H17 Morphoscopic Traits and the Statistical Determination of Ancestry

Stephen D. Ousley, MA, PhD, Smithsonian Institution, Department of Anthropology, PO Box 7012, NMNH MRC 138, Washington, DC 200137012; and Joseph T. Hefner, MA, Department of Anthropology, CA Pound Human Identification Laboratory, University of Florida, Gainesville, FL 32605*

This presentation will impact the forensic community and/or humanity by showing how they can utilize non-metric traits in a statistical approach and immediately start using these techniques to assess ancestry. Their opinions will be more reliable because they can state a probability of group membership and will know the samples and sample sizes their results and conclusions are based on.

The primary goal of the presentation is to provide statistical guidelines and results for using nonmetric traits in the determination of ancestry. It will be made clear that several statistical methods provide enhanced utilization of non-metric traits in the classification of unidentified human crania.

Hefner [1, 2] has shown empirically that the "typical" nonmetric traits of African-, Asian-, and European-derived groups are not found at frequencies suggested by forensic anthropologists such as Bass, Byers, Gill, and Rhine. These traits were highly valued by Earnest Hooton, and come predominantly from his "Harvard List" of traits. Hooton prepared illustrations of the "Harvard List" but apparently, neither he nor any of his students published group trait frequencies. While the extreme trait expressions are not as frequent as suggested and thus, not as reliable for estimating ancestry, the traits were nonetheless demonstrated to have value in discriminating between American Blacks and Whites. Hefner ***et al.*** [3] introduced a Binary Optimized Aggregate Score (BOAS) statistic that classified a sample of 165 American Blacks and 160 American Whites into ancestry 80% correctly. BOAS weighed each variable equally: it was simply the sum of binary trait scores (0 or 1) heuristically derived from the ordinal trait scores.

Discriminant functions (DF) optimize the weights of each variable to maximize group separation. However, Linear Discriminant Functions (LDFs) usually require multivariate normality and approximately the same level of variation in order to produce reliable results. LDFs are aided by the central limit theorum, which tends to increase the normality of data as more and more individuals are added. Some statisticians [4] have suggested that dichotomous or ordinal variables could be used in a LDF as if they were metric with certain caveats. One criterion of a valid two-way DF is that the DF scores of each group should be approximately normally distributed. This way, the sectioning point is guaranteed to be the optimum and should be valid for future samples, and posterior and typicality probabilities will represent the correct values. However, sectioning points can still be derived whether or not the DF scores are normally distributed. For example, if a measurement or function completely separated groups, it could be used no matter what the distribution was. One caveat is that the posterior and typicality probabilities may be misleading.

Logistic regression (LR) has been used to discriminate among groups and does not require assumptions of group variation and multivariate normality. LR can incorporate categorical data as well as continuous data, and can utilize both in analyses. LR calculates a probability of group membership directly, rather than first calculating a composite score, then comparing group scores, and only then calculating a probability, as in LDFs. Similarly, two non-parametric DFs, Kernel DFs (KDFs) and Nearest Neighbor DFs (NNDFs), calculate probabilities directly, independent of group distributions.

For this study, the same ordinal data from Whites and Blacks were analyzed using LDF and a curious pattern emerged. The DF scores and probabilities for each group were skewed but in opposite directions in nearly every analysis, including LDF, KDF, NNDF, and LR of the original variables, as well as LDF of the principal components of the conventional and polychoric correlation matrices of scores. The principal component loadings of the polychoric and Pearson correlation matrices of the traits were remarkably similar (rho = 0.92). Transformations of scores did not appreciably affect the distributions. Cross-validated accuracies up to 90% were achieved using as few as three traits.

LR was the best method overall, in three-group analyses as well, but LR does not produce typicality probabilities. To substitute for typicality probabilities, nonparametric methods such as ranked probabilities and ranked interindividual similarity measures can be employed.

Analyses using KDFs and NNDFs also produced very good results, with crossvalidated accuracies approximately 88%. These results were better than using a basic Bayesian approach, which assumes independence of traits and weighs each trait equally. Problems with a simple Bayesian approach were illustrated by using a trait such as postbregmatic depression, which was present in 44% of the Blacks and 17% of the Whites. Postbregmatic depression was attenuated in LDF and LR through much lower function coefficients.

Our results show that the nonmetric traits examined can be analyzed in ordinal or binary scales (minimizing interobserver error) using DFs. LR produced slightly better results and provides a more robust

 ** Presenting Author*

and reliable prediction of ancestry using nonmetric traits. The results also empirically confirm the value of using nonmetric traits in a statistical framework, as Hooton had hoped. *Daubert* concerns can be addressed through using appropriate and large samples, recorded group frequencies and probabilities, graphs, and cross-validated results, rather than relying solely on the experience and intuition of the observer.

References:

1. Hefner JT. An Assessment of Craniofacial Nonmetric Traits Currently Used in the Forensic Determination of Ancestry. Paper presented at the 54th Annual Meeting of the American Academy of Forensic Sciences, Atlanta, Georgia, 2002.
2. Hefner JT. The Utility of Nonmetric Cranial Traits in Ancestry Determination - Part II. Paper presented at the 55th Annual Meeting of the American Academy of Forensic Sciences, Chicago, IL, 2003.
3. Hefner JT, Ousley SD, and Warren MW. An Historical Perspective of Nonmetric Traits: Hooton and the Harvard List" Paper presented at the 56th Annual meeting of the American Academy of Forensic Sciences, Dallas, TX, 2004.
4. Krzanowski WJ. *Principles of Multivariate Analysis*. Oxford University Press, Oxford, 2000.

Non-Metric Traits, Logistic Regression, Determination of Ancestry

