

Multivariate Analysis and Human Skeletal Populations

RALPH W. ALEXANDER, JR.

The University of Wisconsin-Milwaukee

Milwaukee, Wisconsin 53201

Department of Anthropology

Abstract

The problems of determining the relationships of archeologically-derived human skeletal populations are discussed. Multivariate analysis is examined in terms of its statistical basis, its value as a tool in studies of human skeletal populations, and its development and history of application in anthropology. The specific technique of multiple-discriminant analysis and its application to problems of microevolutionary change is discussed.

One problem in studies of populations based on archeological sites is delimiting the population's boundaries. This may be done on the basis of grouping skeletal materials according to association with particular assemblages of artifactual materials. If different populations resemble one another in a number of diagnostic characteristics, they may be combined into larger groupings. Temporal boundaries are established with the aid of analysis of differences found in cultural materials at different stratigraphic levels, and from the dating of specific periods in the overall complex. Both temporal and spatial delimitation provide the framework for the establishment of relationships of the population. The remains of a specific culture are assumed to constitute an intrabreeding population which may be characterized by a cluster of characteristics and form a recognizable stabilized entity, whose pattern of variation is known. Only after analyzing the physical characteristics of the available remains and comparing the results with other delimitable populations can particular samples be identified as having a certain likelihood of relationship with a particular archeological population or well-defined tribal grouping. Since one purpose of many studies is the selection, identification, and evaluation of traits having diagnostic phyletic significance for the samples involved, attention must be given to the selection of traits that have historic meaning at a particular level of differentiation. The premise is that these traits, and functional combinations of them, will best reflect the genetic continuity of the populations in accord with other evidence. Once traits have been identified, the populations may be compared to a number of other series. It should be emphasized that combinations of traits rather than individual traits, whether continuous or discontinuous in nature, serve to delimit and identify a group.

Of the characteristics which can be studied in prehistoric bone material, little is known of the specific genetic factors involved. Despite this limitation, in almost every case nothing else remains of the biology of early man, and as much reliable and useful information as possible must be obtained. It is apparent from the nature of the patterns of distribution of the values of the measurements and observations that these traits are polygenic in nature. Although the genetic and environmental components

are not presently quantifiable, this variation is in no way detrimental to the use of these traits for comparative purposes; it is rather the nature of varietal groupings to exhibit a degree of overlapping in their characteristics, and the statistical methods employed in descriptive and inductive presentations in general make use of, assess, and in fact require this variation for their proper application. Indicial traits, demonstrating an absolute relationship of proportion between two measurements, may be employed along with measurements and observations. Change on a varietal level may occur in the form of an overall reduction or increase in gross size, with the form of the cranial proportions remaining constant. This has been demonstrated for Amerind varietal groups (1), and may reveal relationships that are not obvious from measurements alone.

In an analysis that deals with relationships and similarities of human skeletal populations it is appropriate to use a tool such as multivariate analysis, which considers the total physical variation present, and assesses the relative contributions of each variable in distinguishing between groups. A condition which must be met for such an analysis is that the same kinds of data be collected for all individuals of all groups under consideration. Assumptions which must be made are that the various samples have a multivariate-normal distribution, a common dispersion matrix, and a linear correlation between variables. A multivariate analysis is characterized by its consideration of a number of characteristics simultaneously in assessing the relationships or discriminating traits between two-or-more groups or sample series. The important distinction of this statistical approach is that the totality of biological variables are considered in combination, as systems. The interrelationship of all the morphological characteristics of the individual and the relative significance of each variable is considered. The significance for evolutionary change of individual traits can only be assessed in terms of their relationships to other traits. All biological characteristics interact. A bone or a tooth or a skull is a unit, and should be dealt with as such rather than as an assembly of individual traits. Bronowski and Long (4) said that the correct statistical method for the study of discrimination in anthropology must treat the set of variables as a single coherent matrix, a capsule summary of the multiple-discriminant technique. The individual is preserved as a vector of all his characteristics, taken together with all they convey as to size and shape by both absolute magnitude and by covariation. The relative contribution of each variable in discriminating between groups is assessed, and "weights" are assigned accordingly. Therefore, only what is really significant in the particular comparison is utilized, and exactly which variables are important and in what magnitude is brought out by the technique. The technique deals with the possibility of a particular variable, which in itself is not a good discriminator between groups, emerging as highly valuable when taken in combination with other variables.

The amount of calculation involved in this type of analysis has severely limited its use of anthropology until recently; a computer is a necessity. Pearson's "coefficient of racial likeness" or C_2 was one of the earliest statistics to utilize multiple measurements, and was a test of the statistical

significance of difference between populations rather than a true measure of biological distance. This statistic was later revised to be a measure of biological distance, and to handle intercorrelations. Mahalanobis (9) developed a still useful statistic referred to as D^2 , or the generalized distance statistic. The D^2 procedure measures the extent and direction of morphological separation between groups, and makes possible an analysis of relationships between continuous variables, discounting the correlation existing between them. Fisher (7) considered D^2 as the best available statistic for estimating biological similarities and differences between groups. In 1936, the concept of discriminant function analysis was introduced by Fisher, a new method of testing significance for multiple measurements. A procedure was devised for estimating test weights in such a way that a linear combination of the weighted scores (called discriminant function coefficients) would provide a maximum discrimination between groups of individuals. The compounded measurements maximize the variation between groups, relative to the variation within groups. The use of multiple-discriminant analysis in anthropological research is only now becoming common, and is not as yet widespread. With the increasing availability of computer facilities the calculation necessary for its application will become less of a problem. A number of investigators have employed it in research of an anthropological nature (3,8,10). Fortran computer programs for multiple-discriminant analysis are available (5,6). Dixon's BMD 04M program computes means of the variables for each group and the mean difference, variance-covariance matrix, inverse of the variance-covariance matrix, discriminant function coefficients, Mahalanobis' D^2 and associated F statistics, and the discriminant scores for the groups with their means, variances, and standard deviations. Scaled vector values to show the relative contributions of the variables to each function may be obtained from the Fortran V program Multivariate Discriminant Analysis (2). Scaled vector values are discriminant function coefficients computed after standardization of variables, and give an accurate picture of the relative contributions of the particular measurements or indices contributing to discrimination. In applications of the method to hybridization or classificatory studies, a high scaled vector value indicates a major contribution to group separation, or a low relative similarity between groups for the particular variable. A low scaled vector value indicates that the variable does not contribute greatly to group separation, and is a biological feature which has not changed between, for example, a hypothetical ancestral population and a hybridized descendant group. A "morphological score" is computed for each individual hybrid skull on the basis of the discriminant function coefficients obtained from the comparison with potential ancestral or contributor series. The biological position of each individual is assessed, in addition to the relationships between the series considered as units. The relative contributions of each series, with respect to each variable used, is obtained. In classificatory studies, the position of each individual may be assessed and the specific factors accounting for this position identified. An initial set of variables may be reduced to a smaller number of variables for further studies, and the overall direction of morphological trends in the samples obtained.

Since in a multivariate analysis it is possible to treat a skull as a unit, rather than as a series of single traits, and to compare these traits in population terms, inter-related measurements likely to bring out aspects of size and shape characterizing populations should perhaps be the goal in mind. Characteristics which would be useful in characterizing populations may be identified which are not at all those which have been traditionally used in univariate skeletal comparisons. In this regard, there is much need for continued experimentation.

Literature Cited

1. ALEXANDER, R. W. JR., and G. K. NEMUANN. 1969. On the origin of the Tutelo-an Eastern Siouan tribe. *Proc. Indiana Acad. Sci.* 78:88-92.
2. ALEXANDER, R. W. JR. 1971. Multi-t-test and multiple discriminant analysis, p. 69-70. *In* S. A. WEST (ed.) *Directory of computer programs for anthropologists*. W. Mich. Univ., Kalamazoo, Mich. 71 p.
3. ————. 1971. The hybrid origin of the Arikara Indian of the Great Plains. Unpublished Ph.D. Dissertation. Indiana University. 233 p.
4. BRONOWSKI, J., and W. M. LONG. 1952. Statistics of discrimination in anthropology. *Amer. J. Phys. Anthropol.* 10:385-394.
5. COOLEY, W. W., and P. R. LOHNES. 1962. *Multivariate procedures for the behavioral sciences*. John Wiley & Sons, New York, N. Y. 211 p.
6. DIXON, W. J. 1964. *Biomedical computer programs*. Health Sciences Computing Facility, Univ. Cal. Press, Los Angeles, Cal. 620 p.
7. FISHER, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annu. Eugen.* 7:179-188.
8. HOWELLS, W. W., and J. M. CRICHTON. 1966. *Craniometry and multivariate analysis*. Papers, Peabody Mus. Cambridge, Mass. 67 p.
9. MAHALANOBIS, P. C. 1930. On tests and measures of group divergence, theoretical formulae. *J. Asian Soc. Bengal.* 26:541-588.
10. RIGHTMIRE, G. P. 1970. Iron age skulls from southern Africa re-assessed by multiple discriminant analysis. *Amer. J. Phys. Anthropol.* 33:147-167.

