RACIAL CLASSIFICATION BASED ON NON-METRIC SKELETAL TRAITS

by

KEVIN B. COOPRIDER

B.A., Kansas State University, 1975

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1978

Approved by:

Richard M. Robinson
Major Professor

ACKNOWLEDGMENTS

I would like to express my gratitude to several individuals for their aid in the preparation of this report. Dr. S. K. Perng and Dr. George Milliken have provided comments and recommendations in their capacities as committee members. Stephen McGuire has given valuable programming assistance. Mrs. Patricia Stewart has typed the final copy of this report.

The Department of Statistics of Kansas State University has provided the computer time necessary for these empirical tests.

A special expression of thanks I extend to Dr. Michael Finnegan who not only served on the report committee and critiqued the anthropological portions of the text, but also graciously permitted me the use of his non-metric data and library.

Finally, to Dr. R. M. Rubison, I wish to express my deepest and most sincere gratitude. His guidance, inspiration, and knowledge have been invaluable and have made this study personally, as well as academically, rewarding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

As reflected in his classification schemes, such as the Linnaean taxonomy, man has sought to bring order to his environment as a means of understanding it. So even has he attempted to classify the variations of his own form in terms of intelligence, lineage, and population affinity. It is this latter variety of identification that is to be examined, that is, how an individual is associated with a specific group, and the criteria by which this association is determined.

Human identification is generally made by means of a cultural label (such as a name) or in terms of physical characteristics including sex, race, and age. Forensic cases and archaeological recoveries, situations where the question of identification is generally raised, usually involve individuals of unknown surname. In fact, all cultural trappings such as clothing, adornment, and other personal effects may be absent. In these cases, classification of a body must be based solely on the cadaver remnants. An accurate description of the individual might result if some soft tissue has survived. Sex and race in particular might possibly be determined. On the other hand, skeletal material may well be the only remnant available for analysis. This is the case in all archaeological problems (save the rare instances of mummification or natural preservation) and in many forensic situations. No infallible methods of identification for the factors of race, sex, or age yet exist for osteological material; thus, means of identification must be employed which minimize the possibility of classification error.

This study proposes several methods for classifying an unknown individual to a specific group with similar characteristics based solely on osteological material. Empirical tests are made using museum specimens of documented origin to determine the reliability of these methods for the factor of racial affinity.

It is important to recognize the distinction between this topic and others examined recently. The question is not one of numerical taxonomy, as no new categories are to be created. An individual must be assigned to one of a finite number of defined populations. Neither is this a procedure which attempts to associate individuals as a group with a defined population; each cadaver is classified individually. The techniques of numerical taxonomy as defined by Jardine and Sibson (1971) and Sokal and Sneath (1963) have, therefore, not been employed. Work which has appeared in recent journals (Berry and Berry 1967, Sjøvold 1977, Finnegan 1978, and Finnegan and Cooprider 1978) concerning the relative "genetic" distance between populations is related to this classification problem only with respect to the assumptions concerning the basic data. Finally, although the term "identification" brings to mind an attempt to find a specific John Doe, the analysis here is only a portion of that exact personal identification. Dental records, fractured bones, and osteological pathologies, so crucial in mass disaster cases, will not be considered in this study. Stated concisely, the problem is to classify an individual to a specific racial group on the basis of his skeletal remains.

POPULATION DIVERSITY AND NON-METRIC TRAITS

An assumption upon which the foundation of this study is built is that the population discriminating variables are genetic in origin, rather than environmental. An inbreeding population should, therefore, exhibit similarities with respect to these variables when compared to other populations. However, this factor of inbreeding does not always hold. It has been estimated that the accumulated American White gene flow into the American Black population is between 10-30% (Glass and Li 1953, and Glass 1955) based on serological studies. Roberts (1955) further suggested that American Indian gene admixture is included in this influx estimate. This drifting from pure inbreeding leaves the American Black population quite different from its African ancestors and the present Negro occupants of the African continent. It is crucial to note this aspect of population composition; cultural and social association may be as important as a person's genetic background. A person who considers himself a member of a particular racial stock and who is generally considered by others to be a member of that same stock, should be classified to that group by the following classification methods. These techniques are not meant to identify any genetic scrambling, but are rather intended to associate an individual to his correct racial (cultural) population.

For this study a population is defined as a set of humans which is considered homogeneous with respect to one or a number of genetic (physical) and/or cultural characteristics. Populations may range from broad groupings such as Negroid, Caucasoid, and Mongoloid, to the very specific archaeological populations effectively determined by geographic,

temporal, or cultural differences. In every case, inbreeding is assumed to at least take precedence over interbreeding; it may exclude it.

History of Quantified Population Diversity

The first individuals to note the fine differences in skeletal morphology were actually setting the stage for more sophisticated techniques of discrimination. Brothwell (1968) recognized studies of cranial variation dating as early as 1839. The anthropological and anatomical literature abounds with studies of fine morphological variation, but those deserving particular notice are by Le Double (1903, 1906), Todd and Tracy (1930), Sullivan (1922), Wood-Jones (1931, 1934), and Russell (1900). Others concentrated on skeletal variations which could be measured quantitatively. This latter group of scholars actually provided the first systematic means of population discrimination. Included in this vanguard were Mahalanobis (1936), Penrose (1954), and Pearson (1926). Corruccini (1973) compared these various distance statistics. With the formulation of the linear discriminant function (Fisher 1936), classification of individual cadavers became possible. The variables involved in this technique consisted of bone dimensions and indices. The statistics required for its implementation were estimated from samples of known racial origin. Examples by Giles and Elliot (1960, 1962) yielded empirical misclassification probabilities of 6-20% for White, Black, and Indian samples. In each case, separate functions were computed for males and females, thus requiring a skeleton to be correctly sexed before applying the discriminant function and establishing a conditional aspect to the classification procedure. In support of the technique, however, a racially mixed sample of 572

individuals (of known sex) not involved in the parameter estimation were used to test the derived discriminators. Ninety-nine misclassifications resulted, an error rate of 17%. The prior probabilities of an individual belonging to any one population were assumed equal; therefore, no population had any beginning advantage over another.

Bone is not the only source of morphological variation in the human skeleton. The teeth are the last portion of the body to yield to destructive elements; they are often the sole remnants in archaeological excavations. This fact alone indicates the importance of dental studies. Berry (1976) discussed the variations in dentition in regard to temporal populations and sex dimorphism. Her study focused on the qualitative characteristics of human teeth. Quantitative or metric analysis was carried out by Dahlberg (1963) to demonstrate racial diversity. Neither study attempted to classify individuals to populations. However, both demonstrated the discrimination potential of odontology. Moreover, it is the general consensus that teeth are excellent determinants of age for children and also may provide positive personal identification when dental records or roentgenograms are available.

Even roentgenograms have provided a means of recognizing population differences. Schulter and Finnegan (1977) used Mahalanobis' D^2 statistic to demonstrate this variation in three broad racial categories. Indices, angles, and linear measurements determined from the roentgenograms constituted their variables of discrimination. No individual classification was attempted, but again, the racial diversity suggested such a discrimination might well be feasible using the linear discriminant function.

Serology is yet another area which has shown promise in separating populations. The aforementioned articles by Glass and Li (1953), Glass (1955), and Roberts (1955) are only a small sample of the investigations of genetic variation through blood groups. Cavalli-Sforza and Bodmer (1971) explained the modes of inheritance and the resulting population diversity in detail. This field was once proposed for cases of discrimination involving the skeleton, but Krogman (1962) suggested the results would be quite inaccurate. He also reported varying opinions on the discrimination value of bone density.

Non-Metric Traits

Of all the means of quantifying population variation listed, only metric analysis has progressed statistically to a level of individual cadaver classification. Two less sophisticated classification procedures have been attempted employing morphological skeletal traits and simplified metric traits. These morphological traits are also known as epigenetic, quasi-continuous, discontinuous, or most commonly, non-metric traits. While the traits are not truly discontinuous, they are often scored in a binary, zero-one fashion. The purpose of this report is to examine and test several techniques of classification using a specified set of these non-metric traits.

First, to clarify the distinction between the two types of variants, one need only note that a non-metric trait is scored only as present or absent. Examples include the presence or absence of torii, accessory bones, and unfused sutures. Figure one illustrates several of these traits. Continuous traits, on the other hand, are always present; they

are variable in degree of expression. Degree of prognathism, shape of the glabella, and inter-orbital distance represent a few of these traits. At least in theory, a discontinuous trait involves a well defined dichotomy, while the continuous trait may require a precise description and considerable experience in the recognition of several states of the trait. Metric traits are the major subset of continuous traits.

It is the simplicity in scoring that makes the use of non-metric variables so attractive. In addition, these traits do not require a skeleton to be reconstructed before analysis. This proves to be a critical advantage in many archaeological situations, since assembling fragmented and deteriorated remains can be a painstaking, if not impossible task. In practice, the scoring dichotomy is not as well-defined. Finnegan (personal communication) suggests intra-observer error may be as high as 5% for some traits. Certainly the inter-observer error would be higher. However, the continuous traits are also subject to error in measurement and judgment. Superiority of variable type is still subject to debate, but the relative analytical ease provided by the non-metric traits in the case of damaged or deformed material is sufficient justification for further study.

Of the two experiments mentioned utilizing morphological traits for classification purposes, Anderson (1968) used strictly discontinuous variables for his classification scheme. His approach was quite intuitive, involving the examination of the trait sequence for an individual and the tallying of supports for populations based on overall trait frequencies for these populations. The individual cadaver was classified into the population which had accumulated the plurality of tallies. Anderson did

not give figures for his example, but he claimed to have classified all individuals in his two samples correctly. This technique is explained more thoroughly and then tested subsequently in this report.

Larnach and MacIntosh (1966, 1967) employed a combination of continuous and discontinuous variables in their discrimination method, even expressing metric traits as discrete random variables. Twenty cranial traits were selected for analysis, each trait represented by three states. Weights of 0, 1, 2, or 3 were variously selected for each state trichotomy, and these weights were summed for each individual's trait sequence. Scatter plots of the summed weights were then examined for population clusters, and discrimination based on the limits defined by these clusters. Results for this technique depended upon the traits selected from the original set of twenty, upon the distinctions between the three states for each trait, and, perhaps most importantly, upon the weights assigned to these states. The procedure gave great advantage to the separation of native Australians from other groups and performed this task well, even though it sacrificed discrimination of Europeans and Mongoloids.

It would appear that the advantages affected by this directional weighting are overridden by the losses in generality. The three states seem to be determined in a somehwat arbitrary fashion, particularly for the cephalic index (< 75, 75 to 80, > 80). Similarly, the distinction between trace and distinct states is unclear and thus subject to considerable scoring error. If these state criteria were chosen to maximize the separation for Aborigines, then exclusion of the Australians would imply the necessity of new criteria. It is in this sense that the

generality of the technique suffers; there is no clear system for the assignment of state demarcation points and weights. Larnach and MacIntosh also recognized the shortcomings of discrimination rules based on small samples (only 8 Europeans and 17 Mongoloids were used to establish the clusters for the classification problem of 1967).

Genetics

Prior to the examination of any new discrimination procedures, it is necessary to understand something of the expression mechanisms for the traits, since several assumptions required for the application of these procedures are dependent upon the origin of the trait. The genetic foundation of non-metric traits, often assumed by early investigators, was not demonstrated experimentally until Grüneberg (1951,1952) identified the repetition of several non-metric traits in successive inbred strains of mice. He concluded that the variables under study were influenced by not one, but a number of interacting genes, and that eventual suppression or expression of a trait was determined by a continuous-model threshold system. Subsequent studies by Searle (1954) and Deol and Truslove (1957) provided evidence for some environmental influence on trait appearance as well. Sjøvold (1977) noted Howe and Parsons' (1967) view that this non-genetic effect could be minimized by examining sizable trait sequences.

Unfortunately, the human species is not readily adaptable to experimental studies such as these; hence the information and results derived from the mice studies have been assumed concordant with human physiology. Certainly the trait sequences for the mice (Grewal 1961)

involved variations quite similar to those in man. Laughlin and Jørgensen (1956) were the first to make use of this genetic basis of non-metric cranial traits to quantify the diversity between human populations. Their efforts inspired work which continues through the present in the area of genetic distance. Again with mice, Grewal (1961) proposed a statistic derived by C.A.B. Smith based on transformed trait frequencies, which became the core for many later studies. Berry and Berry (1967) initially applied this statistic to human populations, while others (Green and Suchey 1976, Finnegan 1972, Sjøvold 1973, and Constandse-Westermann 1972) proposed improvements or modifications to the original derivation. Finnegan (1978) also examined the use of infracranial morphological variation to identify racial diversity, where previous work had been almost exclusively limited to the cranium. This direction of study is mentioned not only to establish the chronology of non-metric trait analysis, but also to note that the assumptions required for the classification procedures have come under close scrutiny.

Trait Frequency Estimation

Population trait frequencies must be estimated by sample frequencies in any analytical procedure. Three basic methods have been employed in this estimation:

Method 1. Tally every occurrence of the trait, treating each side of the skeleton as a separate entity if the trait is subject to bilateral (left and right of the sagittal plane) expression. Divide by the total number of sides (not individuals) examined. This method assumes that the

population trait frequencies are equal for the two sides in bilateral cases. This assumption is satisfied in most situations (Howe and Parsons 1968, Korey 1970, Finnegan 1972, Finnegan 1978) although Sjøvold (1977), with red fox data, discovered discrepancies between some traits' bilateral proportions. While this estimation method has been frequently used, it does not consider the high correlation encountered in bilateral traits. If there is high correlation between sides, then the scoring both repeats information and may incorrectly reduce the variance of the population frequency estimate by as much as a factor of two. In addition, those traits which do not satisfy the assumption of equal side frequency are abandoned, perhaps causing the loss of considerable discrimination power.

Method 2. Tally a bilateral trait when it is expressed on either or both sides. Divide by the number of observed individuals. Again the assumptions of equal bilateral occurrence must be satisfied. The effect of correlation is glossed over with this method also, and no distinction is made between dual or single expression. Probably the most crucial drawback to this method is that of missing observations. Individuals with one damaged side may be totally misrepresented in the statistic (one side unobservable, the other side not expressed).

Method 3. Score one side of a bilateral trait only. Divide by the number of observations. Certainly a great deal of information is lost through the unscored side. However, an unbiased estimate of the proportion for the recorded side results, and missing observations can be ignored without serious consequences.

In each method of estimation, unilateral (midline) trait frequencies are estimated based on the number of observations. Note should be taken that the dichotomy "present-absent" has been avoided. This is to eliminate confusion with trait descriptions. For example, the trait mastoid foramen is tallied when absent. Precondylar tubercle is tallied when present. Thus, tallying indicates the trait definition has been expressed. Missing observations are those damaged beyond recognition or not available for analysis.

Intertrait Correlation

With high side-to-side correlation in bilateral traits, and with the genetic explanation of trait expression suggesting multi-trait influence by a single gene, the question of intertrait correlation has also been examined. Truslove (1961), in analyzing mice data, found a number of associated characters and delineated these further in the next generation. Over half of the resulting correlated trait pairs could be explained physiologically. Although Hertzog (1968) found a notable proportion of significantly related pairs, an analysis of the same data by Benfer (1970) suggested fewer correlations and low statistical association for those that remained significant. Corruccini (1974), treating sexes and races separately, found fewer significant correlations than might randomly occur while using a X^2 test. He then employed Bartlett's test at a different significance level (.01) and found approximately 2% of the pairs significantly correlated. However, even these pairs had low correlations, suggesting a true, but minor relationship between some traits. This same conclusion can be drawn from the work of Sjøvold (1977),

who used red fox data. Consensus thus holds that few traits can be shown to be statistically correlated; those that are can either be explained biologically, or they show a very low degree of association. This follows the genetic model well; independent genes, acting collectively, yield generally disassociated traits.

(Note: Caution must be observed when employing the X^2 test for independence, since small samples might well produce zero frequencies in the 2x2 contingency tables. In those situations, the X^2 test is not appropriate.)

Sex and Age Effects

The findings of Berry and Berry (1967) and Korey (1970), that sex has no significant effect on trait proportions, has been questioned by several authors. In the studies of Corruccini (1974), Sjøvold (1977), and Finnegan (1972, 1978), results indicate significant sex dimorphism for some traits in some populations. Each of these latter researchers has also found that skeletal age might well be a variable influencing the appearance of a trait and thus the population trait frequency. This age effect should not be confused with pathological change; non-metric traits are assumed to be variations which do not significantly affect health or survival. Whether sexes can be pooled and age differences ignored are queries which must be suitably answered prior to any parameter estimation.

With this background of non-metric traits in mind, the assumptions and situations pertaining to the discrimination problem are now confronted. Concerning both the sex and age considerations, one must recall that the cadaver is the only evidence on which to base identification. Therefore,

sex and age must be estimated by imperfect methods just as the race variable is determined. While techniques for sex and age determination have markedly improved in the last two decades, mistakes still occur. More fundamentally, with the factor of age in particular, sample sizes from archaeological populations are too small to yield "good" estimates of trait frequencies for a number of age categories. Nevertheless, if reasonable assurance is obtained for the sex and/or age of a cadaver and substantial sample sizes can be obtained for the corresponding age/sex subset, the classification results should be improved. The empirical tests reported here, however, involve pooled sex and age categories within racial samples to obtain estimates of trait proportions.

The fact that there have been only a few instances of demonstrated low intertrait correlation proves extremely valuable when modeling discrimination procedures. Several of the examined techniques are based on the assumption of trait independence. If each trait is assumed to follow a Bernoulli distribution, the uncorrelated traits are assumed independent, and the likelihood of any individual trait sequence is the product of estimated proportions for scored and unscored traits.

One of the three estimation methods must still be selected. The procedure chosen must be conformable to the classification analysis. Hence, traits should be treated in accordance with the assumptions and conventions listed above. Thus, the decision is somewhat more involved than simply generating distance statistics. Method one will be excluded, since treating the sides of a bilateral trait independently is not statistically valid. The second method encounters obstacles with missing observations which occur frequently in the samples studied here. The

final technique of estimation has been employed at the risk of losing the unused information from the opposing side. In the material analyzed, sides have been treated separately; that is, only one side has been used for estimation and classification, strictly left or strictly right.

Because of the relatively small samples in these data sets and the infrequent (or frequent) expression of some traits, proportion estimates occasionally reach the limits zero and one. Unless the zero and one probability estimates are in some way modified, several of the discrimination procedures become intractable due to attempts of division by zero or extraction of a logarithm of zero. While the actual trait frequency in the population may be very low (or high), it is highly unlikely that any trait is uniformly excluded or exhibited in any sizable population. Thus the modification involves a question of the magnitude of the frequency replacement value. Dillon and Goldstein (1978) followed the advice of Moore (1973) and replaced zero by 10^{-5} and one by $1-10^{-5}$. This facilitated likelihood computation, while maintaining an extreme estimated frequency. However, the choice of this modification seems somewhat arbitrary. Choice of even smaller values better reflects the sample results. As smaller corrections are chosen, the trait theoretically becomes a better discriminator.

A most conservative correction has been chosen for use in these tests, in order to avoid the artificial increase in discrimination power of traits with zero or one sample frequencies. Zero frequencies are replaced by $\frac{1}{n_{ij}+1}$ and ones by $1 - \frac{1}{n_{ij}+1}$ where n_{ij} is the number of observations for trait $\underline{i}$ in population $\underline{j}$. This decidedly conservative correction suggests that the next future observation will exhibit the previously unseen state.

DISCRIMINATION METHODS

Five discrimination procedures have been tested. These are labeled as the tally method, Bayes' theorem, weight of evidence, the linear discriminant function (LDF), and the trait-distance technique. The assumptions required for each procedure, as well as the advantages and disadvantages of each, are explained. Discrimination rules are stated and the empirical results discussed. In every case, the samples are assumed to be randomly selected from the racial population. The individuals in these samples are then used to obtain the estimated population proportions and are then classified based on the derived discrimination functions. Misclassification estimates are determined from these results.

Define the following variables:

x_{ij} = value of observation for trait $\underline{i}$ population $\underline{j}$

$$= \begin{cases} 0 \text{ (untallied)} \\ 1 \text{ (tallied)} \\ \text{missing} \end{cases}$$

$\underline{x} = (x_1, x_2, x_3, \ldots, x_k)$ = trait sequence for an unknown individual

k = number of traits

s = number of populations in the data set

m_{ij} = number of tallies for trait $\underline{i}$ in population $\underline{j}$

n_{ij} = number of observed individuals for trait $\underline{i}$ in population $\underline{j}$

n_j = number of individuals in sample from population $\underline{j}$ $[n_{ij} \leq n_j]$

$\pi_{ij} = P(x_i = 1 | w_j)$ = probability that the $\underline{i}$th trait is tallied in population $\underline{j}$ = trait frequency

$1-\pi_{ij} = P(x_i = 0 | w_j)$ = probability that the $\underline{i}$th trait is not tallied in population $\underline{j}$.

w_j = population $\underline{j}$

$$P_{ij} = \frac{m_{ij}}{n_{ij}} = \text{sample estimate of } \pi_{ij}$$

$$1-P_{ij} = \frac{n_{ij} - m_{ij}}{n_{ij}} = \text{sample estimate of } 1-\pi_{ij}$$

$P(w_j)$ = prior probability of an individual belonging to population j

The Tally Method (Anderson 1968)

As mentioned before, this relatively simple procedure is intuitively appealing. It assumes independence of traits, though this was not made clear by Anderson. Establish a table of estimated trait frequencies for all populations in the data set:

Trait 1	p_{11}	p_{12} ···	p_{1s}
Trait 2	p_{21}	p_{22} ···	p_{2s}
.	.		
.	.		
.	.		
Trait k	p_{k1}	p_{k2} ···	p_{ks}

Select an individual for classification, $\underline{x} = (x_1, x_2, x_3, \ldots, x_k)$.

If x_i is a missing observation, proceed to the next trait.

If x_i is a 1, place a tally or a unit of support for the population j which has the largest proportion p_{ij}.

If x_i is a 0, place a tally or a unit of support for the population j which has the smallest proportion p_{ij}.

When all k traits have been examined, classify the individual to the population which has accumulated the largest number of tallies. When a tie occurs between populations with the same accumulated supports, classify to the population with the largest sample. A different criterion also based on prior population membership should be employed in

classifications of truly unknown individuals when ties occur.

The simplicity of this technique is its main asset. But the fact that all traits are given equal discriminatory power is a serious defect that overrides simplicity. One way to improve the technique would be to establish some limiting rule on the number of traits used, perhaps only those which show significant frequency differences between populations. However, as the number of populations in the data set increases, fewer traits could be excluded. This alternative has not been employed. Another way to handle the analysis would be to introduce some weighting mechanism for the traits to replace the unit tally.

Bayes' Theorem (Duda and Hart 1973)

Define $P(w_j|\underline{x})$ as the probability that an individual with trait sequence $\underline{x}$ belongs to population $\underline{j}$. By Bayes' rule,

$$p(w_j|\underline{x}) = \frac{P(\underline{x}|w_j)P(w_j)}{\sum_{t=1}^{s} P(\underline{x}|w_t)P(w_t)}$$

Since the denominator remains constant for any given discrimination procedure regardless of the chosen population w_j, the classification rule is based on $P(\underline{x}|w_j)P(w_j)$. The individual will be classified to the population $\underline{j}$ for which the probability of his membership is highest.

Estimates for the priors $P(w_j)$ will be based strictly on sample sizes for this study.

$$P(w_j) = \frac{n_j}{\sum_{t=1}^{s} n_t}$$

In the case of a truly unknown individual, these priors could well be determined by demographic factors of the area in which the body was found. A sociologist or anthropologist might also estimate this prior membership probability based on socio-cultural elements.

To estimate $P(\underline{x}|w_j)$, assume the traits are independent. Then, since each trait follows a Bernoulli distribution,

$$P(x_{ij}|w_j) = P_{ij}^{x_{ij}} (1-P_{ij})^{1-x_{ij}}$$

and the maximum likelihood estimate of $P(\underline{x}|w_j)$ is:

$$P(\underline{x}|w_j) = \prod_{i=1}^{k} P_{ij}^{x_{ij}} (1-P_{ij})^{1-x_{ij}}$$

where missing observations are ignored and the proportion estimates have been substituted for the parameters. The natural logarithm of $P(x_{ij}|w_j)P(w_j)$ is a monotonic increasing function and is more amenable to digital computers. The classification rule is then to classify an individual to the population $\underline{j}$ which has the largest value $g(w_j)$:

$$g(w_j) = \sum_{i=1}^{k} [x_i \log P_{ij} + (1-x_i) \log (1-P_{ij})] + \log P(w_j)$$

This function can be readily interpreted, since one of the two terms within the summation operator drops out for each trait. The weighting factor absent in the tally method is here based on the trait frequency. Finally, the term involving the prior may be seen as an adjustment factor. These are distinct advantages in favor of this procedure, as long as reasonable estimates of the priors can be obtained. The disadvantage of this technique is its slight weakness in assuming trait independence.

Note that this classification procedure can be written as a likelihood ratio when only two populations are involved:

$$f(\underline{x}) = \frac{P(\underline{x}|w_1)P(w_1)}{P(\underline{x}|w_2)P(w_2)}$$

$$f^*(\underline{x}) = \sum_{i=1}^{k} [x_i \log (\frac{P_{i1}}{P_{i2}}) + (1-x_i) \log (\frac{1-P_{i1}}{1-P_{i2}})] + \log (\frac{P(w_1)}{P(w_2)})$$

Classify to w_1 if $f^*(\underline{x}) > 0$

to w_2 if $f^*(\underline{x}) \leq 0$

The weighting effect is even more evident here. The larger the discrepancy between P_{i1} and P_{i2}, the more discriminating power is associated with trait $\underline{i}$.

This procedure has also been called the first order Bahadur method (Dillon and Goldstein 1978) based on Bahadur's (1961) interaction model with interactions excluded and a prior term added.

Weight of Evidence (Osteyee and Good 1974)

This classifier is very similar in its final decision rule to the preceding Bayes' technique. The assumption of intertrait independence is again necessary. The authors describe this procedure only for classifications between two populations as follows:

$W(w_1|w_2:\underline{x})$ = weight of evidence in favor of w_1 over w_2 given a trait sequence $\underline{x}$.

$$= \log \frac{P(w_1|\underline{x})}{P(w_2|\underline{x})} - \log \frac{P(w_1)}{P(w_2)}$$

where $P(w_j|\underline{x})$ and $P(w_j)$ are as defined previously.

Now $$\log \frac{P(w_1|\underline{x})}{P(w_2|\underline{x})} = \log \frac{P(\underline{x}|w_1)P(w_1)}{P(\underline{x}|w_2)P(w_2)}$$

$$\text{Thus } W(w_1|w_2:\underline{x}) = \log \frac{P(\underline{x}|w_1)}{P(\underline{x}|w_2)} + \log \frac{P(w_1)}{P(w_2)} - \log \frac{P(w_1)}{P(w_2)}$$

$$= \sum_{i=1}^{k} [x_i \log \frac{P_{i1}}{P_{i2}} + (1-x_i) \log (\frac{1-P_{i1}}{1-P_{i2}})]$$

Classify to w_1 if $W(w_1|w_2:\underline{x}) > 0$

$W(w_1|w_2:\underline{x}) \leq 0$

The discriminator is exactly the same as that based on Bayes' Theorem with the adjustment term due to priors excluded. Although no extension for this method for more than two populations has been given in the literature, a test will be made using the Bayes classifier without the prior adjustment term:

$$G(w_j) = \sum_{i=1}^{k} [x_i \log P_{ij} + (1-x_i) \log (1-P_{ij})]$$

with the same decision rule as that for Bayes' Theorem.

The value of this technique is clear when an investigator has no reason to believe one population is any more or less probable than another.

Linear Discriminant Function (Morrison 1976)

This technique has been derived in detail in the statistical literature and will not be discussed in detail here. The assumptions necessary for use of this discriminant function are by no means satisfied; thus, it should not yield optimal results. Primarily, the data do not follow a multivariate normal distribution. In fact, each trait is dichotomous taking on a value of 0 or 1. Also, the assumption of equal covariance matrices for the respective populations are not equal since, for any trait i in population j,

$$Var(x_{ij}) = \pi_{ij}(1 - \pi_{ij})$$

and for population $q \neq j$, $\pi_{ij} \neq \pi_{iq}$ or else there is no power of discrimination. Studies by Moore (1973) have shown that these assumptions are not as crucial when the intertrait correlations are small. His work suggests that results using this procedure should correspond closely with those of the Bayes procedure.

All calculations for the linear discriminant function were performed through the Statistical Package for the Social Sciences (SPSS) (Nie, et al., 1975) and the computations involved with this program point up another deficiency of the procedure. All observation vectors which have at least one missing observation are not included in estimates of population trait means and covariance matrices. It has been noted that many skeletons are incomplete when received for investigation, and those making up these samples are no exception. Upwards of 80% of some samples are excluded in parameter estimation.

One advantage of the linear discriminant function is that it takes into account the correlations between traits in establishing the discrimination rule. It is surprising then that only when there is low intertrait correlation is the technique effective.

Trait-distance Technique (Kullback 1959)

This procedure was developed from Kullback's work in information theory in an effort to distinctly identify the discriminating power of traits. It assumes trait independence. The method is defined only for two populations and will be tested only for population pairs.

The information for each trait i in favor of respective populations is defined by:

$$I(w_1:w_2) = (1-P_{i1}) \log \left(\frac{1-P_{i1}}{1-P_{i2}}\right) + P_{i1} \log \left(\frac{P_{i1}}{P_{i2}}\right)$$

$$I(w_2:w_1) = (1-P_{i2}) \log \left(\frac{1-P_{i2}}{1-P_{i1}}\right) + P_{i2} \log \left(\frac{P_{i2}}{P_{i1}}\right)$$

The distance between w_1 and w_2 for trait $\underline{i}$ is:

$$D_i = J(w_1:w_2) = I(w_1:w_2) + I(w_2:w_1)$$

$$= (P_{i1} - P_{i2}) \log \frac{P_{i1}(1-P_{i2})}{P_{i2}(1-P_{i1})}$$

$$= (P_{i1} - P_{i2}) \left[\log \left(\frac{P_{i1}}{P_{i2}}\right) + \log \left(\frac{1-P_{i2}}{1-P_{i1}}\right)\right]$$

The larger the difference between P_{i1} and P_{i2}, the larger is D_i, and the better is that trait as a discriminator. The discriminant function is:

$$Q(\underline{x}) = \sum_{i=1}^{k} (a_i) D_i$$

where a_i is directional as follows:

a_i	conditions	support for
1	$P_{i1} > P_{i2}$ $x_i=1$	w_1
-1	$P_{i1} > P_{i2}$ $x_i=0$	w_2
-1	$P_{i1} < P_{i2}$ $x_i=1$	w_2
1	$P_{i1} < P_{i2}$ $x_i=0$	w_1

The discrimination rule is:

classify to w_1 if $Q(\underline{x}) > 0$

w_2 if $Q(\underline{x}) \leq 0$

It is a simple matter, after determining the distances for any population pair, to select those traits which have the most ability to

distinguish between these populations. Test results reported here used fifteen traits, since that number accounted for most of the discriminant function sum.

MATERIALS

Uniformity among authors involved with non-metric osteological traits, not only in terms of their analytical procedures, but also regarding their trait lists used in these analyses, is virtually non-existent. Corruccini (1974) scored seventy-two cranial variants while Suchey (1975) utilized only twenty-seven. Forty-two cranial traits defined by Finnegan (1972), thirty-five bilateral and seven midline, are used for one data set in this report. The set consists of six Hungarian archaeological samples from the Avar period (A.D. 555-796) previously studied by Finnegan and Marcsik (1979). As evidenced in the trait lists and sample frequencies of Tables 1-2, many of the individuals lack complete trait sequences.

A second data set has been scored for thirty infracranial traits defined by Finnegan (1978). All of these traits are subject to bilateral expression. The samples constituting this set are American Blacks and American Whites from the Terry collection and Yukon River Eskimos, Coast Eskimos, and Aleutians from the assemblage housed at the United States National Museum of Natural History. The Terry collection is essentially a well-documented group of dissection room cadavers; thus, there are but a very few missing traits in its two samples. The National Museum samples, on the other hand, are composed of archaeological specimens which often possess incomplete trait sequences (Tables 3-4).

Examination and scoring of all the individuals involved in this study was performed by Dr. Michael Finnegan to whom this author expresses sincere gratitude.

DISCUSSION

Results of Pairwise Discrimination Procedures

All misclassification rates are the "apparent errors" (Dillon and Goldstein 1978) since all parameters are estimated and the observation vectors used to obtain these estimates are then classified by the derived discriminator.

Scrutiny of the results by classification method (Tables 5-8) indicates that, at least for these two data sets, the Bayes and weight of evidence procedures are in general the most effective discriminators. Interestingly, there is very little difference in correct overall classification probabilities between these two techniques and no preference can be stated for either if the penalty for misclassification is equal for the two populations. This seems to hold for all cases of variation in sample size (which has determined the priors for the Bayes method). There is, however, a predictable difference between the two procedures when the misclassifications are broken down and examined. The larger samples are consistently favored by the Bayes method, and classification to the smaller group reflects the necessary sacrifice. Conversely, the weight of evidence technique distributes the misclassifications more proportionately.

The classification rate for the linear discriminant function can be seen to fluctuate wildly about that of the Bayes and weight of evidence methods. Work by Moore (1973) and Krzanowski (1977) has implied that under the assumption of trait independence, the LDF performs quite competitively with these other procedures. Ironically, while the LDF

considers intertrait correlations, it is a poor discriminator (compared to other techniques) when even moderate correlations exist. Simulations by Dillon and Goldstein (1978) have borne out this theory. It remains then to explain why the classifier has behaved so erratically here. The answer is most probably related to the mechanics of parameter estimation. The incomplete individuals in a sample are classified based on the estimated parameters obtained from those which are complete. In the most extreme case, only one of the twenty-two individuals in one Hungarian sample is involved in the estimation. This problem can be eliminated only by the accumulation of large samples of complete individual trait sequences. (Such groups are not presently available for analysis.) In several pairwise comparisons, incompleteness prevents determination of the linear discriminant function since too few individuals are involved to allow inversion of the covariance matrix. These comparisons are denoted in the tables by NP (not possible). Until sufficiently large samples are assembled to allow for reliable parameter estimation, the linear discriminant function should not be selected as a classifier. It should be noted that the difference in overall misclassification rates is negligible for the cases of included priors and eliminated priors in the LDF.

The trait-distance procedure frequently performs less effectively than the Bayes and weight of evidence procedures, and the misclassifications follow no pattern with regard to sample sizes. The error rate does not fluctuate so widely as that for the LDF since the missing observations do not cause such an estimation problem. One advantage of this method is that it clearly identifies traits with superior discriminating power. Nevertheless, this factor of clarification does not compensate for inferior classification rates.

With the preceding five classification techniques based on statistical methodology, one might have a tendency to believe that any one of the five would prove superior to the strictly intuitive tally method. It is somewhat surprising then that this latter procedure proves to be a competitive, albeit inferior, classifier when compared with the other five methods. As with the trait distance procedure there is no predictable pattern to misclassifications when considering sample size, although small samples appear to be at a severe disadvantage in many of the cranial tests. Error rates can be best described as varying erratically between 15% and 45% among population pairs for the tested data.

It is clear from the results that the techniques tested perform quite differently when comparing left and right side analyses. This can be attributed largely to the varying trait sample sizes caused by missing observations. Certainly a procedure which takes account of the bilateral correlations would help to alleviate this apparent classification discrepancy, but this report does not investigate any such procedure through empirical testing.

While comparison of the pairwise methods with regard to sides of a bilateral trait employed clearly indicate differences, it is difficult (if not impossible) to contrast the effectiveness of cranial and infracranial osteological traits with the information available for these tests, since different populations and different individuals are involved. At any rate, neither data set presents a conclusive preference for either cranial or infracranial scoring.

Finally, within the infracranial data set, discrimination between the Amerindian groups is less precise than the discrimination between each

of these groups and the Black and White samples. This reflects the genetic relationship of the Eskimo groups.

Results of Multi-Population Classification Procedures

These are by far the most interesting and useful results of the study, since in practical situations, consideration of several populations is more likely than is just a pair. It is interesting to note that past studies and simulations have dealt almost exclusively with the pairwise classification problem.

It is again evident that the Bayes and weight of evidence methods stand out as the most effective classification procedures (Tables 9-13). As in the pairwise population comparisons, the overall misclassification rates are virtually indistinguishable for the two procedures. Investigation of each table indicates the same Bayes advantage to large samples and more proportional error rates for the weight of evidence method. As mentioned previously, it may be difficult to establish any prior population membership probability. These results imply that there is no significant increase in misclassifications if equal priors are assumed.

Of the remaining techniques, the linear discriminant function with proportional priors performs consistently better than that function with equal prior membership probabilities, although the superiority in every case is slight. As expected, the method involving equal priors yields lower misclassification rates for the smaller samples while the technique accounting for varying population likelihoods favors the larger samples most often. It is unlikely that the increased misclassifications using the two discriminant function classifiers as compared to the Bayes and weight of evidence procedures can be largely attributed to the

information in incomplete trait sequences which is lost to the LDF techniques.

The final and least effective technique tested is the tally method. This unsatisfactory performance is not surprising upon noting that the method ignores all but two population frequencies for each trait. The pairwise analysis allows each population in question the possibility of support. Only two of the groups in the multi-population comparisons, those with the largest and smallest frequencies for each trait, are eligible for a tally. The possibility thus exists that some population with intermediate trait frequencies could receive little or perhaps no chance at accumulating support. The cranial data set results reflect such a situation with only the largest sample receiving assignments. By no means should this procedure be employed for multi-population comparisons.

Discrimination based on the cranial traits is notably inferior (in terms of errors) to that based on the infracranial scoring. However, this difference can be attributed to several factors, only one of which is the trait lists. The samples which compose the cranial set vary chronologically, not geographically, and could thus be related in some undocumented micro-evolutionary sense. On the other hand, the infracranial data consist of three clearly defined racial stocks. It is therefore quite possible that the Hungarian cranial samples appear more homogeneous as a group than do the five American samples. In addition, the effects of smaller samples, more samples, and archaeological collection (with subsequent frequency of missing observations) must be considered as sources of the large discrepancy in the error rates between the two data

sets. This discrepancy can be seen to hold for each of the five discrimination techniques tested.

Other Discrimination Methods

The six pairwise procedures examined here are by no means the sole discriminators available. Brown (1971), Dillon and Goldstein (1978), Hills (1966), and Moore (1973) are a few of the researchers who have examined or designed classification models for binary type data. All of these methods deal with the interactions between variables (traits). However, the simulations by Dillon and Goldstein as well as those of Brown imply the clear superiority of the linear discriminant function and the Bayes (first-order Bahadur) procedures when variables are independent.

Of the interaction models proposed, it would seem that the full-multinomial model defined by Moore which accounts for interactions of all orders would yield the most satisfactory results. But the generality of the technique is also its downfall, since for each population in question, 2^k (where k = number of traits) probabilities must be estimated. Missing observations further complicate the implementation of the method, allowing for 3^k possible trait sequences. Present data sets are definitely inadequate for even attempting this full multinomial method since the vast majority of trait sequence probabilities could not be estimated. Furthermore, the lack of even simple correlations between traits does not bode well for the discovery of many significant higher order interactions.

While the full multinomial model appears unreasonable, even for future employment, those methods which consider two-way interactions may

prove more promising if both sides of a bilateral trait are to be considered. The second-order Bahadur reparametization in particular deserves trial. Higher-order interactions could again be ignored due to low intertrait simple correlations and the close bilateral intratrait association.

Another technique which utilizes the close association of bilateral trait sides may possibly salvage the advantages of the Bayes and weight of evidence procedures since the design follows those very closely. Proposed by Rubison (1978), the procedure follows.

For a population w_j, $\underline{x} = (x_1, x_2, \ldots, x_k)$ and $\underline{y} = (y_1, y_2, \ldots, y_k)$ represent the left and right trait sequences for one member of that population. Let

$$P_{i,qr,j} = P\{(x_i = q,\ y_i = r) | w_j\}$$

$i=1,2,\ldots,k$
$q=0,1$
$r=0,1$
$j=1,\ldots,s$ (s = # of populations)

Then assuming independence of traits,

$$P(\underline{x},\ \underline{y}|w_j) = \prod_{i=1}^{k} p_{i,00,j}^{(1-x_i)(1-y_i)}\ p_{i,01,j}^{(1-x_i)y_i}\ p_{i,10,j}^{x_i(1-y_i)}\ p_{i,11,j}^{x_i y_i}$$

The decision rule is to classify to the population $\underline{j}$ which has the largest value $g(\underline{x},\ \underline{y}|w_j)$ where:

$$g(\underline{x},\ \underline{y}|w_j) = \log P(\underline{x},\ \underline{y}|w_j) + \log P(w_j)$$

and $P(w_j)$ is the prior probability of population $\underline{j}$.

The drawback to this classification procedure is that four probabilities must be estimated from a sample. Again the reality of limited

samples presents an obstacle to empirical tests. Nevertheless, preliminary results by McGuire (1978 personal communication) using the two data sets of this study indicate a great deal of potential for this technique.

RECOMMENDATIONS

Many of the problems associated with these tests result directly from the inadequacy of the sample sizes. Thus, it is imperative to assemble substantial sets of individuals, particularly cadavers which yield complete trait sequences. Information must be accumulated from widely scattered, presently existing osteological collections, since addition of new skeletal material occurs only very slowly. The larger samples are necessary not only for the complete trait sequences, but also for the determination of better trait frequency estimates.

Before this data accumulation can take place, several agreements must be reached among anthropologists. Foremost among these are standardized descriptions of the traits. The ideal situation would have one or perhaps two osteologists perform all of the scoring. Secondly, the trait list or at least a core trait list should be established for this analysis and all subsequent analyses. Finally, since cranial and infracranial traits have each demonstrated discrimination ability individually, studies should be conducted utilizing both types of traits combined into one comprehensive trait sequence. Maximum information must be derived and employed in discrimination procedures, a contention supported by this report.

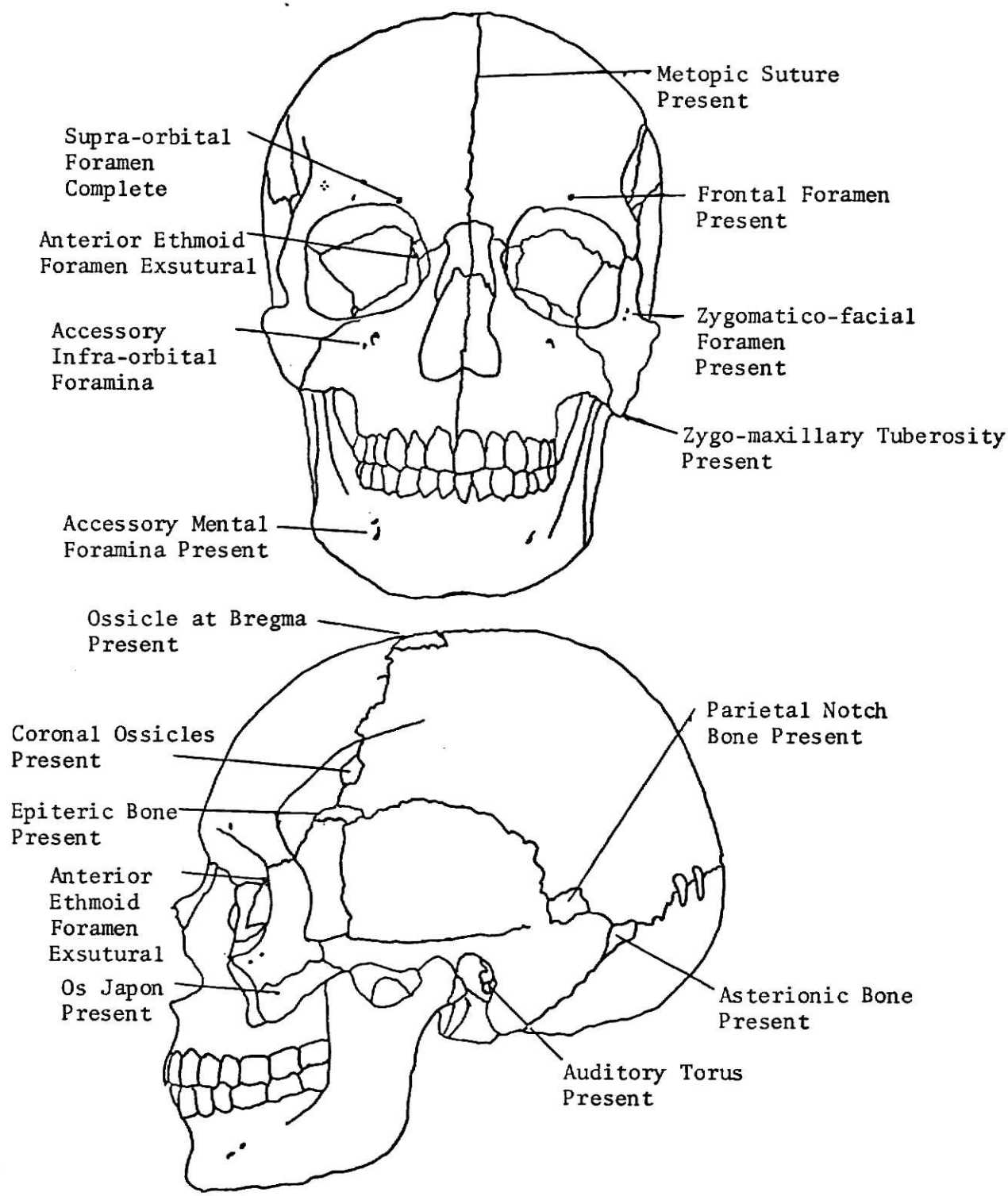

Figure 1. Non-Metric Cranial Traits

Reprinted with permission from A Guide to Osteological Analysis by Michael Finnegan (1978).

TABLE 1. Cranial Traits and Frequencies (Left Side and Midline)

Traits	Populations 1	2	3	4	5	6
1. Highest nuchal line present	16/30	13/32	14/20	23/44	80/96	48/75
2. Coronal ossicles present	0/28	0/31	0/18	1/47	2/100	2/80
3. Ossicle at bregma present	0/28	0/29	0/18	1/47	1/99	0/81
4. Sagittal ossicles present	1/29	0/32	1/19	1/47	16/100	3/81
5. Ossicle at lambda present	1/30	1/33	4/20	3/47	20/98	15/78
6. Lambdoid ossicles present	7/28	7/31	8/16	15/44	46/98	29/76
7. Os inca present	0/31	0/33	0/20	0/46	0/100	0/78
8. Parietal foramen present	15/31	18/32	6/18	23/47	42/100	38/78
9. Parietal notch bone present	1/30	0/31	0/13	0/44	17/99	6/79
10. Asterionic bone present	0/29	1/29	0/12	1/44	3/100	6/77
11. Auditory torus present	0/30	0/28	0/16	0/46	0/100	0/81
12. Malar tubercle present	3/29	0/22	0/17	0/38	0/100	0/80
13. Os japon present	1/27	0/23	0/13	0/43	1/97	1/80
14. Pterion form non H-form	2/28	1/27	0/11	1/44	3/100	5/78
15. Epiteric bone present	1/28	0/26	2/11	5/44	26/99	13/78
16. Infra-orbital foramen present	24/24	21/21	12/12	32/32	99/99	80/80
17. Supra-orbital foramen present	30/30	27/28	20/20	45/45	99/99	81/81
18. Frontal foramen present	11/30	3/27	3/20	9/44	15/99	15/81
19. Metopic suture present	3/31	0/30	1/20	1/46	10/98	5/78
20. Mandibular foramen double	0/27	1/27	0/11	4/44	4/98	4/74
21. Mylohyoid groove closed	1/27	0/27	0/10	0/45	2/99	3/74
22. Mandibular torus present	8/29	0/30	0/18	3/46	5/99	3/78
23. Mental foramen double	3/29	2/30	1/15	4/44	4/99	2/77
24. Palatine torus present	4/27	1/22	1/14	5/42	16/99	24/80
25. Lesser palatine foramen present	20/20	14/14	14/14	24/29	89/96	70/73
26. Foramen of Vesalius present	7/20	3/12	1/12	7/35	20/96	14/71
27. Foramen ovale present	25/25	21/21	14/14	45/45	97/97	75/75
28. Foramen spinosum present	25/26	20/20	14/14	45/45	93/95	73/75
29. Foramen of Huschke present	7/30	6/26	4/15	12/46	5/99	5/81
30. Condylar facet double	0/23	1/25	0/13	0/38	3/96	2/70
31. Posterior condylar foramen patent	7/23	12/26	11/14	28/38	55/95	40/70
32. Precondylar tubercle present	1/25	1/28	0/17	3/44	8/100	8/73
33. Anterior condylar foramen bipartite	6/25	4/29	4/14	6/44	22/99	11/72

TABLE 1 (Contd.)

Traits	Populations 1	2	3	4	5	6
34. Mastoid foramen absent	6/28	5/27	5/16	5/46	21/98	10/77
35. Mastoid foramen exsutural	6/28	5/27	6/16	14/46	28/98	17/77
36. Paramastoid process dependent	23/25	16/19	7/7	28/31	84/96	56/67
37. Diagastric groove double	9/29	8/28	5/16	13/44	26/98	19/76
38. Stylomastoid foramen present	30/30	27/27	13/13	47/47	99/99	81/81
39. Zygo-maxillary tuberosity present	16/27	10/24	4/10	21/41	70/99	61/78
40. Zygo-facial foramen absent	3/27	12/28	3/15	8/45	21/98	19/80
41. Anterior ethmoid foramen exsutural	3/12	5/10	5/5	7/22	13/76	14/67
42. Posterior ethmoid foramen absent	1/18	3/10	0/5	3/25	14/85	5/69

TABLE 2. Cranial Traits and Frequencies (Right Side)

Traits	Populations 1	2	3	4	5	6
1. Highest nuchal line present	14/30	12/32	14/20	22/44	78/96	49/76
2. Coronal ossicles present	0/28	0/28	0/17	0/47	2/100	4/80
3. Ossicle at bregma present	--	--	--	--	--	--
4. Sagittal ossicles present	--	--	--	--	--	--
5. Ossicle at lambda present	--	--	--	--	--	--
6. Lambdoid ossicles present	4/26	11/30	5/16	18/44	46/98	35/77
7. Os inca present	--	--	--	--	--	--
8. Parietal foramen present	19/31	20/31	9/20	29/46	47/100	46/77
9. Parietal notch bone present	5/30	1/25	0/13	2/42	6/100	4/80
10. Asterionic bone present	0/29	0/24	0/11	5/40	9/100	10/78
11. Auditory torus present	0/31	0/26	0/16	0/46	0/100	0/80
12. Malar tubercle present	2/28	0/21	0/17	0/35	0/100	0/80
13. Os japon present	0/26	0/19	0/16	0/38	1/99	0/77
14. Pterion form non H-form	1/26	0/22	0/10	0/42	1/99	4/80
15. Epiteric bone present	2/26	1/21	2/9	6/42	29/99	17/80
16. Infra-orbital foramen present	23/23	16/16	12/12	28/28	100/100	79/79
17. Supra-orbital foramen present	29/29	28/29	17/17	48/48	99/100	80/81
18. Frontal foramen present	12/29	7/29	1/17	12/48	16/100	21/81
19. Metopic suture present	--	--	--	--	--	--
20. Mandibular foramen double	0/29	0/27	1/10	1/42	7/99	2/78
21. Mylohyoid groove closed	2/29	2/30	0/10	1/43	4/99	2/79
22. Mandibular torus present	9/30	0/29	0/17	3/46	6/99	6/78
23. Mental foramen double	4/30	1/29	1/16	1/46	12/99	5/78
24. Palatine torus present	--	--	--	--	--	--
25. Lesser palatine foramen present	22/23	13/16	11/12	25/28	84/95	73/76
26. Foramen of Vesalius present	6/20	2/13	3/10	4/32	22/97	9/72
27. Foramen ovale present	30/30	24/24	10/10	40/40	100/100	76/76
28. Foramen spinosum present	30/30	24/24	11/11	40/40	95/98	76/77
29. Foramen of Huschke present	7/31	5/24	2/15	11/46	7/100	2/80
30. Condylar facet double	0/25	1/20	0/15	0/40	2/95	4/72
31. Posterior condylar foramen patent	11/22	12/23	11/15	26/38	63/96	45/72
32. Precondylar tubercle present	--	--	--	--	--	--
33. Anterior condylar foramen bipartite	3/24	2/23	3/16	4/43	21/100	13/72

TABLE 2 (Contd.)

Traits	Populations 1	2	3	4	5	6
34. Mastoid foramen absent	3/28	9/24	3/18	9/40	18/97	13/76
35. Mastoid foramen exsutural	11/28	5/24	7/18	7/40	22/97	11/76
36. Paramastoid process dependent	21/23	13/16	8/8	29/31	86/97	59/68
37. Diagastric groove double	9/29	6/25	3/17	10/42	35/98	24/77
38. Stylomastoid foramen present	30/31	26/26	14/14	45/45	99/99	80/80
39. Zygo-maxillary tuberosity present	16/28	9/19	7/16	18/36	70/99	54/79
40. Zygo-facial foramen absent	3/26	5/20	0/15	9/41	27/100	15/77
41. Anterior ethmoid foramen exsutural	9/14	8/11	3/4	10/24	16/81	12/63
42. Posterior ethmoid foramen absent	1/20	2/9	0/5	5/25	20/86	6/64

TABLE 3. Infracranial Traits and Frequencies (Left Side)

Traits	Populations				
	Coast	Yukon	Black	White	Aleut
1. Allen's fossa present	29/57	19/49	28/99	30/94	15/46
2. Poirier's facet present	10/49	8/48	66/99	74/94	20/45
3. Plaque formation	14/52	22/49	39/99	34/94	11/46
4. Hypotrochanteric fossa present	24/59	12/51	9/99	8/95	18/47
5. Exostosis in trochanteric fossa	13/59	4/51	24/99	35/96	9/46
6. Third trochanter present	4/58	7/50	7/99	15/96	5/47
7. Medial tibial squatting facet	2/54	0/51	6/100	3/96	10/43
8. Lateral tibial squatting facet	23/54	29/51	24/100	11/96	25/43
9. Supracondyloid process present	0/55	0/51	0/100	4/96	0/46
10. Septal aperture present	10/55	10/51	20/100	15/96	3/46
11. Acetabular crease present	14/61	21/50	10/100	37/95	17/51
12. Pre-auricular sulcus present	27/61	26/51	54/99	42/95	24/50
13. Accessory sacral facets present	9/61	7/51	18/99	24/95	9/51
14. Acromial articular facet present	9/48	20/47	41/100	32/95	14/39
15. Suprascapular foramen present	0/51	0/49	1/100	6/96	0/39
16. Circumflex sulcus present	22/50	17/49	50/100	49/96	17/38
17. Vastus notch present	15/35	8/37	33/99	24/94	22/40
18. Vastus fossa present	30/35	28/36	73/99	70/94	34/40
19. Emarginate patella	0/35	0/36	1/99	0/94	4/41
20. Os trigonum present	4/49	3/44	18/100	11/96	4/33
21. Medial talar facet present	16/50	17/43	27/100	40/96	13/33
22. Lateral talar extension present	10/51	7/44	21/100	24/96	12/33
23. Inferior talar articular surface	18/49	19/44	60/100	81/96	7/33
24. Anterior calcaneal facet double	52/52	43/45	89/99	85/96	34/34
25. Anterior calcaneal facet absent	20/35	19/30	51/100	71/95	18/32
26. Peroneal tubercle present	4/55	5/47	11/100	12/96	2/42
27. Atlas facet form double	4/54	3/47	13/100	6/96	5/41
28. Posterior bridge present	4/52	5/47	3/100	3/96	5/41
29. Lateral bridge present	28/57	22/48	45/100	49/96	14/46
30. Transverse foramen bipartite	3/56	0/47	5/100	3/96	1/48

TABLE 4. Infracranial Traits and Frequencies (Right Side)

Traits	Populations				
	Coast	Yukon	Black	White	Aleut
1. Allen's fossa present	26/57	17/47	31/99	24/94	13/46
2. Poirier's facet present	14/51	7/46	68/99	74/94	19/46
3. Plaque formation	14/52	18/47	43/99	39/94	10/46
4. Hypotrochanteric fossa present	18/59	10/50	7/100	6/94	16/48
5. Exostosis in trochanteric fossa	10/58	1/50	22/100	18/96	10/48
6. Third trochanter present	6/59	4/49	4/100	11/96	6/48
7. Medial tibial squatting facet	3/58	0/49	6/100	4/95	9/46
8. Lateral tibial squatting facet	21/57	30/49	25/100	15/95	27/46
9. Supracondyloid process present	0/55	0/50	0/100	1/96	1/44
10. Septal aperture present	6/55	4/49	14/100	11/96	2/44
11. Acetabular crease present	13/59	22/52	6/100	34/96	15/46
12. Pre-auricular sulcus present	21/59	28/52	55/99	43/96	23/46
13. Accessory sacral facets present	15/59	11/52	18/99	23/96	9/47
14. Acromial articular facet present	8/56	18/42	39/100	31/95	12/39
15. Suprascapular foramen present	0/58	0/50	2/100	4/94	1/42
16. Circumflex sulcus present	25/58	16/51	51/100	52/95	24/41
17. Vastus notch present	15/31	10/39	31/100	18/95	19/43
18. Vastus fossa present	26/31	24/38	63/100	64/95	35/43
19. Emarginate patella	1/31	1/39	0/100	0/95	7/43
20. Os trigonum present	6/45	3/45	18/100	9/96	4/36
21. Medial talar facet present	15/44	18/43	33/100	31/96	15/36
22. Lateral talar extension present	8/44	6/44	18/100	30/96	8/36
23. Inferior talar articular surface	15/45	16/44	61/100	71/96	7/36
24. Anterior calcaneal facet double	47/48	42/45	93/100	90/96	37/37
25. Anterior calcaneal facet absent	15/32	15/27	41/100	73/95	21/35
26. Peroneal tubercle present	9/55	3/47	15/100	10/96	2/42
27. Atlas facet form double	2/55	4/47	14/100	6/96	2/41
28. Posterior bridge present	6/55	6/47	0/100	2/96	4/41
29. Lateral bridge present	22/57	26/47	50/100	51/96	20/46
30. Transverse foramen bipartite	2/56	1/47	5/100	3/96	1/48

TABLE 5. Pairwise Classification - Infracranial Data Set (Right Side)

Population Pair	Sample Size	Tally	Bayes' Theorem	Weight of Evidence	LDF Prop. (Priors)	LDF Equal (Priors)	Trait-Distance
Coast	62	28	15	15	0	0	22
		28.70	26.09	25.22	36.52	36.52	28.70
Yukon	53	5	15	14	42	42	11
Coast	62	17	18	13	41	41	21
		19.14	16.67	16.05	25.93	25.93	19.14
Black	100	14	9	13	1	1	10
Coast	62	22	13	11	34	32	18
		19.62	12.66	12.03	22.15	20.89	15.82
White	96	9	7	8	1	1	7
Coast	62	5	15	16			12
		36.28	24.78	25.66	NP	NP	35.40
Aleut	51	36	13	13			28
Yukon	53	8	13	8	27	26	9
		28.76	13.73	15.03	19.61	19.61	18.95
Black	100	36	8	15	3	4	20
Yukon	53	8	14	11	14	12	9
		18.12	14.09	14.09	13.42	13.42	16.72
White	96	19	7	10	6	8	16
Yukon	53	10	10	10	36	36	1
		27.88	20.19	20.19	37.50	37.50	37.50
Aleut	51	19	11	11	3	3	38
Black	100	61	25	25	19	20	20
		35.20	22.45	21.94	17.86	18.37	22.96
White	96	8	19	18	16	16	25
Black	100	8	6	8	1	2	3
		17.22	9.27	7.95	13.91	14.57	17.88
Aleut	51	18	8	4	20	20	24
White	96	8	2	7	1	1	3
		16.33	5.44	7.48	16.33	14.97	14.29
Aleut	51	16	6	4	23	21	18

Cell values are:
number of misclassifications in sample A;
overall misclassification percentage;
number of misclassifications in sample B.

TABLE 6. Pairwise Classification - Infracranial Data Set (Left Side)

Population Pair	Sample Size	Method: Tally	Bayes' Theorem	Weight of Evidence	LDF Prop. (Priors)	LDF Equal (Priors)	Trait-Distance
Coast	62	18	13	16	0	0	17
		33.91	25.22	26.09	33.91	33.91	27.83
Yukon	53	21	16	14	39	39	15
Coast	62	12	17	11	38	36	17
		26.54	18.52	20.37	29.63	29.63	28.40
Black	100	31	13	22	10	12	29
Coast	62	13	12	9	20	17	17
		22.15	14.56	15.82	16.46	15.19	18.35
White	96	22	11	16	6	7	12
Coast	62	3	8	13	51	51	4
		42.48	22.12	22.12	46.90	46.90	29.20
Aleut	51	45	17	12	2	2	29
Yukon	53	12	17	10	27	23	10
		30.07	18.30	20.92	20.26	19.61	22.22
Black	100	34	11	22	4	7	24
Yukon	53	3	10	5	12	11	6
		34.9	12.08	10.07	9.40	9.40	14.09
White	96	49	8	10	2	3	15
Yukon	53	5	12	12	33	33	3
		30.77	23.08	22.12	35.58	35.58	29.81
Aleut	51	27	12	11	4	4	28
Black	100	20	21	22	19	20	16
		33.16	20.92	21.43	18.37	18.37	26.53
White	96	45	20	20	17	16	36
Black	100	20	11	18	5	7	10
		21.85	17.22	15.89	20.53	18.54	21.85
Aleut	51	13	15	6	26	21	23
White	96	21	7	11	3	4	1
		17.01	11.56	10.88	12.24	12.24	12.24
Aleut	51	4	10	5	15	14	17

Cell values are:
number of misclassifications in sample A;
overall misclassification percentage;
number of misclassifications in sample B.

TABLE 7. Pairwise Classification - Cranial Data Set (Right Side)

Population Pair	Sample Size	Method: Tally	Bayes' Theorem	Weight of Evidence	LDF Prop. (Priors)	LDF Equal (Priors)	Trait-Distance
1	31	26	6	5			12
		40.00	16.92	15.38	NP	NP	24.62
2	34	0	5	5			4
1	31	1	3	4	0	0	13
		37.74	16.98	15.09	39.62	39.62	26.42
3	22	19	6	4	21	21	1
1	31	31	12	10	1	1	16
		38.75	22.50	23.75	47.50	47.50	23.75
4	49	0	6	9	37	37	3
1	31	27	14	5	16	13	7
		20.61	13.74	13.74	16.79	16.79	19.08
5	100	0	4	13	6	9	18
1	31	30	13	7	8	7	15
		26.79	14.29	16.96	13.39	13.39	18.75
6	81	0	3	12	7	8	6
2	34	0	3	7			10
		39.29	21.43	25.00	NP	NP	28.57
3	22	22	9	7			6
2	34	31	17	13			8
		38.55	27.71	28.92	NP	NP	32.53
4	49	1	6	11			19
2	34	17	14	7	25	24	2
		16.42	15.67	17.91	22.39	22.39	25.37
5	100	5	7	17	5	6	32
2	34	18	12	7	26	26	4
		16.52	13.04	14.78	28.70	28.70	19.13
6	81	1	3	10	7	7	18
3	22	21	15	9			8
		29.58	22.54	28.17	NP	NP	19.72
4	49	0	1	11			6
3	22	16	18	9	21	21	7
		13.11	16.39	19.67	19.67	19.67	21.31
5	100	0	2	15	3	3	19

TABLE 7 (Contd.)

Population Pair	Sample Size	Method: Tally	Bayes' Theorem	Weight of Evidence	LDF Prop. (Priors)	LDF Equal (Priors)	Trait-Distance
3	22	20	14	7	21	21	7
		19.42	14.56	18.45	25.24	25.24	17.48
6	81	0	1	12	5	5	11
4	49	21	23	15	32	27	3
		22.15	23.49	24.83	30.20	32.21	37.58
5	100	12	12	22	13	21	53
4	49	12	26	13	32	32	28
		26.92	27.69	23.08	33.08	34.62	25.38
6	81	33	10	17	11	13	5
5	100	4	16	20	23	26	80
		35.36	24.86	24.31	25.41	25.97	46.96
6	81	60	29	24	23	21	5

Cell values are:
number of misclassifications in sample A;
overall misclassification percentage;
number of misclassifications in sample B.

TABLE 8. Pairwise Classification - Cranial Data Set (Left Side)

Population Pair	Sample Size	Method: Tally	Bayes' Theorem	Weight of Evidence	LDF Prop. (Priors)	LDF Equal (Priors)	Trait-Distance
1	31	27	5	5	11	11	12
		41.54	18.46	20.00	35.38	35.38	30.77
2	34	0	7	8	12	12	8
1	31	1	4	7	0	2	6
		39.62	16.98	20.75	41.51	41.51	18.87
3	22	20	5	4	22	20	4
1	31	30	14	9	6	6	13
		37.50	23.75	22.50	67.50	67.50	27.50
4	49	0	5	9	48	48	9
1	31	28	11	7	18	17	7
		21.37	12.98	19.85	20.61	22.90	24.43
5	100	0	6	19	9	13	25
1	31	28	15	9	15	15	13
		25.00	17.86	16.96	20.54	20.54	16.07
6	81	0	5	10	8	8	5
2	34	0	1	3			3
		39.29	21.43	23.21	NP	NP	23.21
3	22	22	11	10			10
2	34	33	13	10	5	5	10
		39.76	24.10	25.30	65.06	65.06	32.53
4	49	0	7	11	49	49	17
2	34	26	11	4	29	28	3
		19.40	11.94	15.67	23.13	22.39	30.60
5	100	0	5	17	2	2	38
2	34	17	14	5	24	24	5
		15.65	17.39	16.52	30.43	30.43	25.22
6	81	1	6	14	11	11	24
3	22	22	16	11	7	7	13
		30.99	28.17	26.76	42.25	42.25	28.17
4	49	0	4	8	23	23	7
3	22	22	18	8	21	21	9
		18.03	18.03	16.39	21.31	21.31	18.85
5	100	0	4	12	5	5	14

TABLE 8. (Contd.)

Population Pair	Sample Size	Method: Tally	Bayes' Theorem	Weight of Evidence	LDF Prop. (Priors)	LDF Equal (Priors)	Trait-Distance
3	22	22	16	8	14	14	7
		21.36	18.45	15.53	34.95	34.95	20.39
6	81	0	3	8	22	22	14
4	49	42	17	9	28	27	6
		30.20	18.12	18.79	28.86	20.20	33.56
5	100	3	10	19	15	18	44
4	49	25	19	9	15	14	19
		23.85	20.77	16.15	20.77	21.54	26.15
6	81	6	8	12	12	14	15
5	100	0	23	31	24	25	69
		44.2	30.94	30.94	31.49	30.94	41.99
6	81	80	33	25	33	31	7

Cell values are:
number of misclassifications in sample A;
overall misclassification percentage;
number of misclassifications in sample B.

TABLE 9. Multi-Way Classification - Tally Method

Infracranial

Left

(predicted population)

(actual population)

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	1	32	22	7	0	62	
Yukon	0	37	11	5	0	53	
Black	0	23	61	16	0	100	Error Rate = 57.46%
White	0	24	21	51	0	96	
Aleut	0	20	19	8	4	51	

Right

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	20	10	31	0	1	62	
Yukon	10	22	20	1	0	53	
Black	8	1	85	6	0	100	Error Rate = 57.46%
White	15	7	51	23	0	96	
Aleut	10	11	24	2	4	51	

Cranial

Left

	1	2	3	4	5	6	Sample Size	
1	0	0	0	0	31	0	31	
2	0	5	0	1	28	0	34	
3	0	0	0	1	21	0	22	Error Rate = 66.56%
4	0	0	0	1	48	0	49	
5	0	0	0	0	100	0	100	
6	0	0	0	0	81	0	81	

Right

	1	2	3	4	5	6	Sample Size	
1	0	0	1	0	30	0	31	
2	0	3	1	2	28	0	34	
3	0	1	0	0	21	0	22	Error Rate = 67.19%
4	0	0	0	1	48	0	49	
5	0	0	0	0	100	0	100	
6	0	0	0	0	81	0	81	

Tabled values are the resulting individual classifications.

TABLE 10. Multi-way Classification - Bayes' Theorem

Infracranial

Left

(predicted population)

(actual population)

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	38	7	8	5	4	62	
Yukon	10	23	8	6	6	53	
Black	6	8	61	20	5	100	Error Rate = 41.16%
White	3	4	16	70	3	96	
Aleut	7	9	9	5	21	51	

Right

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	30	7	12	8	5	62	
Yukon	9	25	7	9	3	53	
Black	5	6	64	23	2	100	Error Rate = 39.78%
White	3	5	16	71	1	96	
Aleut	9	8	3	3	28	51	

Cranial

Left

	1	2	3	4	5	6	Sample Size	
1	10	3	1	5	6	6	31	
2	0	19	1	6	4	4	34	
3	0	2	2	4	11	3	22	Error Rate 51.42%
4	1	5	2	17	14	10	49	
5	3	3	1	5	66	22	100	
6	2	5	0	5	29	40	81	

Right

	1	2	3	4	5	6	Sample Size	
1	13	0	1	5	6	6	31	
2	0	14	0	10	8	2	34	
3	1	2	2	0	12	5	22	Error Rate = 46.37%
4	2	4	1	16	15	11	49	
5	0	2	1	7	79	11	100	
6	0	1	1	7	26	46	81	

Tabled values are the resulting individual classifications.

TABLE 11. Multi-way Classification - Weight of Evidence

Infracranial

Left

(predicted population)

(actual population)

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	34	12	4	3	9	62	
Yukon	9	29	5	4	6	53	
Black	9	12	50	21	8	100	Error Rate = 41.16%
White	5	4	14	69	4	96	
Aleut	5	9	4	2	31	51	

Right

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	31	8	5	6	12	62	
Yukon	9	29	4	7	4	53	
Black	7	8	60	22	3	100	Error Rate = 39.78%
White	3	7	15	68	3	96	
Aleut	9	8	1	3	30	51	

Cranial

Left

	1	2	3	4	5	6	Sample Size	
1	16	4	2	5	3	1	31	
2	3	20	2	6	1	2	34	
3	1	2	7	6	3	3	22	Error Rate = 50.79%
4	4	7	4	23	8	3	49	
5	7	9	7	7	51	19	100	
6	6	7	2	5	22	39	81	

Right

	1	2	3	4	5	6	Sample Size	
1	17	1	1	5	3	4	31	
2	2	19	3	6	2	2	34	
3	2	4	6	2	5	3	22	Error Rate = 48.58%
4	3	9	7	16	7	7	49	
5	7	6	6	10	63	8	100	
6	4	3	6	10	16	42	81	

Tabled values are the resulting individual classifications.

TABLE 12. Multi-way Classification - Linear Discriminant Function [Proportional Priors]

Infracranial

Left

(actual population) \ (predicted population)

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	16	16	17	8	5	62	
Yukon	10	24	14	5	0	53	
Black	9	4	67	16	4	100	Error Rate = 46.41%
White	5	2	17	69	3	96	
Aleut	6	11	9	7	18	51	

Right

(actual population) \ (predicted population)

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	20	4	23	13	2	62	
Yukon	14	14	17	7	1	53	
Black	6	2	72	20	0	100	Error Rate = 44.75%
White	0	4	12	79	1	96	
Aleut	14	9	9	4	15	51	

Cranial

Left

	1	2	3	4	5	6	Sample Size	
1	8	0	1	4	10	8	31	
2	2	5	0	7	4	16	34	
3	1	0	1	4	9	7	22	Error Rate = 54.26%
4	0	1	0	21	10	17	49	
5	3	3	2	8	70	14	100	
6	3	2	2	8	26	40	81	

Right

	1	2	3	4	5	6	Sample Size	
1	12	0	1	1	5	12	31	
2	5	3	0	3	8	15	34	
3	1	0	1	3	9	8	22	Error Rate = 53.63%
4	8	1	0	11	8	21	49	
5	1	1	1	11	68	18	100	
6	1	2	2	7	17	52	81	

Tabled values are the resulting individual classifications.

TABLE 13. Multi-way Classification - Linear Discriminant Function [Equal Priors]

Infracranial

Left

(predicted population)

(actual population)

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	19	18	12	7	6	62	
Yukon	13	24	10	5	1	53	
Black	15	8	57	14	6	100	Error Rate = 47.79%
White	5	2	16	70	3	96	
Aleut	7	12	7	6	19	51	

Right

	Coast	Yukon	Black	White	Aleut	Sample Size	
Coast	22	6	21	11	2	62	
Yukon	15	17	15	5	1	53	
Black	7	6	68	19	0	100	Error Rate = 45.58%
White	2	7	10	75	2	96	
Aleut	15	10	7	4	15	51	

Cranial

Left

	1	2	3	4	5	6	Sample Size	
1	8	0	1	6	8	8	31	
2	2	7	1	9	4	11	34	
3	2	0	1	4	8	7	22	Error Rate = 55.21%
4	0	1	1	24	7	16	49	
5	4	4	3	10	65	14	100	
6	5	4	3	9	23	37	81	

Right

	1	2	3	4	5	6	Sample Size	
1	14	1	1	3	3	9	31	
2	8	4	0	4	5	13	34	
3	1	1	1	4	6	9	22	Error Rate = 55.84%
4	10	2	0	14	6	17	49	
5	1	4	2	16	54	23	100	
6	1	2	2	13	10	53	81	

Tabled values are the resulting individual classifications.

BIBLIOGRAPHY

Anderson, J.E. 1968. Skeletal "anomalies" as genetic indicators. In The Skeletal Biology of Earlier Human Populations, ed. by D.R. Brothwell. Oxford: Pergamon Press. pp. 135-148.

Bahadur, R.R. 1961. A representation of the joint distribution of responses to n dichotomous variables. Studies in Item Analysis and Prediction, ed. by H. Solomon. Stanford University Press, California.

Benfer, R.A. 1970. Associations among cranial traits. American Journal of Physical Anthropology, Vol. 32, pp. 463-464.

Berry, A.C. 1976. The anthropological value of minor variants of the dental crown. American Journal of Physical Anthropology, Vol. 45, #2, pp. 257-268.

Berry, A.C. and R.J. Berry. 1967. Epigenetic variation in the human cranium. Journal of Anatomy, Vol. 101, #2, pp. 361-379.

Brothwell, D.R. 1968. Introducing the field. In The Skeletal Biology of Earlier Human Populations, ed. by D.R. Brothwell. Oxford: Pergamon Press, pp. 1-18.

Brown, A. 1971. Classification using dichotomous responses. Ph.D. dissertation, University of Pittsburgh.

Cavalli-Sforza, L.L. and W.F. Bodmer. 1971. The Genetics of Human Populations. San Francisco: Freeman and Co.

Constandse-Westermann, T.S. 1972. Coefficients of Biological Distance. Oosterhout N.B., The Netherlands.

Corruccini, R.S. 1973. Size and shape in similarity coefficients based on metric characters. American Journal of Physical Anthropology, Vol. 38, #3, pp. 743-753.

_____. 1974. An examination of the meaning of cranial discrete traits for human skeletal biological studies. American Journal of Physical Anthropology, Vol. 40, pp. 425-446.

Dahlberg, A.A. 1963. Analysis of the American Indian dentition. In Dental Anthropology, ed. by D.R. Brothwell. Pergamon Press. pp. 149-177.

Deol, M.S. and G.M. Truslove. 1957. Genetical studies on the skeleton of the mouse. XX. Maternal physiology and variation in the skeleton of C57BL mice. Journal of Genetics, Vol. 55, pp. 288-312.

Dillon, W.R. and M. Goldstein. 1978. On the performance of some multinomial classification rules. Journal of the American Statistical Association, Vol. 73, #362, pp. 305-313.

Duda, R.O. and P.E. Hart. 1973. Pattern Classification and Scene Analysis. New York: John Wiley and Sons.

Finnegan, M. 1972. Population definition on the Northwest coast by analysis of discrete character variation. Ph.D. dissertation, University of Colorado.

_____. 1978. A Guide to Osteological Analysis. 2nd edition. Manhattan, Ks.: Kansas State University.

_____. 1978. Non-metric variation of the infracranial skeleton. Journal of Anatomy, Vol. 125, #1, pp. 23-37.

Finnegan, M. and K. Cooprider. 1978. Empirical comparison of distance equations using discrete traits. American Journal of Physical Anthropology, Vol. 49, #1, pp. 39-46.

Finnegan, M. and A. Marcsik. 1979. A non-metric examination of the relationships between osteological remains from Hungary representing populations of Avar period. I and II. Acta Biologica Academiae Scientiarum Hungaricae (in press).

Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics, Vol. 7, pp. 179-188.

Giles, E. and O. Elliot. 1960. Negro-white identification from the skull. Actes du VI^{e} Congres International des Sciences anthropologiques et ethnologiques, Paris.

_____. 1962. Race identification from cranial measurements. Journal of Forensic Sciences, Vol. 7, #2, pp. 147-157.

Glass, B. 1955. On the likelihood of significant admixture of genes from the North American Indians in the present composition of the Negroes of the United States. American Journal of Human Genetics, Vol. 7, #4, pp. 368-385.

Glass, B. and C.C. Li. 1953. The dynamics of racial admixture - an analysis based on the American Negro. The American Journal of Human Genetics, Vol. 5, #1, pp. 1-20.

Green, R.F. and J.M. Suchey. 1976. The use of inverse sine transformations in the analysis of non-metric cranial data. American Journal of Physical Anthropology, Vol. 45, pp. 61-68.

Grewal, M.S. 1962. The rate of genetic divergence of sublines in the C57BL strain of mice. Genet. Res. Camb., Vol. 3, pp. 226-237.

Grüneberg, H. 1951. Evidence for genetic drift in Indian rats. Evolution, Vol. 15, pp. 259-262.

_____. 1952. Genetical studies on the skeleton of the mouse. IV. Quasi-continuous variations. Journal of Genetics, Vol. 51, pp. 95-114.

Hertzog, K.P. 1968. Associations between discontinuous skeletal traits. American Journal of Physical Anthropology, Vol. 29, pp. 397-404.

Hills, M. 1966. Discrimination and allocation with discrete data. Applied Statistics, Vol. 16, pp. 237-250.

Howe, W. and P. Parsons. 1967. Genotype and environment in the determination of minor skeletal variants and body weight in mice. Journal of Embryology and Experimental Morphology, Vol. 17, pp. 283-292.

_____. 1968. Morphogenetic homeostasis in man. Acta genetica, Vol. 18, pp. 341-348.

Jardine, N. and R. Sibson. 1971. Mathematical Taxonomy. London: John Wiley and Sons, Ltd.

Korey, K. 1970. Characteristics of the distribution of non-metric variants. M.A. thesis, University of Chicago.

Krogman, W.M. 1962. The Human Skeleton in Forensic Medicine. Springfield, Ill.: Charles C. Thomas.

Krzanowski, W.J. 1975. Discrimination and classification using both binary and continuous variables. Journal of the American Statistical Association, Vol. 70, pp. 782-790.

Kullback, S. 1959. Information Theory and Statistics. New York: John Wiley and Sons.

Larnach, S.L. and N.W.G. MacIntosh. 1966. The craniology of the Aborigines of coastal New South Wales. Oceania Monographs, No. 13.

_____. 1967. The use in forensic medicine of an anthropological method for the determination of sex and race in skeletons. Archaeology and Physical Anthropology in Oceania, Vol. 2, #2.

Laughlin, W.S. and J. Jørgensen. 1956. Isolate variation in Greenland Eskimo crania. Acta Genetica et Statistica Medica, Vol. 6, pp. 3-12.

LeDouble, A.F. 1903. Traite des variations des os du crane et leur signification au point de vue de l'anthropologique zoologique. Paris, Vigot.

LeDouble, A.F. 1906. Traite des variations des os de la face et leur signification au point de vue de l'anthropologique zoologique. Paris, Vigot.

Mahalanobis, P.C. 1936. On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India, Vol. 12, pp. 49-55.

Moore, D.H. II. 1973. Evaluation of five discrimination procedures for binary variables. Journal of the American Statistical Association, Vol. 68, pp. 399-404.

Morrison, D.F. 1976. Multivariate Statistical Methods. 2nd edition. New York: McGraw-Hill.

Nie, N.H., C.H. Hull, J.G. Jenkins, K. Steinbrenner, D.H. Brent. 1975. Statistical Package for the Social Sciences (SPSS). 2nd edition. New York: McGraw-Hill.

Osteyee, D.B. and I.G. Good. 1974. Information, weight of evidence, the singularity between probability measures and signal detection. Lecture Notes in Mathematics #376. Berlin: Springer-Verlag.

Pearson, K. 1926. On the coefficient of racial likeness. Biometrika, Vol. 18, pp. 105-117.

Penrose, L.S. 1954. Distance, size, and shape. Annals of Eugenics, Vol. 18, pp. 337-343.

Roberts, D.F. 1955. The dynamics of racial intermixture in the American Negro - Some anthropological considerations. American Journal of Human Genetics, Vol. 7, #4, pp. 361-367.

Russell, F. 1900. Studies in cranial variation. The American Naturalist, Vol. 34, #405, pp. 737-745.

Schulter, F. and M. Finnegan. Racial distance: a multivariate analysis of roentgenographic measurements in Eskimos, Indians, and Whites. HOMO, Vol. 28, 1977, pp. 227-235.

Searle, A.G. 1954. Genetical studies on the skeleton of the mouse. IX. Causes of skeletal variation within pure lines. Journal of Genetics, Vol. 52, pp. 68-102.

Sjøvold, T. 1973. The occurrence of minor, non-metrical variants in the skeleton and their quantitative treatment for population comparisons. HOMO, Vol. 24, pp. 204-233.

_____. 1977. Non-metrical divergence between skeletal populations. Ossa, Vol. 4, supplement 1.

Sokal, R.R. and P.H.A. Sneath. 1963. Principles of Numerical Taxonomy. San Francisco and London: W. H. Freeman and Co.

Suchey, J.M. 1975. Biological distance of prehistoric central California populations derived from non-metric traits of the cranium. Ph.D. dissertation, University of California-Riverside.

Sullivan, L.R. 1922. The frequency and distribution of some anatomical variations in American crania. Anthropological Papers of the American Museum of Natural History, Vol. 23, part 5, pp. 205-258.

Todd, T.W. and B. Tracy. 1930. Racial features in the American Negro cranium. American Journal of Physical Anthropology, Vol. 15, #1, pp. 53-110.

Truslove, G.M. 1961. Genetical studies on the skeleton of the mouse. XXX. A search for correlations between some minor variants. Genet. Res. Camb., Vol. 2, pp. 431-438.

Wood-Jones, F. 1931. The non-metrical morphological characters of the skull as criteria for racial diagnosis. Journal of Anatomy, Vol. 65, Parts 1,2,3, pp. 179-195, 368-378, 438-445.

_____. 1934. The non-metrical morphological characters of the skull as criteria for racial diagnosis. Journal of Anatomy, Vol. 68, pp. 96-108.

RACIAL CLASSIFICATION BASED ON NON-METRIC SKELETAL TRAITS

by

KEVIN B. COOPRIDER

B.A., Kansas State University, 1975

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1978

Osteological material is often the sole remnant of the human cadaver in archaeological situations and modern forensic cases. Information obtained from this material can determine characteristics of the associated individual with varying degrees of accuracy. Non-metric skeletal traits, long studied as possible clues to genetic relationships among populations, were used to empirically test several techniques proposed to classify an individual to his correct racial group. Assumptions concerning parameter estimation, intertrait correlation, and pooling of age and sex categories were examined in relation to the implementation of the procedures. Discriminant functions were derived from collected individuals of documented racial origin. The same individuals were subsequently classified by these functions to obtain misclassification probability estimates.

Two data sets were employed for this test. The first was composed of six Hungarian groups scored for forty-two cranial traits, and the second consisted of White, Black, and Amerindian collections scored for thirty infra-cranial traits. Results were obtained for six techniques in pairwise population classifications. Included were a technique based on Bayes' theorem, a weight of evidence procedure (similar to the Bayes technique but ignoring the prior membership probabilities), the linear discriminant function (involving equal and proportional priors), a tally method, and a procedure derived from information theory. The former two methods proved generally superior. These same two techniques yielded the most effective classifications for multi-population comparisons as well.

Recommendations for improved discrimination results stress accumulation of existing data and standardization of scoring techniques.

Furthermore, accommodation of the high correlation between opposing sides of bilateral traits should decrease current misclassification rates.