Selection of Variables for Discriminant Analysis of Human Crania for Determining Ancestry

Adam Kolatorowicz, MS -- University of Indianapolis, 1400 East Hanna Avenue, Indianapolis, IN

ABSTRACT

Forensic anthropologists use the computer program FORDISC 2.0 (FD2) as an analytical tool for the determination of ancestry of unknown individuals. There is an almost endless number of measurements that can be taken on the human skeleton, yet FORDISC includes only 78 measurements for its analysis. In particular, the program will only utilize up to 24 measurements of the cranium. These 24 cranial variables are used because they require simple, relatively inexpensive instruments that most biological anthropology laboratories have (spreading and sliding calipers). Also, individuals with a basic knowledge of the anatomical landmarks can take the measurements with relative ease. Unconventional measurements of the cranium require unusual, costly instruments (such as the radiometer and coordinate caliper) and are more difficult to take. This poster will examine which measurements of the human cranium provide the greatest classificatory power when constructing discriminant function formulae for the determination of ancestry and will answer the question of whether the use of variables that require more time, training, and equipment are worth the effort.

Sixty five cranial measurement were taken on 155 adult human crania from three different ancestral groups: (1) African American (n=50), (2) European American (n=50), and (3) Coyotero Apache (n=55). The 65 measurements were broken up into four subsets for statistical analysis: (1) FD2 (1996), (2) Howells (1973), (3) Gill (1984), and (4) All Measurements. A predictive discriminant analysis with a forward stepwise methodology of p = 0.05 to enter and p = 0.15 to remove was run using the statistics computer software package SPSS 13.0. The analysis produced 4 sets of discriminant function formulae. The classificatory power of each set of formulae was determined by comparing the hit-rate estimation (the percent correctly classified) of each of the subsets. First, the resubstitution rate was compared to the leave-one-out (LOO) rate for each subset and then both rates were compared across all subsets. The FD2 subset had a resubstitution rate of 90.3% and LOO rate of 85.8%. The Howells subset had a resubstitution rate of 92.9% and a LOO rate of 90.3%. The Gill subset had a resubstitution rate of 63.2% and a LOO rate of 61.9%. Finally, the All Measurements subset had a resubstitution rate of 95.5% and a LOO rate of 93.5%. The non-standard measurements of the All subset performed the best and the standard FD2 measurements performed third best. Non-standard measurements incorporated in the All formulae included frontal subtense, mid-orbital breadth, bistephanic breadth, bimaxillary breadth, and molar alveolar radius.

The formulae provided the best separation of the Apache group from the other two groups. Stepwise analysis showed that the use of more variables is not necessarily better. Not all of the variables were included in the final formulae. Only 12 of the 24 FD2 measurements, 12 of the 57 Howells measurements, 4 of the 6 Gill measurements, and 15 of the 65 All Measurements were used. Results show that the non-standard measurements can be useful for determining the ancestry of unknown human crania. These measurements could be especially useful for incomplete crania. It is suggested that biological anthropology laboratories purchase radiometers and coordinate calipers to record data that would be missed with spreading and sliding calipers. Standard measurements can be combined with non-standard measurements to produce more powerful discriminant function formulae for the prediction of ancestry.

INTRODUCTION

The term "ancestry" as used by biological anthropologists refers to the population affiliation, based on geographical location, of an individual. The determination of the ancestry of human skeletal remains is a key step in the course of identifying unknown individuals. The predicted ancestry, sex, age, and stature form a biological profile of the decedent which, in turn, helps with the identification process by narrowing down the number of potential matches.

The biological profile is also used by anthropologists to understand human variation in historic and prehistoric populations. Bioarcheologists analyze the skeletons of long-deceased individuals to better understand the history of human populations, including migration patterns, demographic structure, and the effects of disease. Creating an accurate biological profile, whether for recent or ancient remains, is the starting point of all future analysis. Therefore, it is imperative to refine existing methods for determining ancestry and sex and for estimating age and stature in order to improve the accuracy of the biological profile, and, by extension, conclusions based on those profiles.

The purpose of this study is to determine which measurements of the human cranium provide the greatest classificatory power when constructing discriminant function formulae for the determination of ancestry. The study will identify the most effective combinations of standard and non-standard measurements and will answer the question of whether the use of variables that require more time, training, and equipment are worth the effort. For the purposes of this study, "standard" measurements are defined as the twenty-four cranial measurements used in a FORDISC analysis, which require the use of conventional instrumentation. "Non-standard" measurements are those not included in a FORDISC analysis and are less frequently used because they require special instruments to record. Perhaps biological anthropology laboratories should invest in other instrumentation besides spreading and sliding calipers. It is hypothesized that a combination of standard and non-standard measurements will provide higher classificatory power over only standard measurements.

MATERIALS AND METHODS

The Study Sample. This study examined three different ancestral groups from different time periods: recent African Americans (AA), recent European Americans (EA), and prehistoric Coyotero Apache (CA). The Coyotero Apache crania were drawn from the Indiana University Anthropology Department Osteological Collection in Bloomington, IN. The series of crania used in this study are from the Edward Palmer Arkansaw Mounds located just outside of Little Rock, AR. It is a Late Woodland to Mississippian site dating from A.D. 700-950. The African American and European American samples were drawn from the Hamman-Todd Osteological Collection located at the Cleveland Museum of Natural History in Cleveland, OH. The specimens come from the Cuyahoga County Morgue and city hospitals and date from the late 1800s and early 1900's. Table 1 shows the distribution of the sample by collection, age, and sex.

TABLE 1. Distribution of Individuals by Collection, Group, Sex, and Age.

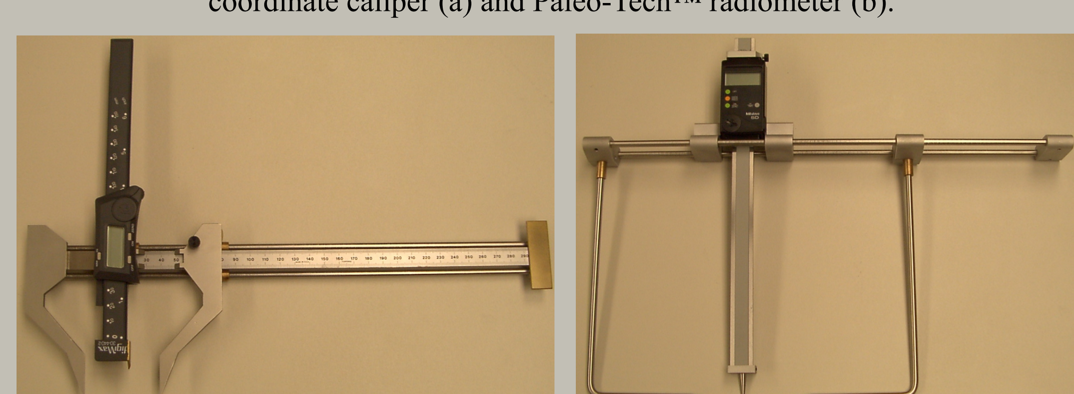
Collection	IU		CMNH			
Group	Coyotero Apache		African American		European American	
Sex	M	F	M	F	M	F
Count	19	36	25	24	26	25
Min. Age	NA		21		19	
Max. Age	NA		46		48	
Mean Age	NA		28.9		33.7	

Taking the Measurements and Instrumentation. A battery of 65 measurements was taken on each cranium. All measurements were recorded to the nearest millimeter and on the left side of the skull in the case of bilateral measurements. If the left side was damaged to the degree that a measurement could not be taken, then the right side was used. If an area was damaged or resorbed by more than 2 mm, then the measurement was taken and recorded on the form in parentheses. If more than 2 mm of bone was missing, the measurement was not taken at all.

The measurements were derived from three sources: (1) FORDISC 2.0 (FD2) (Ousley and Jantz, 1996), (2) Howells (1973), and (3) Gill (1984). The 24 FD2 measurements are those used in a standard FD2 analysis of the cranium.

A Paleo-Tech™ spreading caliper, Mitutoyo™ sliding dial caliper, Paleo-Tech™ PaleoCal-1 coordinate caliper, and a Paleo-Tech™ radiometer were used to take all measurements of the crania from IU. At the CMNH, a Mitutoyo™ direct input tool and foot pedal were used in conjunction with a Mitutoyo™ sliding digital caliper. This caliper was connected directly to a laptop computer so the measurements could be input directly into a Microsoft™ Office Excel 2003 spreadsheet. Figures 1a and 1b show the radiometer and coordinate caliper in use.

FIGURE 1. Non-standard Instrumentation. Paleo-Tech™ PaleoCal-1 coordinate caliper (a) and Paleo-Tech™ radiometer (b).



(a) (b)

Statistical Analysis. Four sets of discriminant function formulae were constructed with a forward stepwise method and unequal prior probabilities using the statistics computer software package *SPSS 13.0*. Predictive discriminant analysis requires that all specimens have all measurements or else they will not be included in the calculation of the functions. Measurements that could not be taken were substituted with the 'linear trend at point' function of *SPSS*. This function replaces missing values with the linear regression estimate for that point. Only 12 (0.001%) of the 10,075 measurements taken were missing and subsequently replaced.
The stepwise selection method followed İşcan and Steyn's (1999) data entry methodology, with p values of p = 0.05 to enter and p = 0.15 to remove.

The crania were divided into three different ancestry groups: (1) African were entered four times to produce four sets of discriminant function formulae for a total of eight functions. The first set of functions used the 24 standard measurements of the cranium included in a FD2 analysis. The second set used the 57 Howells measurements. The third set of formulae used Gill's 6 measurements of the nasal region. The fourth and final set combined all 65 different measurements. The weights of the variables for each function were calculated. Eigenvalues and Wilk's λ values were calculated to test the significance of the discriminant functions. Structure matrices were constructed to determine how heavily the variables loaded on the functions.

The four sets of discriminant function formulae were compared side by side to determine which set was better able to correctly separate groups and sort the individuals. The ability of the discriminant functions to correctly classify individuals was assessed by testing the functions on the original sample via simple resubstitution as well as a leave-one-out method (LOO). Resubstitution rates were used only as a general baseline measure of the formulae's performance. LOO more accurately tests the functions ensuring that the cases used to produce the functions are not used when testing them for classificatory ability. The success of the discriminant functions is expressed as a percentage of cases in the original sample that are correctly classified into their groups.

RESULTS

TABLE 2. Variables Included Via Stepwise Selection. Variables are listed in the order they were selected.. An * indicates a non-standard measurement and a [1,2,3,4] indicates the most heavily weighted variables.

FD2
parietal chord
nasal breadth
basion-bregma height[4]
max. alveolar length
orbital height
max. cranial breadth
max. cranial length[1]
biauricular breadth[3]
min. frontal breadth[2]
basion-prosthion length
foramen magnum breadth
bizygomatic breadth

Howells
frontal subtense*
bimaxillary breadth*[1]
prosthion radius*
zygoorbitale radius*[3]
parietal chord
bistephanic breadth*
bifrontal breadth*[2]
cranial base length[4]
orbital height
biasterionic breadth*
nasal breadth
subspinale radius*

Gill
mid-orbital breadth*[1]
maxillofrontal breadth[2]
alpha chord*[3]
naso-zygoorbital breadth*[4]

All Measurements
frontal subtense*[3]
mid-orbital breadth*
parietal chord
bistephanic breadth*
bimaxillary breadth*[2]
bifrontal breadth*[1]
foramen magnum breadth
orbital height
molar alveolar radius*
naso-zygoorbital subtense*
nasal breadth
biasterionic breadth*
bizygomatic breadth[4]
occipital fraction*
basion-bregma height

FIGURE 2. Canonical Map: FD2 Measurements.

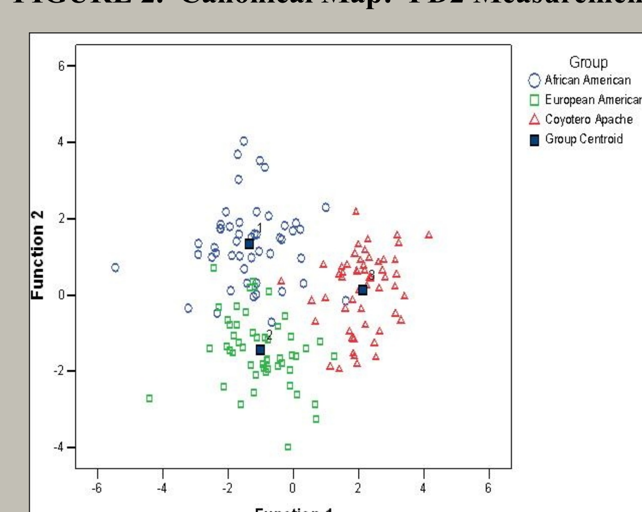


FIGURE 3. Canonical Map: Howells Measurements.

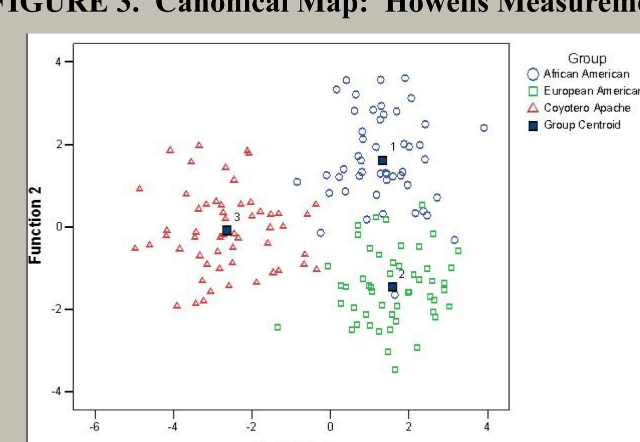


FIGURE 2. Canonical Map: Gill Measurements.

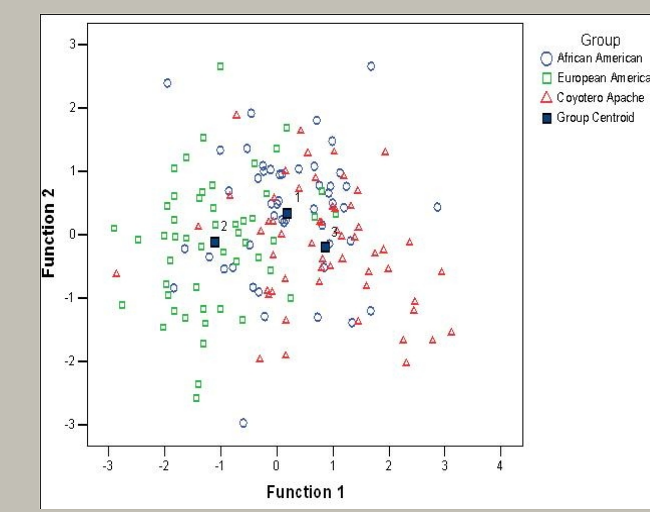


FIGURE 2. Canonical Map: All Measurements.

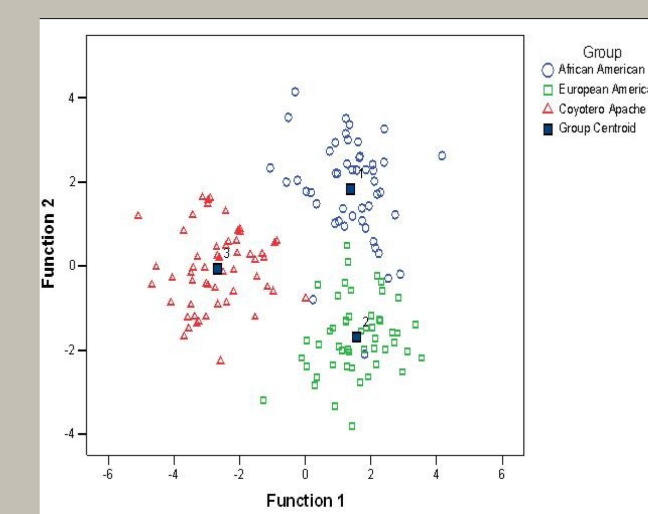


TABLE 3. Average Classification Accuracy of Measurement Sets.

Measurement Set	Resubstitution	Leave-One-Out
FD2	90.3%	85.8%
Howells	92.9%	90.3%
Gill	63.2%	61.9%
All Measurements	95.5%	93.5%

TABLE 4. Classification Accuracy of Discriminant Functions Via Leave-One-Out. All values are in %. Red values indicate percent correctly classified.

Measurement Set	Actual Group	Predicted Group		
		AA	EA	CA
FD2	AA	81.6	12.2	6.1
	EA	9.8	86.3	3.9
	CA	1.8	9.1	89.1
Howells	AA	87.8	10.2	2.0
	EA	7.8	90.2	2.0
	CA	3.6	3.6	92.7
Gill	AA	42.9	26.5	30.6
	EA	15.7	76.5	7.8
	CA	23.6	10.9	65.5
All Measurements	AA	89.8	8.2	2.0
	EA	3.9	94.1	2.0
	CA	1.8	1.8	96.4

DISCUSSION

The use of a forward stepwise variable selection method shows that a large number of variables is not necessary when constructing discriminant functions. Only 12 of the 24 FD2 variables, 12 of the 57 Howells variables, 4 of the 6 Gill variables, and 15 of the 65 All Measurements variables were included in the final analysis (Table 2). For the Howells, Gill, and All Measurements sets, the stepwise method selected combinations of standard and non-standard measurements to classify the crania. It is proposed here that all discriminant analyses employ a stepwise process for variable selection so as to only include statistically significant variables and eliminate any 'statistical noise' resulting from superfluous variables.

The success of the discriminant functions is assessed by their classification accuracy (Tables 3 & 4). The set with the lowest accuracy is the Gill set, followed by the standard measurements of the FD2 set. The second best set was the mixed measurements of the Howells set, and the set with the highest classification accuracy is the mixed measurements of the All Measurements set. The combination of standard and non-standard measurements increased the classificatory power of the linear discriminant function formulae; therefore, the original hypothesis is accepted.

Taking a closer look at the performance of each measurement set in predicting group membership of the crania via the leave-one-out method, patterns within the discriminant functions emerge. The FD2, Howells, and All Measurements functions performed best with the Coyotero Apache (CA) crania, with 89.1%, 92.7%, and 96.4% correct classification, respectively. The separation of the CA from the European (EA) and African Americans (AA) can be seen in the scatterplots of the discriminant function scores (Figures 2, 3, 4, & 5).

Surprisingly, the Gill set performed best on the EA crania, with 76.5% correct classification. Gill originally used the six measurements of the nasal region to differentiate Native American crania from other groups, yet only 65.5% of the CA crania were correctly classified. This result is most likely due to the fact that Gill specifically used Northwest Plains Indian crania to develop the technique, while the current study examined Native American crania from the Midwest Plains. All four sets of functions had a difficult time classifying the African American crania, with consistently lower hit rates compared to the other groups. Perhaps the AA subsample is more variable in cranial form than either the EA or CA samples. This greater variability can be seen by examining the spread of the AA specimens on the discriminant function score canonical maps. The AA group is the least tightly clustered group of the three, with overlap into the EA and CA clusters.

CONCLUSIONS

Twelve non-standard measurements were selected in this study as being important discriminating variables. These include frontal subtense (FRS), bimaxillary breadth (ZMB), prosthion radius (PRR), zygoorbitale radius (ZOR), bistephanic breadth (STB), bifrontal breadth (FMB), biasterionic breadth (ASB), subspinale radius (SSR), mid-orbital breadth (ZOB), alpha chord (ALC), naso-zygoorbital subtense (NZS), and molar alveolus radius (AVR). FRS, NZS, ZMB, FMB, ZOB, and ALC were taken with the coordinate calipers. PRR, ZOR, SSR, and AVR are all radial measurements that required the use of the radiometer. STB and ASB were taken using a sliding caliper. Although ZMB, FMB, ZOB, and ALC were taken with the coordinate calipers, one can take these four measurements using a sliding caliper. The coordinate caliper was used in this study because a subtense was measured immediately after and directly from the breadth/chord measurement. This fact demonstrates the diversity of the coordinate caliper. It is possible to take up to three measurements at once: chords/breadths, subtenses, and fractions. With regards to time constraints and data collection, it would take longer to switch instruments for each measurement instead of using one instrument to take them all.

The twelve non-standard measurements provide a detailed description of the cranium that is not achieved with standard measurements. ZMB, ZOR, FMB, ZOB, ALC, and NZS are all measures of the face. It is not surprising that these measurements were selected because the facial skeleton tells researchers the most about the population affiliation of an individual. PRR, SSR, and AVR all measure the projection of the maxilla. These three more fully describe which portions of the maxilla project the farthest. FRS, STB, and ASB measure aspects of the vault and add more detail to the overall shape of the vault. Although standard measurements are able to discriminate between groups with a relatively high level (85.8%) of accuracy by themselves, the inclusion of non-standard measurements increases the accuracy level (93.5%) of the functions by ferreting out more subtle shape and size differences.

It is suggested here that biological anthropology laboratories should purchase instrumentation such as coordinate calipers and radiometers so that non-standard measurements can be taken to record data that would be missed by just using spreading and sliding calipers.

REFERENCES

Gill GW (1984) A forensic test case for a new method of geographical race determination. In TA Rathbun and JE Buikstra (eds). Human Identification: Case Studies in Forensic Anthropology. Springfield: Charles C. Thomas, pp. 229-339.

İşcan MY, Steyn M (1999) Craniometric determination of population affinity in South Africans. Int. J. Legal Med. 112:91-97.

Howells WW (1973) Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference Among Recent Human Populations. Papers of the Peabody Museum of Archaeology and Ethnology, Harvard University, vol. 67.

Ousley SD, Jantz RL (1996) FORDISC 2.0: Personal Computer Forensic Discriminant Functions. Knoxville: University of Tennessee.

ACKNOWLEDGMENTS

- *The Connective Tissue* Graduate Student Research Award (2004-007)
- Lyman Jellema, Cleveland Museum of Natural History
- Dr. Della C. Cook, Indiana University
- Noel D. Justice, Glenn A. Black Laboratory of Archaeology, Indiana University
- Dr. Stephen P. Nawrocki, University of Indianapolis