Use of Clustering Techniques in a Social Stratification Study

Jean Stockard

NSF Undergraduate Research
Dec. 19, 1968
Gerald A. King, advisor

edited 6/13/49 - gs -

In The Conduct of Inquiry Abraham Kaplan states that "what makes a concept significant is that the classification it institutes is one into which things fall as it were of themselves."[1] It is common to hear people speak of social class as though the distinctions between the groups were as clear as Kaplan's significant concepts. Respondents to questionnaires are often asked to check one of three to five social classes of which they consider themselves a member. And in everyday conversations we glibly toss around phrases such as"middle-class" or "lower-class" as descriptive categories representing distinct value systems, economic groups, or ways of life. We seem to assume the existence of a stratification system ~~in~~ which embodies ~~there exist~~ discrete classes characterized by unique socio-economic attributes and life-styles.

Several social scientists feel this conception of discrete groups is valid and have supported their contentions with various studies, of which the work of Lloyd Warner is perhaps the most famous.[2] On the other hand, some sociologists have suggested continuum theories of social class ~~which~~ defined by Cuber and Kenkel ~~define~~ as the "idea that there are several privilege, power, and status ranges, more or less continuous from top to bottom with no clear lines of demarcation."[3] They claim that acceptance of this concept provides a sounder theoretical ground for social science. ~~An~~ One piece ~~example~~ of research supporting this theoretical perspective is Lenski's study of a New England village.[4]

Because the literature is indeed contradictory and there is dispute over the nature of social class, there is a clear need to approach the question in an empirical and objective manner which would meet the criterion suggested by Kaplan in his definition of a significant concept. Thus we decided to pursue the question by using a procedure designed to find the natural divisions within groups. The technique used may be called clumper analysis and belongs to the general group of clustering techniques which includes factor analysis, linkage analysis, multidimensional scaling, and hierarchic cluster analysis. Until recently, these techniques, with the exception of factor analysis, have been used almost exclusively by natural scientists, especially biologists and taxonomists who have employed them in checking the validity of previously determined classical biological groups.[5]

The rationale behind clustering techniques are fairly simple. They are basically methods for grouping variables or members of a group through a multidimensional analysis in an attempt to determine the natural divisions within the group. The particular techniques used in this study were first programmed by Sharon Roof of the SLCC at the University of Oregon and were later modified by Gerald A. King. Briefly it involved finding the Euclidian distance between the elements of a group within an n-space based on the indices of n selected variables. (The n in this study equaled 24.) The distance between each pair of individuals was then placed in a matrix of size (n-1)x(n-1). Clumps of similar individuals were then determined by either finding the closest pair of individuals and using this pair as a nucleus for adding other close individuals, or alternatively, finding the pair which had the greatest separation and splitting the groups into clumps from this basis. The objective of this study was to analyze a

set of data on the basis of socio-economic status variables using the clumper analysis techniques to see the extent to which distinct groups would appear in a way similar to that described by Kaplan, in other words, to examine the natural divisions within the sample on the basis of socio-economic characteristics.

Because the literature is so inconclusive regarding the nature of social class, no hypothesis representing an expected outcome was formulated. However, several possible results of the process were suggested, all centering on the extent to which distinct groups would appear. First it would be possible that no groups would appear at all; the sample would be homogeneous to an excessive degree. Secondly, groups could appear, but though the characteristics of ~~each~~ the groups could be ordered on an SES scale, they would not be unique to each group, but would tend to overlap, lending support to the continuum theories. The third suggested result was that distinct groups would appear and that socio-economic characteristics would be unique to ~~that~~ each group.

The question also arase as to whether life-style characteristics would differ between the various groups. Roger Brown has emphasized that the validity of a social class structure is high only if style of life characteristics differ between and are unique within the classes.[6] Thus we would look for unique style of life characteristics within the groups of the third possible result and overlapping or homogeneous characteristics within groups of the other suggested results.

The data used in the study was collected by Dr. Ted Johannis in 1953 and was not gathered for purposes of a stratification analysis. The total sample involved 1500 tenth grade students from the high schools of Tampa,

In a statistical analysis of differences of the means (which test to show the data) if data involving ch. affiliation + years mother worked is eliminated, it is found that the average distribution of the groups is not significantly different.

Florida from which a subsample of 130 members was selected for the cluster analysis, using the different schools of the city as a proportional basis for the selection. For the clustering process twenty-four variables deemed indicative of socio-economic class were chosen.

Three main groups seemed to emerge from the clumper analysis. To determine the nature of the groups the means, standard deviations, and distributions for each variable were calculated. Table 1 shows part of the results. For most of the variables a lower mean indicates a higher place on the socio-economic scale. The table reveals that one group tends to have the lowest educational and occupational levels for the parents, the lowest type of income source, homes in the worst condition and in the most undesirable neighborhoods. The largest group is midway to the other two when most of the means are compared and the third group is the highest in terms of the socio-economic variables. The sample as a whole seems to be skewed toward the lower socio-economic echelons. Thus the highest group may represent only what is termed an "upper-middle class." Also those at the extremely lower end of the scale are missing, for few youngsters from that background stayed in school until the tenth grade. Though not shown on the table, the middle group also tended to subdivide, with one group higher than the other on thatsocio-economic scale.

It must be noted that the standard deviations for the indices of each group are not greatly, if at all, reduced from those of the total sample. There appears to be a great deal of overlapping between groups on the distribution of each variable. It seems then that although distinct groups could be determined through the clumping process, they are by no means completely homogeneous in nature. Instead an overlapping of characteristics of the groups seems to occur, which would lend support to the continuum theories.

The few differences that could be determined in the limited review possible of information relating to life styles were mainly in the area of home life. Table 2 lists some of the variables studied. Of all the respondents, those in the middle socio-economic range seemed to live with their own parents more and to perceive their home life as happier than members of other groups did. Note again the very slight reductions in standard deviations in each group.

Several criticisms may logically be leveled at the study. First the data is fifteen years old, from a southern town, and is perhaps inapplicable to today's situation. The absence of members of the extreme upper and lower ends of the socio-economic spectrum is also regrettable. Secondly the data is based upon the impressions of tenth graders of their conditions and surroundings and the validity of this information for the purpose of this study could be questioned. Most important, we must seriously question the validity of some of the indices used as indicators of social class and the nature of the scales employed.

The analysis does seem to indicate that the ordering of respondents on a socio-economic scale occurs when an objective method of looking at the data is used. Differences in a few life-style characteristics between the groups were also noted. However, the overlapping between the groups is so great that it is actually impossible to determine discrete groups, lending support to the continuum theories of the nature of social class.

Despite the drawbacks of this study we feel that the methods employed are ones which should be investigated further, and that with intensive and careful studies would have to be seriously considered.

Table 1

Means and standard deviations on selected status variables for clumper analysis groups.

( ) indicates # in ~~si~~ data + printouts + most of discussion

Variable	(I) Groups — High SES I (N=22) mean (s.d.)	(III) Mid SES II (N=58) mean (s.d.)	(II) Low SES III (N=30) mean (s.d.)	Total Sample mean (s.d.)
Educational Level Father (1-7)	4.562 (1.499)	5.289 (1.512)	5.444 (1.133)	5.227 (1.458)
Occupational level (1-7)	3.937 (1.391)	4.434 (1.370)	4.759 (1.194)	4.441 (1.354)
Church affiliation - father (11-35)	20.857 (6.501)	12.500 (2.062)	17.667 (6.493)	14.730 5.248
Number of organizational Memberships - father (1-9)	2.933 (1.482)	2.481 (1.640)	2.370 (1.636)	2.513 (1.629)
Educational Level Mother (1-7)	4.222 (1.474)	5.171 (1.293)	4.464 (1.239)	4.869 (1.367)
Occupational Level Mother (1-6)	6.190 (1.735)	6.922 (1.620)	6.467 (2.093)	6.695 (1.788)
Type House (1-7)	2.048 (1.214)	2.104 (1.100)	2.464 (1.476)	2.175 (1.222)
Part of Town (1-7)	2.600 (0.735)	2.720 (0.684)	2.889 (1.030)	2.738 (0.787)
Income Source (1-6)	4.158 (0.670)	4.333 (0.827)	4.532 (0.674)	4.357 (0.781)

Table 2

Means & standard deviations on selected life style variables for cluster analysis groups

Variable	I (N=22) Hi SES mean (s.d.)	II (N=72) Mid SES mean (s.d.)	III (N=30) Low SES mean (s.d.)	Total sample mean (s.d.)
who respondent lives with (1-9) 1-one parent	3.318 (2.182) 3	1.423 (1.354) 1	2.367 (1.906) 2	1.962 (1.808)
family size act. number (1-9)	2.818 (1.850) 1	3.013 (1.613) 2	3.200 (1.492) 3	3.023 (1.634)
where father raised (1-9) 1 local 9-for.	2.818 (2.587) 1	2.610 (2.523) 2	2.241 (2.095) 3	2.562 (2.452)
# times of resp with father (1-5) 1 hrs/week	1.706 (0.892) 3 2	1.442 0.860 1	1.704 (0.853) 2 3	1.537 0.872
Relations with Mother (1-5) 1-hrs/week	1.952 (1.090) 3	1.513 (0.888) 1	1.517 (0.815) 2 1	1.586 (0.923)
Relations with siblings (1-5) 1-hrs/week	2.952 (1.731) 3	2.675 (1.694) 2	2.517 (1.221) 1	2.685 (1.611)
Satisfaction with home life (1-5) 5-best	3.952 (1.045) 3	4.321 (0.913) 1	4.167 (0.860) 2	4.225 (0.934)
Perceived happiness of parents' marr. (1-5) 5-best	4.500 (0.500) 1	4.218 (1.033) 2	4.200 (0.800) 3	4.252 (0.937)
Disagreement w/ family, result in (1-4) 1-auth.	3.000 (0.926) 1	2.608 (0.970) 3	2.714 (1.119) 2	2.679 (1.003)

Footnotes

[1] Abraham Kaplan, The Conduct of Inquiry; Methodology for Behavioral Science (San Francisco, Chandler Publishing Co., 1964).

[2] See Lloyd Warner, Yankee City (New Haven, Yale University Press, 1963), Social Class in America (New York, Harper Torchbook, 1960), Democracy in Jonesville, (New York, Hay, 1949). Also see Davis, Allison, Gardner and Gardner, Deep South (Chicago, Univ. of Chicago Press, 1941) and Bendix and Lipset, Class, Status, and Power (New York, 1966).

[3] J.F. Cuber and W.F. Kenkel, Social Stratification in the United States (New York, Appleton-Century-Crofts, 1959).

[4] G.E. Lenski, "Prestige, Status, & Wealth," American Journal of Sociology, (Sept., 1952), pp. 139-144.

[5] See Final Report, Conference on Cluster Analysis of Multivariate Data New Orleans, La., Dec 9,10,11, 1966 (Washington, D.C., Catholic University of America, June, 1967); Geoffrey H. Ball, "A Comparison of Some Cluster Seeking Techniques," (Stanford Research Institute); Ball and D.J. Hall, "A Clustering Technique for Summarizing Multivariate Data," Behavioral Science 12, (March, 1967) pp. 153-155; and S.C. Johnson, "Hierarchical Clustering Schemes, Psychometrika 32, (Sept., 1967), pp. 241-254.

[6] Roger Brown, Social Psychology, (New York, 1965).