H16 Morphoscopic Traits and the Statistical Determination of Ancestry II

Joseph T. Hefner, BA, MA, Department of Anthropology, University of Florida, Gainesville, FL 32611; and Stephen D. Ousley, PhD, Department of Anthropology, Smithsonian Institution, Washington, DC 20560*

The primary goal of the presentation is to provide the audience with a new statistical method for utilizing nonmetric traits in the determination of ancestry in unidentified human crania.

This presentation will impact the forensic community and/or humanity by enhancing both the process of identifying the ancestry of human remains and the courtroom presentation by attaching probabilities to the analysis.

Ancestry determination is generally regarded as the most difficult aspect of the biological profile, primarily because of the traditionally subjective, experienced-based emphasis that has been taught in the past. In addition to subjectivity, another weakness of the classic nonmetric approach is that it does not provide any posterior probabilities, which indicate how likely the individual comes from one group as opposed to the other groups. Nonmetric (categorical) trait analysis is often devoid of all but the most simplified statistical analysis (e.g., frequency distributions and chi-square tests). Nonetheless, Ousley and Hefner (2005) demonstrated that morphoscopic traits could be analyzed statistically with classification accuracy as high as 90% using kernel probability densities. In addition, several other methods (including logistic regression and linear discriminant functions), confirm the value of using various statistical methods in nonmetric trait analysis. The statistical techniques that utilize metric data provide posterior probabilities, usually because the data are normally distributed, but nonmetric data are not normally distributed, and quantifying nonmetric traits is not practical and may not provide additional significant information. A new multivariate statistical method recently introduced for use with categorical data is presented that not only addresses these concerns, but also addresses the *Daubert* issue using posterior probabilities, graphs, and cross-validated results.

In the present contribution, a total of thirteen separate distance measures were applied to 12 nonmetric variables recorded for American Whites (n = 89) and American Blacks (n = 105). The data were analyzed using a constrained ordination procedure (CAP) proposed by Anderson and Willis (2003) which relies on a canonical discriminant analysis (CDA) on the principal coordinates (PCO) of the dataset. The analysis can be performed using *any* calculated inter-individual distance matrix, and the distances need not follow any particular distribution or even be Euclidean. Numerous distance calculations can be used, so distance calculations can be fine tuned (i.e., each method highlights various aspects of the multivariate data), and judged by their classification accuracy. A "leave-one-out" resampling post hoc test of the CDA is then used to (1) calculate classification accuracy and (2) test the significance of additional PCO axes (m). This method is more applicable to ancestry determination than frequency distributions, as it accounts for correlation among variables, provides classification error rates, and allows for classification of a new observation. In these regards it is more consistent with traditional multivariate discriminant function analysis (DFA). An additional benefit of this method is that while conventional DFA can be applied to macromorphoscopic data (Ousley and Hefner 2005), the CAP approach is more appropriate for this type of data.

Several CAP analyses offered interesting results, demonstrating the flexibility of the technique. Multiple distance measures (e.g., chi-square, Euclidian), were selected for the analysis in order to determine the weaknesses and strengths of each one of them. Among these, chi-square distances appear to be the most appropriate measure for categorical, macromorphoscopic data. For example, the chi-square distance measure with $m = 10$ produced the lowest mis-classification error, and the first three PCOs alone explained over half of the total observed variation. In fact, a CAP analysis using the five best variables (INA, NBS, ANS, IOBW) and the chi-square distance measure with $m = 3$ has cross-validated accuracies of 86%.

These results indicate that the analysis of nonmetric traits through robust statistical models, particularly constrained ordination procedures like CAP, greatly enhances the process of identifying the ancestry of human remains. This method also addresses issues related to the *Daubert* decision by calculating classification accuracy and posterior probabilities of the classification, and enhances courtroom presentation through the obtained bivariate plot, an easily understood graphical representation.

Nonmetric Traits, Canonical Discriminant Analysis, Ancestry Determination

* *Presenting Author*