as these effects are measured by linear regression. It is important to note that the covariate scales need not be direct causal agents of the variates but may for example, merely reflect characteristics of the environment that also influences the variate scales.

In order to determine the dependence between the two sets of student test scales canonical correlation analysis is used to determine the correlation between the two sets of variables. The Chi-square test of significance developed to test the hypothesis that the p covariates are unrelated to the q variates is used in this study. All three of the correlations are significant at the .02 level. Hence, the domains are significantly related. The major variate is Numbers-Whole and the major covariates are Lorge-Thorndike Verbal and Lorge-Thorndike Nonverbal.

Multivariate Analysis of Covariance

We attempt to understand the nature of the clusters by looking at differences between the groups not only on measures of school-community and teacher characteristics; but also in terms of the students' mathematics achievement. Significant cluster differences are a reflection that the schools are not equally effective across clusters as measured by the students' mathematics achievement, after adjustments are made for competencies of the students. Whereas, nonsignificant differences are a reflection that the schools' characteristics do not influence the achievement level of the students.

The multivariate analysis of covariance results produced an F value of 9.14 using 6 and 5,986 degrees of freedom. Hence, the hypothesis of equality of treatment means following covariance adjustment is rejected at the .01 significance level. The means and standard deviations for the three clusters are presented in Table 1. The results of the univariate tests (Table 2) reveal that the most significant variate is Conversion followed by Algebra-Sentences. Numbers-Whole did not discriminate between the groups.

The "lower average" group produces the lowest student achievers as evidenced by the adjusted mean performances of the students on scales Numbers-Whole and Algebra-Sentences. The "average" group produces the lowest achievers on Conversion; and the "upper average" group produces the highest achievers on both Algebra-Sentences and Conversion. (See Table 3).

Hence, the mathematics achievement of the students across the clusters cannot be considered the same after adjustments have been made for differences in aptitude and initial understanding of mathematical concepts. Therefore, we conclude that the schools may not be considered equally effective. The observed differences between the adjusted means cannot be explained by the competencies of the students; but must be attributed at least in part to the effect of the school to which the student is assigned.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR EACH CLUSTER AND THE TOTAL GROUP

Student Test Variable	Total Possible Score	"Lower Average Group" Cluster I $N_1 = 119$		"Average Group" Cluster II $N_2 = 1898$		"Upper Average Group" Cluster III $N_3 = 978$		Total Group N = 2995	
Covariates		Mean	SD	Mean	SD	Mean	SD	Mean	S
Lorge-Thorndike Verbal--X_1	40	20.78	5.90	20.59	6.39	22.7	5.97	21.31	6.2
Lorge-Thorndike Nonverbal--X_2	58	33.42	7.99	33.20	9.81	37.22	8.34	34.53	9.2
Rationals-Computation--X_3	6	2.76	1.31	3.56	1.47	3.56	1.37	3.53	1.4
Rationals-Noncomputation--X_4	11	3.45	2.08	4.22	2.26	4.77	2.17	4.37	2.2
Whole Numbers--X_5	9	5.69	1.57	5.58	1.68	6.03	1.53	5.73	1.6
Geometry--X_6	4	1.00	0.98	1.19	1.02	1.25	1.01	1.20	1.0
Variates									
Numbers-Whole--Y_1	8	3.72	2.02	4.16	2.22	4.68	2.08	4.31	2.1
Algebra-Sentences--Y_2	6	2.13	1.63	2.67	1.82	3.14	1.81	2.80	1.8
Conversion--Y_3	12	5.41	3.25	5.33	3.38	6.67	3.30	5.77	3.3

TABLE 2

F-VALUES FOR DIFFERENCES BETWEEN CLUSTERS ON EACH VARIATE

		F^{2}_{2986}	P
Y_1:	Numbers-whole	0.86	.42
Y_2:	Algebra-sentences	4.07	.02
Y_3:	Conversion	20.43	.01

TABLE 3

THE ADJUSTED MEANS FOR THE THREE VARIATES

	Cluster I	Cluster II	Cluster III
Y_1--numbers-whole	4.13	4.32	4.30
Y_2--algebra-sentences	2.45	2.81	2.84
Y_3--conversion	6.04	5.53	6.20

BIBLIOGRAPHY

Ball, G. and Hall, D. (1965), "ISODATA", A Novel Method of Data Analysis and Pattern Recognition," Technical Report, Stanford Research Institute, Menlo Park, California, May.

Friedman, H. and Rubin, J. (1967), "On Some Invariant Criteria for Grouping Data", Journal of American Statistical Association, December, 1159-77.

Friedman, H. and Rubin, J. (1967) "A Cluster Analysis and Taxonomy System for Grouping and Classifying Data", IBM Corporation, New York Scientific Center, New York, August.

Johnson, S. (1967), "Hierarchical Clustering Schemes", Psychometrika, 32,241-54.

Singleton, R., and Kautz, W. (1965), Technical Report, Stanford Research Institute, Menlo Park, California.