Paul Switzer Stanford University

1. Of the many different problems in clustering and pattern recognition we have chosen to examine several which have the following common features:

The objective is to partition the individuals of a specified population into subsets.

The permitted partitions may depend only on the values of a specified measurement variable X defined for this population.

2. The properties above imply that our problems are equivalent to partitioning the space of all possible values of X; in particular, two individuals with the same X value must not be in different subsets of the partition. Quite a variety of clustering and pattern recognition problems can be regarded in this framework and they are distinguished by what we hope to accomplish with our chosen partition.

3. Definitions: A subset of X values will be called a <u>stratum</u>, especially when it refers to one of the subsets of our chosen partition. A collection of individuals whose X values all belong to the same stratum will be called a <u>cluster</u>. So, for a given stratification of X and a given group of individuals, there is a unique decomposition of the group into clusters. The definition of a cluster and a stratum are equivalent only in the case of a finite population where no two individuals assume the same X value.

4. Often, what we hope to accomplish with a chosen partition is not well articulated, as we shall see later. Occasionally, we can be quite explicit however, as the next several examples will illustrate.

Equal probability partitions. Here the ob-5. jective is to get k strata each with probability content 1/k; i.e., each stratum contains an equal fraction of the population. Without further requirements this objective is easily achieved, in general, if we know enough about the distribution of X values. The problem becomes interesting when this distribution is unspecified. Suppose X has a continuous distribution on a Euclidean space, and we have a random sample of m individuals with observed values X,X, ,...,X,; Fraser [1] has given a general class of partitioning procedures, based on the sample, with the property that the expected probability content of each of the strata is r/(m+1) where r is any specified divisor of m+1. If X is real-valued, for example, the sample points themselves partition the line into m+l intervals, each with expected probability content 1/(m+1).

6. As a second example with an explicit objective consider the k-means problem - as it is sometimes called. What we need to specify is k, the number of strata, and a distance function ρ defined on the space of X. Once ρ has been specified, the mean μ_h of stratum h is defined to be that point in X space which minimizes $V_h = E[\rho^2(X,\mu_h)|X\varepsilon$ stratum h]. If W_h is the probability content of stratum h, then the total dispersion within strata is taken to be $V = \Sigma W_h V_h$. The objective is to choose a partition which minimizes this total within strata dispersion. (For example if ρ is Euclidean distance then μ_h is the usual stratum center of gravity.)

7. MacQueen [2] has discussed this problem and indicated how it might arise in real situations. One illustration not developed there concerns estimation of the mean of X when X is real-valued: If the allocation of a random sample to strata is proportional to the stratum weights, then the mean-square-error of the sample mean is indeed proportional to V if $\rho^2(a,b)$ is taken to be $(a-b)^2$.

7a. In practice, when we are estimating the unknown mean value of X in a population, it will not be possible to stratify the sample on the values of X itself. Instead what is usually done is to find a variable Y which is strongly correlated with X and whose distribution is pretty well known, then stratify according to Y values. In principle, there exists an optimum partition of Y which will minimize the V quantity for the X variable.

8. If the allocation of the sample to strata in the preceding example is according to Neyman optimal allocation (see Cochran [3]) then the appropriate quantity to minimize is $\Sigma W_h \sqrt{V_h}$. Dalenius and Hodges [4] have, for this case, obtained an approximate solution to the partitioning problem: On the assumption that k is fairly large, the distribution of X (assumed continuous) will be nearly uniform within strata. Using this uniformity, the optimum stratum boundaries can be approximated by taking equally spaced points on the scale

$$Z(X) = \int_{-\infty}^{X} \sqrt{f(t)} dt ,$$

where f is the density function of X.

9. We now turn to some partitioning problems where the objective is not so explicit; the ones we will be discussing frequently go under the name of cluster analysis. When one tries to articulate his objective in doing cluster analysis, one tends to come up often with something like to chop up a supposed multimodal distribution of X values by putting one mode in each stratum. Often underlying this objective is the notion of trying to represent the population as a mixture of intrinsically interesting subpopulations each with its distinctive distribution of X values.

10. We might try to model the situation of the last paragraph in the following approximate way: There are k simply connected sets (islands) in X-space which are at a positive distance from each other and on which the density of X is positive; everywhere else the density is zero. See Fig. 1. The objective is to have each stratum of our chosen partition contain one island. It is assumed, of course, that a distance has been defined on X space. If the population is finite this model may not seem too helpful, but in such cases it is sometimes possible to think of the population as having been sampled from a superpopulation with a continuous X distribution of the above type. This model is surely an oversimplification of most real situations; but if our partitioning procedures can not do well in such simplified problems, they are not likely to do well in the less well resolved practical situations. Such a model might, therefore, be useful in a comparative study of partitioning procedures.

11. If we had this kind of population in mind and a random sample from it, what sort of procedures are reasonable for dividing the sample into clusters (recall the definitions of Par. 3). If the sample is correctly clustered, then we might expect the individuals within a cluster to have X values close to each other, whereas distances between clusters should be fairly large. Considerations of this kind have led various people to fix on within-or between-cluster distance criteria as a means of evaluating any given partition. We will call such criteria C criteria, generically, and give some examples



Fig. 1. The shaded region is where the density of X-values is positive; it is an example of the model mentioned in Par. 10. below. The corresponding procedure then tries to find a partition which optimizes (maximize or minimize) the criterion C - as a substitute for the less explicit objectives of Par. 9 and 10.

12. One example of a C criterion is the total within strata dispersion, the quantity V of Par. 6. A similar more widely used C criterion is W, the determinant of the pooled within strata covariance matrix, for X variables defined on a Euclidean space. Specifically, if Sh is the covariance matrix conditional on X being in stratum h, and W_h is the weight of stratum h, then $W = det(\Sigma W_h S_h)$. We would then try to find a minimum W or a minimum V partition. Minimizing W or V also has a tendency to maximize the average distance between strata as measured by the distance between the stratum means. The V and W and similar criteria have been explored and used in this way by Friedman and Rubin [5] among others. For example, the criterion trace $(\Sigma W_h S_h)$ is sometimes suggested; it is not often recognized that this is equivalent to MacQueen's V when Euclidean distance is used to compute V.

13. Whether minimizing V or W gives us the kind of partition we really want is at least sometimes open to question. Unfortunately, it is all too easy to construct examples where such minimizing does not divide up multimodal distributions in the desired way. In particular, let X be real valued and have zero density everywhere except on the intervals (0, .3) and (.4), 1) where it is constant. See Fig. 2. If we were smart enough to have chosen k = 2, then the articulation of the objective of Par. 9 would be to partition the line into two halflines with the boundary point between 0.3 and 0.4. However, this partition actually gives a larger (worse) value for the V or W criterion than the partition with boundary point at 0.5, say. One might well regard this as peculiar or unsatisfactory.

14. Another example is illustrated in Figure 3. It shows a finite population of 10 individuals taking on X values in the unit square. Two selected partitions for k=2 are also shown. Although we might feel happier with the first of the two partitions, it is the second one which gives a better (smaller) value for V or W. This example hints at the care one must exercise in choosing a C-criterion, if this is the desired approach. Part of the problem here is that both V and W may not work well when the islands differ appreciably in size or shape.

15. The actual calculation of an optimum C partition, given the distribution of X, is in general a very difficult mathematical task. If the population is finite, or if it is infinite but there are only a finite number of different X-values in the population, then the number of possible partitions of the X values into k sets is also finite. In principle, it is then possible to compute C for all possible partitions and thus find the optimum C partitions. However, if there are m different X-values then there are $k^{m}/k!$ partitions to enumerate for each fixed k - an unreasonable task if m is any size at all.

16. Instead of enumerating to get an optimum C partition in this finite case, Fortier and Solomon [6] have examined the possibility of using the partition which optimizes C among a random sample of partitions. Suppose there are about 10^{10} possible partitions and we take a sample of 10^{2} of them. The probability that the sample will contain the overall optimum C partition (assume it is unique) is still minute, viz. 10^{-7} ; but the probability that the best partitions is very high, viz. $1-(.99)^{1000} = .99995$.

17. Despite the fact that the sampling method is likely to give a partition with a very good C, it seems from the Fortier and Solomon study that very good is not good enough. In the first place partitions with even better C can be obtained by various simple ad hoc procedures to be discussed in Par. 32. In the second place it seems that the distribution of C values has a very long thin tail, or that even slightly suboptimal C values may correspond to unappealing partitions.

18. Although Fortier and Solomon used a particular C criterion not discussed here, we feel that these conclusions about the sampling of partitions have more general validity. However, if we know that an optimum C partition must have certain properties, then we can restrict our sampling to those partitions possessing such properties. For example, the clusters of an optimum partition may have disjoint convex hulls. There will be relatively very few such partitions and our sampling effort is likely to be much more successful when so restricted. Of course, it is probably hard to come by such properties of optimal C partitions; but we can make them up anyway so as to be consistent with the real objective (See Par. 9) - which, after all, is not the optimization of C.

19. As a simple example consider the finite population of Figure 3 again. With k=2 there are 512 possible unrestricted partitions. With the convexity restriction of the preceding paragraph there remain only m(m-1) partitions to consider, or in this example only 90.

20. Other approaches to find optimum C partitions have been explored by Friedman and Rubin [5] and a good list of references is contained there. Common to these approaches is that one starts with a given partition, modifies it according to a fixed procedure, and ends up with new partition with a better C value. The modification procedure is iterated until it reaches a stable partition, i.e., one which is unaffected by the modification. It is then said to have converged and the resulting stable partition becomes a candidate. For example, for a given partition into k clusters of a finite population in Euclidean space, the cluster means and the pooled within clusters covariance S are computed. The Mahalanobis distance of each individual to each of the cluster means is then also computed. The modification procedure consists of reassigning each individual to the cluster to whose mean it is closest. This guarantees that the value of the criterion function W of Par. 12, the determinant of the pooled within clusters covariance matrix, will be reduced. However, there are many stable partitions for any given configuration of X values, some good and some not so good. Any one of them can be reached by a suitable input partition. Of course, the minimum W partition is a stable partition, also.

21. Although the modification procedure guarantees a sequence of partitions with decreasing W, the final stable partition of the sequence may be nowhere near a minimum W partition. Of course, the modification procedure could be used with several different input partitions, then we can choose the best of the output stable partitions. Indeed, such modifications routines can be combined with Fortier-Solomon sampling of input partitions. Figure 3 shows the final result of using the modification procedure of the last paragraph, starting from two different initial partitions. A recent application is reported by Demirmen [7] who used the technique to "improve" rock classifications. Demirmen also wrote an efficient IBM 360 computer program to implement the technique.

22. In summary, we see that replacing a general clustering objective of the type in Par. 9



Fig. 2. The X distribution of Par. 13 and two different partitions of X. If we were trying to minimize W (or V), the second partition would be judged better. by a specific objective which seeks to optimize the value of V or W may or may not be a reasonable thing to do. The state of the art is such that we don't always know when it is reasonable. Furthermore, we don"t yet have proven efficient methods for finding optimum V or W partitions, even when we do think it is reasonable.

23. As an alternative to V or W, we might look for a C criterion which is especially tailored to do well for models of the type in Par. 10. Recall that the islands are assumed to be a positive distance apart. It seems more reasonable then to measure distance between strata in terms of the minimum distance between any pair of individuals from different strata - rather than in terms of the distance between stratum means which is implied in the use of V or W.

24. Specifically, let the minimum distance between stratum h and stratum i be denoted by R_{hi} . For any subgroup of m > 1 individuals there will be a subdivision into two strata which is optimal in the sense that R is maximized over all partitions of the mX values into two sets. This maximum value of R could be taken as a measure of the within subgroup distance for that subgroup of m individuals. For example, in Figure 4 the within distance for cluster 1 is 21mm, for cluster 2 it is 26 mm, and the R distance between the two clusters is 22mm. (Using Euclidean distance.) For a chosen partition of the whole population we define ${\rm R}_{\rm h}$ to be the within stratum distance for stratum h. If a stratum has only one point take $R_h = 0$. As an alternative to W type criteria, we suggest using something like $Z = \min R_{hi} - \max R_{h}$. The objective now is to find the maximum Z partition of the population.

25. Maximizing criteria like Z is likely to be consistent with the clustering objectives of paragraph 9 and 10; especially for fairly large samples. This is suggested by the following consideration: Suppose our partition did correspond to the "islands" of the model, then splitting an island would greatly decrease minR_{hi} whereas merging two islands would greatly increase max R_h . The merit of a Z criterion should not be much affected by diversity in the sizes and shapes of the islands. The maximum Z partition for the population of Figure 3 is the first one shown there (A).

26. Z also has the satisfying property that, for any partition, it is computible from the pairwise distances between the points of the sample. In fact it has the property that for any specified number k of strata, there is a unique partition for which the Z value is positive. The last statement assumes there are no ties among the pairwise distances and implies that this unique partition is the optimum one. The proof is a little involved and rests on the following observations: If any sample is divided into two clusters so that the R distance between them is maximized then this is an optimum Z partition for k=2, and Z is positive; the optimum Z partition for k' > k strata is always a refinement of the optimum Z partition

for k strata; if we have subsets of two strata then the distance between the subsets is not less than the distance between the strata; finally if a single stratum is divided into several subsets, then there will be a pair of subsets with distance betweeen them not less than the within stratum R-distance as defined in Par. 24. The construction of a proof is now left to the reader.

27. The refinement property of maximum Z partitions, noted in the last paragraph, suggests a simple algorithm for finding them. For, suppose we have the best partition for k+l strata; if we merge the two nearest strata (in terms of R distance), it follows that we will now have the best partition for k strata. In particular, we can start by assuming that each of m distinct X-values of the sample is a cluster unto itself. Then, by repeating the process of merging nearest clusters m-k times, we will end up with the maximum Z partition for k clusters. So we see that this particular criterion can be optimized in an extraordinarily simple manner. It also follows that the quantity $\max R_h$ for optimum k clusters is equal to the quantity $\min R_{h_1}$ for optimum k+l clusters, hence the sequence of maximum Z values for each k is itself easily computed. This would be useful if our choice for the number of strata was based on comparing Z values for different k.

28. There is an even more direct and general approach to the partitioning problem when the population is believed to resemble the model of Par. 10. Suppose we could estimate the density of X somehow. Then the estimated "islands' could be those connected sets in X-space on which the estimated density exceeded a certain threshhold value; the chosen partition would then put one island in each stratum. Note that the number of strata is not fixed in advance. A naive method of thresholding the density at a point X_0 in X space is to count the number of sample points in a fixed neighborhood around X_{o} . This count can then serve as the threshholding device. This procedure has been applied to the finite population of Figure 3 with circular neighborhoods of radius 11 min and a threshhold of 1 point. The resulting partition is shown in Figure 5.

29. The density threshold approach and the maximum Z approach are both clearly tailored to the partitioning of populations which are more or less of the type in Par. 10. We have tried to argue that clustering of populations is often attempted only because it is believed that such a description roughly obtains. But it is clear that both of the last named approaches can fall down in many situations.

30. For one, if the sample is small both procedures become rather sensitive to the actual configuration of sample X values. The maximum Z approach is, in general, sensitive to outlying X values; this could possibly be remedied by weighting the R distances by the number of points involved in the pair of clusters, or a similar device. A problem with the model occurs when there are relatively high density "bridges" between pairs of islands. See Figure 6. We may still want such a pair of islands to show up in different strata, but the tendency of our last suggested procedures would be to lump such connected islands into a single stratum, even in large samples.

31. We now return to re-examine the algorithm used to find maximum Z partitions, as described in Par. 27. This algorithm belongs to an interesting class of stepwise partitioning algorithms very much in the spirit of the one used by King [8] and by Sokol and Sneath [10]. We will call algorithms of this type stratum-merging procedures.



32. Suppose we already have a partition of the population into k strata (based on X values), then we can obtain a partition into k-l strata by merging two of the existing strata into a single stratum. If we specify a function D which measures distance between any pair of strata, then the natural thing to do would be to merge the two strata which are nearest each other. Such a merging procedure can be iterated to obtain a partition with any number of strata less than k. In particular, for populations taking on only a finite number m of different X values, we can get started by considering each X-value to be a stratum by itself. If we iterate the merging routine with a specified D function





Fig. 3. Four partitions of a population of ten individuals into two clusters. The partitions A^* and B^* are stable by the method of Par. 20 and can be reached from initial partitions A and B, respectively, in one iteration. The W values are (in increasing order) .0047 - B*, .0056 - A*,



.0158 - A, .0388 - B. The V values are (in increasing order) .2515 - B*, .3341 - A*, .4112 - A, .4662 - B. The only partition with a positive Z value is partition A, with Z = .04. See Par. 24. The scale is 1 inch = .24 units.

exactly m-k times, we will then end up with exactly k strata - for any specified k.

33. It is clear that when D is taken to be the R-distance of Par. 24, then this exactly describes the procedure used to find maximum Z partitions. But for any arbitrary D measure it may not be evident which, if any, criterion is being maximized by such a stepwise procedure (in the sense of Par. 11). Indeed, the V or W criterion discussed in Par. 12 are unlikely to be optimized in this way for any choice of the cluster distance function D: We might try anyways by choosing D to be some measure of distance between the stratum means - but the stepwise procedure will tend to maximize the minimum distance between any pair of means, whereas V or W tends to maximize the average distance between means. Notwithstanding these divergent tendencies, the stepwise procedure may still give as good a value of V or W as that obtained by the more cumbersome methods of Par. 16 or Par. 20. Also, such stratum merging procedures are always relatively quick and simple to execute, and have the virtue of automatically generating a whole sequence of suggested partitions, one for each possible value of k, the number of strata.

54. Solomon [9] has obtained satisfactory partitions using cluster-merging on at least three sets of data, in each case basically using distances between means to measure distances between strata. King himself was partitioning real-valued variables $(Y_1, Y_2, \dots say)$, and for a distance measure between clusters A and B he used

 $D^{2} = (\sum_{j \in B} \sum_{i \in A} S_{ij})^{2} / \sum_{j \in A} \sum_{i \in A} \sum_{j \in B} \sum_{i \in B} S_{ij}$

where S_i is the covariance between Y_i and Y_i.^{ij}This can be recognized as the square^j product-moment correlation between the average of variables in A and the average of variables in B. So, essentially, all he needs to be able to do is to estimate the covariance matrix.

35. Throughout the entire discussion thus far we have had frequent occasion to refer to the computation of distances - for the V criterion of Par. 6, for the definition of neighborhoods in the density estimation of Par. 28, and in many other places. Very little was said explicitly about how distances should be defined and it seems that little can be said. This is an especially thorny problem when the X variable is of high dimension and it becomes very easy to make very poor choices. We would probably want certain simple transformations of the X-variable to preserve the order relation among distances, and we might choose our distance function with this in mind. Monotonic transformations of the distance function will not affect minimum Z partitions; however, this is not true for the other procedures discussed here. 36. If the covariance matrix of a population in Euclidean space is S, then the distance between points X_0 and X_1 in this space can be taken as (X_0-X_1) 'S⁻¹ (X_0-X_1) . This is what is called Mahalanobis distance with respect to the overall covariance; it is totally invariant under linear transformations of X. This distance has been used with satisfaction by Solomon [9] for example, but one can construct examples where it is obviously inappropriate. In general, we would have a much better idea of an appropriate distance measure if we already had the clusters, but this just begs the question. The cluster improvement procedures discussed in Par. 20 implicitly modify an arbitrary distance function to make it more appropriate.

37. The sensitivity of a partitioning procedure to the choice of a distance function will tend to decrease as the sample gets large and as the dimensionality of X goes down. While we may not be able to control our sample size, it might be useful to attempt to reduce dimensionality before fixing on a distance function. Broadly speaking we will need some luck in this anyways.

38. We have made no attempt so far to cast any of the foregoing partitioning problems in a decision theoretic framework. If we allow ourselves to postulate the existence of an "ideal" partition, then our loss could be measured by the extent to which the chosen partition differs from an ideal one. The measure of discrepancy should depend, of course, on the nature and objectives of the particular problem in hand. In general, it will not be possible to calculate the loss without knowing an ideal partition, and if we knew of such a partition we would naturally use it and incur a minimum loss. Losses are being introduced, therefore, not because they are a calculable quantity in real problems, but for two other reasons: first, to put the formulation of the objectives of partitioning procedures on a firmer conceptual basis; second, to provide a means of evaluating recommended partitioning procedures on artificial examples (with specified ideal partitions) as a clue to how they would perform on different kinds of real populations. Finally, in certain special cases involving random sampling from populations, it may actually be possible to calculate the expected loss of a given procedure, even when an ideal partition is not known.

39. As a first example consider the problem of partitioning a continuous X distribution into sets of equal probability content, the problem of Par. 5. If P_1, P_2, \ldots, P_k are the actual (but unknown) probabilities of the k chosen strata, then our loss might be measured by $L_1 = \max P_h - \min P_h$. If we use the Fraser [1] method of partitioning into strata, it is actually possible to compute the expected loss for this method, since the P_h 's so generated will have a completely specified Dirichlet joint distribution.

40. If the k-means problem of Par. 6 has the minimum mean-square-error objective of Par. 7, then it might be reasonable to measure the loss by something like $I_2 = \log[V_0/\min V] - \text{where } V_0$

is the computable V value for the chosen partition and min V is generally not computable except in artificial examples. However, if we are trying to find minimum V partitions as an approximation to the general <u>clustering</u> objectives of Par. 9, it is probably inappropriate to measure the loss by L_2 or anything like it in which only V values are involved. A similar remark applies to procedures that look for partitions with small W values. See Par. 12.

41. Here is an approach to the specification of an appropriate loss function when general clustering objectives are involved. Let us call individuals who belong to the same stratum of an ideal partition of X - friends; individuals belonging to different strata of an ideal partition will be called enemies. If, in the chosen partition, an individual finds that he has f friends outside his stratum and e enemies within his stratum, his loss is e+f. The total loss, L_z , for the chosen partition may then be taken proportional to the sum of the e+f losses over all individuals in the population. As an example, if the partition A of Figure 3 is considered to be "ideal", then partition A* has loss $I_2/100 = .18$, B has loss .48, and B* has loss .32. For infinite populations the expected value of an L₅ type loss could be computed for a random sample of standard size.

42. The loss function of the last paragraph has an interesting characterization, whether for finite or infinite populations: Let A_1, A_2 , $\dots A_k$, denote the strata of the ideal partition and let B_1, B_2, \dots, B_k denote the strata of the chosen partition. Note that k and k' are not necessarily the same. Now let P_1 denote the probability content of A_1 , Q_1 the content of B_1 , and P_{11} the content of $A_1 \cap B_1$. Then it turns out that L_2 is proportional to $\Sigma P_1^2 +$ $\Sigma Q_2^2 - 2\Sigma \Sigma P_{11}^2$. The proportionality factor would involve only the population or the sample size. Note that $L_2=0$ if and only if the chosen and ideal partitions coincide. As an example for an infinite population, suppose the ideal partition of the population in Figure 2 (k'=2) has its boundary point somewhere between X = .3 and X = .4; then the suboptimum partition with boundary at X = .5 (k=2) has $L_3 = 16/81$.

43. We conclude by extending the notion of an ideal partition. At the very outset we noted in Par. 1 that our chosen partition of a population must be based on the values of a specified measurement variable X. However, in populations where more than one individual can take on the same X value, it is a non-trivial extension to allow the ideal partition to be arbitrary, i.e., not to depend on X. The implication is that no partition based on X can be perfect in the sense that L₃ can never be zero, though there will still exist a best partition based on X with minimum L₅. The representation for L₃ of the preceding paragraph remains valid in this extended situation.

44. An example is provided if we want to cluster a human population into ethnic backgrounds but our measurement variable X is the surname. Note that this extension plays no role if we are

strictly in the islands situation of Par. 10 and the objective is to isolate the islands. It is when the islands begin to overlap that the extension comes into play. One of the reasons for introducing the extension was to forge a link from the problems we have been discussing to the standard Wald assignment or classification problem. If the strata of the ideal partition have labels 1,2,...,k, and if k is specified, and if we are required to attach labels 1,2,...,k to the strata of our chosen partition - then we have arrived; i.e., it is possible to regard the difference between the Wald problem and the problems of this paper as the difference between having to choose an ordered versus an unordered partition of X space.



Fig.4. This particular partition of the 10 individuals into two clusters has within cluster 1 dispersion $R_1 = 21 \text{ mm}$, within cluster 2 dispersion $R_2 = 26 \text{ mm}$, and between clusters dispersion $R_{12} = 22 \text{ mm}$. This gives Z = -4 mm. See Par. 24.



Fig. 5. The shaded area is where the density of X is estimated to be positive by the method of Par. 28 with circular neighborhoods of radius llmm. The result is a partition into three strata as shown.



Fig. 6. The shaded region is where the density of X values is positive; it is an example of the model mentioned in Par. 30.

References

- Fraser, D.A.S., <u>Nonparametric Methods in</u> <u>Statistics</u>, John Wiley and Sons, Inc., <u>New York</u>, 1958.
- [2] MacQueen, J., "Some methods for classification and analysis of multivariate observations," <u>Fifth Berkeley Symposium</u> (1965) 281-297.

- [3] Cochran, W.G., <u>Sampling Techniques</u> (2nd edition), John Wiley and Sons, Inc., New York, 1963.
- [4] Dalenius, T., and Hodges, J.L., Jr., "Minimum variance stratification," J. of Amer. <u>Statist. Assoc</u>., 54 (1959) 88-101.
- [5] Friedman, H.P., and Rubin, J., "On some invariant criteria for grouping data," <u>J. of Amer. Statist. Assoc</u>., 62 (1967) 1159-1178.
- [6] Fortier, J., and Solomon, H., "Clustering procedures," <u>Multivariate Analysis</u>, Edited by P.R. Krishnaiah, Academic Press, New York (1966).
- [7] Demirmen, F., Unpublished Ph.D. dissertation, Geology Department, Stanford University (1968).
- [8] King, B., "Stepwise clustering procedures," <u>J. of Amer. Statist. Assoc</u>., 62 (1967) 79-85.
- [9] Solomon, H., personal communication.
- [10] Sokol, Robert R., and Sneath, P.H.A., <u>Principles of Numerical Taxonomy</u>, W.H. Freeman and Company, San Francisco, 1963.

MEASUREMENT OF HOUSING QUALITY AND ITS POLICY IMPLICATIONS

ш

Chairman, LAWRENCE N. BLOOMBERG, U. S. Office of Statistical Standards

	Page
Methods of Measuring Housing Quality - ARTHUR F. YOUNG and JOSEPH M. SELOVE, U. S. Bureau of the Census	49
The Uses of Housing Quality Data in Public Policy Formulation -	
FRANK S. KRISTOF, Housing and Development Administration, City of New York	52
Discussion - JOSEPH WAKSBERG, U. S. Bureau of the Census	59
Discussion - E. EVERETT ASHLEY, Health & Welfare Council	60
Discussion - CATHERINE E. MARTINI, National Association of Real Estate Boards	63