ON GROUPING FOR MAXIMUM HOMOGENEITY*

WALTER D. FISHER
Kansas State College

Given a set of arbitrary numbers, what is a practical procedure for grouping them so that the variance within groups is minimized? An answer to this question, including a description of an automatic computer program, is given for problems up to the size where 200 numbers are to be placed in 10 groups. Two basic types of problem are discussed and illustrated.

1. INTRODUCTION

STATISTICIANS are often interested in defining homogeneous groups. Measures of precision of point estimates partly depend on homogeneity within strata from which samples are taken. Tests of significant differences are based on comparisons that also involve such homogeneity within strata, as well as differences between them. Apart from sampling or inference problems, it is often important to know how a population may be decomposed into sub-groups that contrast sharply with each other, individuals of the same group being fairly alike.

This paper deals with the following problem from the viewpoint of statistical description: given a set of K elements, each element having assigned to it a weight, w_i, and a numerical measure, a_i, and given a positive integer G that is less than K; to find a systematic and practical procedure for grouping the K elements into G mutually exclusive and exhaustive subsets such that the weighted sum of squares

$$D = \sum_{i=1}^{K} w_i(a_i - \bar{a}_i)^2 \tag{1}$$

is minimized, where $\bar{a}_i$ denotes the weighted arithmetic mean of those a's that are assigned to the subset to which element i is assigned. This problem will be called a *grouping problem.*[1] The D value, well known as the sum of squares within groups in the sense of the analysis of variance, will here be called *squared distance.* A system of grouping is often called a partition, and a partition associated with the minimum squared distance D will be called a *least squares partition.*

Two subclasses of the grouping problem are distinguished: (1) the *unrestricted problem,* where no restrictions or side conditions are imposed on the partitions allowed; and (2) the *restricted problem,* where such conditions are imposed *a priori* on the basis of previous knowledge, theory, or for convenience. The relevance of each type of problem will be illustrated by an example from

* This paper is the result of work supported by the Social Science Research Council and by the Bureau of General Research of Kansas State College; also by the courtesies extended to the author by the University of Illinois when he visited there during the summer of 1956. The author gratefully acknowledges the helpful suggestions he received from W. A. Neiswanger, Kern Dickman, and D. B. Gillies of the University of Illinois; from J. I. Northam, now with the Upjohn Company; and from Henry Tucker, now of the University of Arizona.

[1] An equivalent geometric problem is: given K weighted points on a straight line, to group the points into G groups so that sum of squared distances of the individual points from their group centers of gravity is minimized.

the literature, and computational methods for handling the simpler types of problem will be presented.

An analogous problem for the case of a continuous frequency distribution has been investigated by Dalenius [3], [5] and Dalenius and Gurney [4]. The special case of the normal distribution has been considered in a recent note by Cox [2]. The methods suggested by these writers are useful in the discrete problem considered here when the number of individuals is large, when their distribution can be approximated by a fairly simple continuous curve, when the number of groups G is fairly small—say five or less—and when no side conditions are put on the admissible partitions. Otherwise, the approach of the present paper is believed to be preferred.

The position taken here that the w_i and a_i are given and known with complete certainty entails a descriptive or non-stochastic approach; yet this approach leads to sampling and other stochastic applications. It is moreover assumed without attempt at justification here that the measure of homogeneity used, D, formed by adding squared deviations, is useful and relevant to many practical problems.[2]

2. THE UNRESTRICTED PROBLEM

The unrestricted problem may be easily understood by considering a familiar situation. Assume that it is desired to find the best method of choosing a given number of strata for proportional-stratified sampling when information is available regarding the relevant variable in the population.[3] In their discussion of stratified sampling Hansen, Hurwitz and Madow [8] present a problem with data on income levels of Atlanta families, based on a previous study by Mendershausen [9]. A frequency distribution of these families, grouped into ten income classes, is shown in Fig. 791. The problem is to combine the ten classes into three larger strata so that the estimate of mean income for all families, based on a stratified sample, has a small variance.[4] Various strata and various methods of sampling are suggested. Here attention will be confined to the various possible combinations of the original classes into strata, assuming proportional allocation of sample numbers between the three strata and random sampling within each stratum. It is also assumed that the sample mean is taken as the estimate of the overall population mean. It has been shown, and is well known, that under these conditions, if w_i denotes the weight of income class i in the population, a_i denotes the mean income of income class i, and $\bar{a}_i$ denotes the mean income of the stratum to which income class i is assigned, then the variance of the estimate is proportional to D as given by equation (1) plus a constant representing the variance within the original classes. To minimize the variance of the estimate it is sufficient to minimize D, which repre-

[2] Savage has given a general theoretical argument in support of the squared error criterion for statistical decision and estimation problems [10, Ch. 15]. In a previous article [6] the present writer deduced the squared error criterion from a specific economic decision problem, in which uncertainty was also introduced.

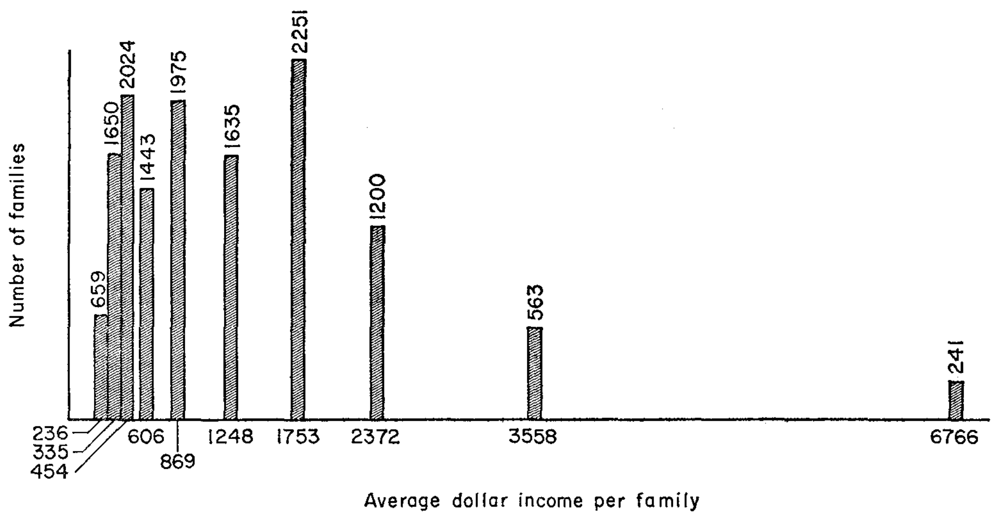
[3] *Strictly speaking, the problem to be discussed is based upon the assumption that the stratification variable is* identical with the variable to be estimated. The stratification method is useful, without this assumption, when an *a priori* stratification variable can be found that is highly correlated with the variable to be estimated.

[4] See [8], Exercise 17.2 to 17.5 inclusive. In our Figure 791 two of the original eleven groups having nearly the same mean have been combined for graphic convenience, the combination being *immaterial* to the present problem.

sents the variance of the ten income class means within the three strata. If we replace "income class" by the word "element," the original grouping problem of the paper has been restated.

The solution to this problem—not obvious from a visual inspection of Fig. 791—happens to be the following one. Numbering the original ten classes from low to high income (from left to right in Fig. 791), put classes 1 to 6 in one stratum, 7 to 9 in another, and 10 in a third by itself. This particular method of stratification is not mentioned by Hansen, Hurwitz and Madow, nor by Dalenius and Gurney, who also discuss this same example.[5] In this small problem

FIG. 791. Atlanta families by income in 1933.



Source: Hansen, Morris H., Hurwitz, William N., and Madow, William G., *Sample Survey Methods and Theory*, Vol. I, New York: John Wiley and Sons, 1953, p. 235.

it is quite feasible to find the solution by hand computation. It is intuitively obvious, and it can be proved, that when the ten original classes are ordered according to income ($i<j$ when $a_i<a_j$), the only partitions that need be considered are *contiguous partitions*, defined for a set of completely ordered elements as a partition that consists entirely of subsets satisfying the following condition: if elements i, j, and k have the order $i<j<k$, and if elements i and k are assigned to the same subset, then element j must also be assigned to that same subset.[6] To find the optimal grouping it is therefore sufficient to compute the D values for each of the 36 possible contiguous partitions of ten elements into 3 groups, and then select one with minimum D.[7]

---

[5] See [8] and [4], pp. 144–146. These writers consider some alternative solutions, including that obtained under conditions of optimal allocation of sampling numbers. For both problems values of the variance function for "nearly optimal" stratifications do not differ greatly from each other.

[6] Proof that a least squares partition is always contiguous is given in the Appendix.

[7] The number 36 is obtained from the formula in the next paragraph. The solution to this problem was actually obtained by an automatic computation to be mentioned in Section 4 below, and checked by hand computation.

The general unrestricted problem of K elements into G groups can, by the same reasoning, be reduced to a consideration of $\binom{K-1}{G-1}$ contiguous partitions.[8]

3. THE RESTRICTED PROBLEM

In the preceding example the investigator, in solving the grouping problem, was at liberty to take the preliminary step of ordering the ten original classes by income level, and then seek a contiguous partition of these ten elements so ordered that minimized the squared distance. Assume for the moment, however, that it was desired to "respect" some *a priori* ordering of these elements which might be different than the income order. In other words, while still seeking a minimal squared distance, a solution would be considered admissible only if it were a contiguous partition of the elements ordered in the *a priori* manner. For example, suppose that the ten original elements were families living on the same street in a known order of location, and that it is still desired to create three groups of families with maximum homogeneity measured in terms of income, but that it is also desired that these groups all be contiguous in terms of location. This latter condition of the problem may be regarded as a side condition or restriction imposed on the minimization of D. It is obvious that this problem is different from the old one; it is possible that the solution to it may not attain as low a D as the solution to the old one.

Other types of side conditions may be imagined. It may be required that the solution involve only a partial ordering of the elements with respect to some criterion, as contrasted with a complete ordering. The given elements could be associated with points in some space of more than one dimension, apart from the values of the a_i, and mathematical restrictions could be imposed on the coordinates of these points.[9] Certain partitions of the elements may be barred explicitly, irrespective of any concept of ordering or spatial location. The grouping problem defined in the second paragraph of this paper will be called a *restricted problem* if any *a priori* restrictions whatever are placed on the set of partitions of the K elements into G subsets that are regarded as admissible for a solution (other than the requirement that the subsets must be mutually exclusive and exhaustive).

Most practical problems will be of the restricted type, since the investigator will almost always wish to inject prior knowledge, or factors of convenience into the conditions of the grouping. In fact, the class of restricted problems is so large that a general approach seems extremely difficult if not impossible. Even a definition of the major categories of restrictions that seem to be significant for practical applications is beyond the scope of this paper. It will suffice to present a larger numerical example that illustrates the solution of a problem having the simple type of restriction first mentioned: a complete *a priori* ordering of the elements that is different than the numerical order of the a_i.

In the discussion of time series in their textbook Wallis and Roberts [11]

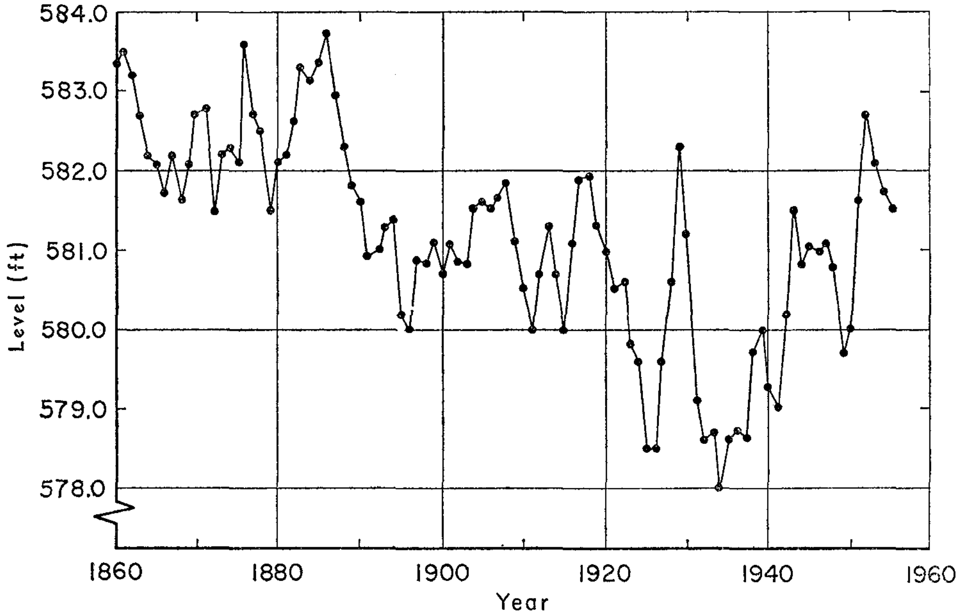
---

[8] A contiguous partition of K completely ordered elements into G subsets may be represented by $G-1$ points of division lying in any of the $K-1$ intervals between adjacent elements, imagined to lie on a line in the specified order. The number of possible contiguous partitions will therefore equal the number of ways of choosing the division points, which is the number of combinations of $K-1$ different things taken $G-1$ at a time.

[9] The notion of "property space" is applicable to such a scheme. See [1]. The numerical example in [6] includes certain restrictions on the K elements in a three dimensional property space.

present an example of the change of lake levels over a time span of 96 years. Their graph, reproduced in Fig. 793, suggests certain epochs when the lake level was high and others when it was lower, although no obvious regularity or periodicity is apparent. Their analysis of the phenomenon is largely in terms of runs and moving averages. Suppose that it were desired to define G epochs such that the variation of lake level within epochs, as defined by squared distance D, is minimized. It is of course required that each epoch comprise only consecutive years in time: this is the *a priori* ordering. Then a restricted prob-

FIG. 793. Lake Michigan-Huron highest monthly mean level, 1860–1955.



Source: Wallis, W. Allen and Roberts, Harry V., *Statistics: A New Approach*, Glencoe, Illinois: The Free Press, 1956, p. 567.

lem of the type formulated above results, with all weights w_i equal to 1. The solution must be a contiguous partition in terms of the ordering in time—not necessarily the ordering according to lake level.

The solution to the problem for G ranging from 1 to 10 is given in Table 794. This solution is provided by an automatic computer program to be mentioned in the next section. Alternative values for G are listed in the first column. The minimized values of D are listed in the second column. Each row of the triangular array headed "P" identifies an optimal partition that yields a solution to the problem for the G value of that row, giving the minimized D value of that row. The optimal partition is identified by the order number of the highest-order element of each subset, except the last (highest), which is always 96. For example, for $G=3$ the three epochs are: years 1 to 30, years 31 to 61, and years 62 to 96. For $G=1$ the solution is of course trivial; the relevant partition is simply the original set, and the D value is the sum of squared deviations from the general mean. It should also be remarked that this program does not yield multiple solutions when they exist. The program is discussed further below.

TABLE 794

AUTOMATIC COMPUTER SOLUTION TO LAKE MICHIGAN-HURON PROBLEM

G	D	P								
1	16614									
2	8673	31								
3	7400	30	61							
4	4969	30	63	83						
5	4463	30	63	71	82					
6	3580	30	63	68	71	82				
7	3080	30	63	68	71	82	91			
8	2857	23	29	63	68	71	82	91		
9	2559	3	23	29	63	68	71	82	91	
10	2359	3	23	29	63	68	71	78	82	91

4. METHODS OF SOLUTION

For small unrestricted problems—say, of the order of $K \leqq 20$, $G \leqq 5$—it is feasible to obtain the solution by complete enumeration of all possible contiguous partitions by hand computation of D, and selection of a partition having minimal D. Some restricted problems of the same order will yield to the same treatment. For the restricted problem the number of admissible partitions for which D values must be computed may be less than for an unrestricted problem of the same size, but the task of applying the restrictions to obtain admissible partitions consumes additional time.

For some unrestricted problems where K is large but G still small it will be possible to obtain the solution or a near-solution by visual inspection of the frequency distribution of the a_i, ordered according to their magnitudes. Divisions between groups may be placed where the data are sparse or the weights of small magnitude. This principle cannot be so readily applied when the number of such regions of sparse data does not correspond with the number of divisions to be made. When the frequency distribution of the a_i can be represented or closely approximated by a continuous function of a fairly simple form—say by one that is not multi-modal—the method of Dalenius [3], [5] can be applied.[10]

For the general problem with arbitrary distributions of the a_i and with larger K or G, when the special devices noted above are not applicable, a combina-

[10] This method is based on the principle that for a continuous frequency distribution a necessary condition for minimum D is equidistance between any point of division between two adjacent subsets and the two means of the subsets. Dalenius outlines an iterative method for attaining this condition from an initial trial division. It has not yet been shown, however, for what class of frequency distributions this necessary condition is also sufficient; and examples can be found for which the condition is not sufficient, even if the usual conditions on the derivatives for a minimum D are also assumed. For example, if the given frequency distribution has extreme tri-modality, the D function for a division into two groups may have two local minima, either one of which may be approached by Dalenius' iterative procedure, and so the *minimum minimorum* may have to be ascertained by further examination. Dalenius has acknowledged this fact [5, p. 165].

torial approach seems indicated.[11] In principle, since the number of possible partitions of K elements into G subsets is finite, it would still be possible to find the solution by consideration of each partition and selection of the one (or those) having minimum D. Unless this approach is modified, however, the number of combinations becomes so large (as indicated by the formula proved in Footnote 10) that consideration of all possibilities becomes impractical, even for the fastest digital computers now in existence. For example, the number of contiguous partitions for a problem of the size of the Lake Michigan-Huron problem, where $K=96$, $G=10$, is slightly over one trillion. A high speed Princeton-type computer, computing D values at the rate of 10 milliseconds per partition, and working 24 hours a day, would require more than 280 years to compare these partitions one by one.

Because of the additive character of the squared distance function, however, it is possible to reduce the computations very substantially by the use of suboptimization procedures. Such procedures are implied by the following lemma.

> *Suboptimization Lemma:* If $A_1 \vdots A_2$ denotes a partition of set A into two disjoint subsets A_1 and A_2, if $P_1{}^*$ denotes a least squares partition of A_1 into G_1 subsets and if $P_2{}^*$ denotes a least squares partition of A_2 into G_2 subsets; then, of the class of subpartitions of $A_1 \vdots A_2$ employing G_1 subsets over A_1 and G_2 subsets over A_2 a least squares subpartition[12] is $P_1{}^* \vdots P_2{}^*$.

In other words, once a least squares partition over set A_1 has been found, this work need not be done over again when testing for various partitions over A_2, providing that suitable records are kept. It is apparent that application of this lemma makes it possible to avoid separate consideration of many possible partitions of the entire set of elements. The extent of the saving of time is indicated by the fact that when the lemma was applied, the solution of the Lake Michigan-Huron problem for all G from 1 to 10 was actually obtained in 3 minutes.

The lemma will also hold if "least squares partition" is found under side conditions on admissible partitions, and hence is applicable to restricted problems.

A program for the "Illiac" automatic digital computer at the University of Illinois has been written and checked by the author for solving the unrestricted grouping problem, or the restricted problem when the elements are com-

[11] It has been pointed out to the writer by George B. Dantzig in correspondence that the unrestricted grouping problem can be formulated as a non-linear programming problem by the use of special variables that assign the elements to groups. The usefulness of this parallel is limited, however, by the non-availability of computational algorithms for the type of programming problem where a strictly concave objective function is to be minimized on a convex set. Allowing for fractional assignment, let x_{hi} denote the fractional part of a_i that is assigned to group $h (h=1 \cdots G;$ $i=1 \cdots K)$, set $\bar{a}_h = \Sigma_{i=1}^K x_{hi} w_i a_i / \Sigma_{i=1}^K x_{hi} w_i$, $S = \Sigma_{h=1}^G \Sigma_{i=1}^K x_{hi} w_i (a_i - \bar{a}_h)^2$, and consider the problem of minimizing S subject to the constraints $x_{hi} \geqq 0$ and $\Sigma_{h=1}^G x_{hi} = 1$. With given a's and w's, S is strictly concave in the GK dimensional space of the x_{hi}, the constraint set is a convex polyhedron in this same space, and minimum S is attained only at extreme points of this constaint set. An extreme point corresponds with a matrix $[x_{hi}]$ having a single unit element in each column, all other elements being zero (x_{hi} being unity when a_i is "completely" assigned to group h, and zero when a_i is not assigned to group h). Then the problem is precisely equivalent to the grouping problem of this paper, $\bar{a}_h$ becomes a group mean, S becomes equivalent to our D, and for a solution no fractional assignment is possible. Moreover, the attainment of a local minimum, in the sense that D cannot be lowered by changing the assignment of any single a_i, does not guarantee that the absolute minimum has been attained. All this emphasizes that the problem is essentiallv a combinatorial one.

[12] Proof: Let P_1 and P_2 denote partitions of A_1 and A_2 into G_1 and G_2 subsets respectively. Let D_1, D_2, D_{12}, $D_1{}^*$, $D_2{}^*$, $D_{12}{}^*$ denote the squared distances associated with partitions P_1, P_2, $P_1 \vdots P_2$, $P_1{}^*$, P_2, $P_1{}^* \vdots P_2{}^*$ respectively. From the definition of least squares partition $D_1{}^* \leqq D_1$ and $D_2{}^* \leqq D_2$. From the definition of D in equation (1) $D_{12}{}^* = D_1{}^* + D_2{}^*$ and $D_{12} = D_1 + D_2$. Hence $D_{12}{}^* \leqq D_{12}$, and so $P_1{}^* \vdots P_2{}^*$ is a least squares partition.

pletely ordered *a priori*,[13] with capacity $K \leqq 200$, $G \leqq 10$. The program solves the problem to the extent of identifying one optimal partition and its associated squared distance for $G=1, 2, \cdots, \bar{G}$, where $\bar{G}$ is specified as the maximum number of subsets to be considered, and cannot exceed 10. Machine running time for the largest possible problem is approximately 14 minutes, including input and output.

Input of the data in the specified order is made on standard teletype tape. If an unrestricted problem is to be solved, then the w_i and the a_i must be put into the computer in order according to the numerical values of the a_i. The basic method of solution is to have the "Illiac" systematically identify and compute D values for all partitions that are relevant after consideration of contiguity and application of the suboptimization lemma. In applying the lemma to the problem of G subsets, systematic use is made of certain results obtained and recorded while working the problem of $G-1$ subsets. The solution to a problem appears in the form of Table 794 and has $\bar{G}$ rows. In fact Table 794 is a precise reproduction of the output of a particular problem as it emerged from the page printer, with the exception that some non-significant decimals of the printed D values have been omitted.

The computational program described has at least three limitations; the magnitude of K and G it will accommodate is still quite limited, and even with modifications to handle further increases in K and G, computation time will press on reasonable limits; the program will not identify multiple solutions or near-solutions; and it will not handle restricted problems other than the special case of complete one-dimensional ordering. It is to be hoped that future progress will overcome these shortcomings.

5. GENERALIZATION

Some ways of generalizing the grouping problem as formulated in this paper will be briefly indicated. A stochastic approach to the problem is presented in [6], as well as a rationale for dropping the assumption of fixed G, making the selection of G a part of the decision, which depends on the value of more detailed information as compared with the extra "cost of detail." Even without such an explicit theory of cost, knowledge of the change in D resulting from change in G (see, for example the second column of Table 794) may assist the investigator in making a decision on what G he wants to use when he is initially uncertain. The "Illiac" program was designed to provide this information.

While mention was made in Section 3 of the possibility of specifying a property space in more than one dimension, the idea of a single dimension for measuring squared distance D was retained. It would of course be most desirable to develop, both theoretically and computationally, a distance criterion that is defined in more than one dimension. An example of the need for such a formulation is shown in a multivariate stratification problem encountered in a sample

[13] Instruction in programming and a key suggestion that made the writing of this program possible was given the author by D. B. Gillies of the Department of Mathematics of the University of Illinois. Invaluable aid in certain aspects of programming and debugging was given by Kern Dickman, Computer Consultant at the University of Illinois. A copy of the instruction tape entitled "Optimal Partition of Discrete Points" is available in the Office of the Computer Consultant, and also in the hands of the author.

survey by Hagood and Bernert [7]. Of course involved in any such approach is a relevant system of weighting the different dimensions to reflect their relative importance in determining distance.

The one-dimensional approach of this paper may be used to provide an approximate solution to a multi-dimensional grouping problem. As a preliminary step the data may be reduced to single measures on each element by extracting the first principal component, or by other methods of factor analysis. Such procedures are now well known and routinized on many types of computing machines. Then the computational methods suggested above may be applied to group the elements into the desired number of groups. The goodness of this approximation will, of course, depend on the degree of dominance of the first principal component in the multi-dimensional scatter.

APPENDIX: PROOF THAT A LEAST SQUARES PARTITION IS CONTIGUOUS

Consider a non-contiguous partition as defined in the text. Let elements i and k belong to a subset having mean $\bar{a}_{ik}$, while j belongs to a different subset having mean $\bar{a}_j$, and where $a_i < a_j < a_k$. Then, whatever be the values of $\bar{a}_{ik}$ and $\bar{a}_j$, at least one of the following three statements is true:

$$|a_j - \bar{a}_j| \geq |a_j - \bar{a}_{ik}| > 0, \tag{1}$$

$$|a_i - \bar{a}_{ik}| \geq |a_i - \bar{a}_j| > 0, \tag{2}$$

$$|a_k - \bar{a}_{ik}| \geq |a_k - \bar{a}_j| > 0. \tag{3}$$

In other words, of the three distinct points, a_i, a_j, and a_k, there exists one whose distance from the mean of its own subset is equal to or greater than its distance from the mean of another subset, both distances being positive. Relabelling such a point as "a," its own subset A with mean $\bar{a}$, and the "foreign" subset B with mean $\bar{b}$, we have

$$|a - \bar{a}| \geq |a - \bar{b}| > 0. \tag{4}$$

From definition (1) of the text the squared distance associated with the given partition may be written

$$D = \sum_{i=1}^{K} w_i a_i^2 - W_A \bar{a}^2 - W_B \bar{b}^2 - R, \tag{5}$$

where $W_A = \sum_{i \epsilon A} w_i$, $W_B = \sum_{i \epsilon B} w_i$, and R denotes a weighted sum of squared means of subsets other than A and B.

Consider the new partition formed by transferring point a from subset A to subset B. Let A' with mean $\bar{a}'$ and B' with mean $\bar{b}'$ denote the new subsets after the transfer. Since from (4) point a was distinct from the mean $\bar{a}$, set A contained at least two points; hence both A' and B' contain at least one point, and the new partition has the same number of subsets as the old. The new means $\bar{a}'$ and $\bar{b}'$ can be determined from the relationships

$$(W_A - w)\bar{a}' = W_A \bar{a} - wa, \tag{6}$$

$$(W_B + w)\bar{b}' = W_B \bar{b} + wa, \tag{7}$$

where w is the weight of point a. The squared distance associated with the new partition is

$$D' = \sum_{i=1}^{K} w_i a_i^2 - (W_A - w)\bar{a}'^2 - (W_B + w)\bar{b}'^2 - R. \tag{8}$$

By subtracting (8) from (5), eliminating $\bar{a}'$ and $\bar{b}'$ by means of (6) and (7), and simplifying, it follows that

$$D - D' = w\left[\frac{W_A}{W_A - w}(a - \bar{a})^2 - \frac{W_B}{W_B + w}(a - \bar{b})^2\right]. \tag{9}$$

From (4) and from the fact that all of the weights are positive with $W_A > w$, the right-hand side of (9) is found to be positive, and hence $D > D'$. Hence any non-contiguous partition can always be altered to give another partition with the same number of subsets and with smaller squared distance. Hence a least squares partition must be a contiguous partition.

REFERENCES

[1] Barton, Allen H., "The concept of property space in social research," in Lazarsfeld, Paul F. and Rosenberg, Morris, *The Language of Social Research*, Glencoe, Illinois: The Free Press, 1955, 40–54.

[2] Cox, D. R. "Note on grouping," *Journal of the American Statistical Association*, 52 (1957), 543–7.

[3] Dalenius, Tore, "The problem of optimum stratification," *Skandinavisk Aktuarietidskrift*, (1950), 203–13.

[4] Dalenius, T., and Gurney, M., "The problem of optimum stratification, II," *Skandinavisk Aktuarietidskrift*, (1951), 133–48.

[5] Dalenius, Tore, *Sampling in Sweden*. Stockholm: Almqvist and Wiksell, 1957.

[6] Fisher, Walter D., "On a pooling problem from the statistical decision viewpoint," *Econometrica*, 21 (1953), 567–85.

[7] Hagood, Margaret Jarman, and Bernert, Eleanor H., "Component indexes as a basis for stratification in sampling," *Journal of the American Statistical Association*, 40 (1945), 330–41.

[8] Hansen, Morris H., Hurwitz, William N., and Madow, William G., *Sample Survey Methods and Theory*, Vol. I, New York: John Wiley and Sons, 1953, 235.

[9] Mendershausen, Horst, *Changes in Income During the Great Depression*, New York: National Bureau of Economic Research, 1946.

[10] Savage, Leonard J., *The Foundations of Statistics*, New York: John Wiley and Sons, 1954, 229–34.

[11] Wallis, W. Allen and Roberts, Harry V., *Statistics: A New Approach*, Glencoe, Illinois: The Free Press, 1956, 565–9.