Discriminant Function Analysis

Discriminant function analysis (DFA) is a statistical procedure that classifies unknown individuals and the probability of their classification into a certain group (such as sex or ancestry group).

From: Research Methods in Human Skeletal Biology, 2013

Related terms:

Progressive Supranuclear Palsy, Positron Emission Tomography, Attention Deficit Hyperactivity Disorder, Parkinson's Disease, Pervasive Developmental Disorder, Ultrasonic Vocalization, Biomarker

TRIAL DESIGN, MEASUREMENT, AND ANALYSIS OF CLINICAL INVESTIGATIONS

Hermine I. Brunner, Edward H. Giannini, in Textbook of Pediatric Rheumatology (Sixth Edition), 2011

Discriminant Function Analysis

Discriminant function analysis is MANOVA reversed. In MANOVA, the independent variables are the groups, and the dependent variables are the predictors. In discriminant function analysis, the independent variables are the predictors, and the dependent variables are the groups. As previously mentioned, discriminant function analysis is usually used to predict membership in naturally occurring groups. It answers the question: Can a combination of variables be used to predict group membership? Usually, several variables are included in a study to see which ones contribute to the discrimination between groups. Discriminant function analysis is broken into a two-step process: One first performs the multivariate test, and, if statistically significant, proceeds to see which of the variables have significantly different means across the groups.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B978141606581410007X

Assessing Performance Validity with the ACS

James A. Holdnack, ... Grant L. Iverson, in WAIS-IV, WMS-IV, and ACS, 2013

Discriminant Function Analysis

Discriminant Function Analysis (DFA) has been used extensively in the past to derive optimal combinations of variables to differentiate groups because of its computational simplicity. However, DFA assumes that the predictors (i.e., tests included in the model) are each normally distributed and the set of predictors has a multivariate normal distribution along with homogeneous variance-covariance matrices (Harrell, 2001). These are strong statistical assumptions that are rarely met in clinical research and with performance validity measures in particular. In addition, there is no consensus regarding whether one should use the discriminant weights (standardized coefficients) or discriminant loadings (structure correlations) when interpreting the DFA model. Finally, DFA has no inferential tests for the individual predictors to determine which are statistically reliable in differentiating the groups.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9780123869340000079

Sex Estimation and Assessment

Megan K. Moore, in Research Methods in Human Skeletal Biology, 2013

Discriminant Function Analysis

Discriminant function analysis (DFA) is a statistical procedure that classifies unknown individuals and the probability of their classification into a certain group (such as sex or ancestry group). Discriminant function analysis makes the assumption that the sample is normally distributed for the trait. The **posterior probability** and **typicality probability** are applied to calculate the classification probabilities (Albanese et al., 2008).[7] The posterior probability is the probability that an unknown case belongs to a certain group based on relative **Mahalanobis' distances** measuring the distance to the center or **centroid** of each group. The typicality probability is how *likely* the unknown case belongs to a group based on variability within all groups. The discriminant function procedure has been programmed into most standard statistical packages for greater applicability.

Not all skeletal measurements are equally effective for sex estimation using DFA and the skill of the researcher plays an important role; practice and exposure to population variation are still crucial. Adams and Byrd (2002) compared 13 different measurements taken by 68 researchers. They discovered high interobserver variability in all measurements by researchers with less than 5 years of experience in osteometrics, but there was no significant improvement *after* 5 years of experience. Clearly, there is some level of subjectivity even within metric sex estimation, requiring some training by the researcher. This can be compared with the many years of training and experience necessary to become familiar with visual sexual dimorphism within a single population, especially to accurately assess sex from the cranial morphology (Buikstra and Ubelaker, 1994). Krogman explains the dichotomy between metric analysis versus descriptive analysis as "experience versus statistical standardization" (Krogman and İşcan, 1986). The merits of metric sex estimation can be summarized with the famous quotation by Sir William Thomson Kelvin (1824–1907): "To measure is to know."

URL: https://www.sciencedirect.com/science/article/pii/B9780123851895000042

Forensic Classification and Biodistance in the 21st Century

S.D. Ousley, in Biological Distance Analysis, 2016

Introduction

Linear discriminant function analysis (LDFA), the first multivariate statistical classification method, was invented by R.A. Fisher in 1936. LDFA is predominantly used in bioarchaeology and biological anthropology to assess biodistance (relationships) among groups (called descriptive discriminant analysis or DDA) and in forensic anthropology to classify unidentified individuals (called predictive discriminant analysis or PDA). Both of these approaches to group relationships are provided in LDFA, because individual classifications are based on a cohesive and limited model of intragroup variation and distances based on them. LDFA is a member of the family of classification methods known as supervised learning, because group membership of individuals is used during the process of maximizing group differences. Overviews of LDFA are available in Affifi and Clark (1997), Hand (1997), Huberty (1994), Huberty and Olejnik (2006), Klecka (1980), and Tabachnik and Fidel (2007). LDFA specific to forensic anthropology is outlined in Ousley and Jantz (2012).

The LDFA model utilizes group-specific means and a pooled within-group variance–covariance matrix (VCM), representing the relationships among measurements within groups, to calculate multivariate distances among groups. In this paper, which is centered on craniometrics, "measurements" will be used often when measurements are meant, and "variables" will be used when any kind of

variables can be referred to, which may be measurements, binary, ordinal, or nominal observations, or transformations of measurements; in most cases the two terms are interchangeable. The measurement means for each group and the pooled VCM are calculated directly from the data, and the pooled VCM is the average VCM from each group weighted by sample size. Using matrix multiplication, vectors of differences in group means are multiplied by the inverse of the VCM to combine and scale all differences into one number, the Mahalanobis distance. If all Mahalanobis distances are significantly greater than zero, then the distances reflect all variation in the measurements in the most objective way. Canonical variates analysis can be used to compare groups using two- or three-dimensional plots that show the most important axes that separate groups, and dendrograms can be constructed from the distances that illustrate complex group relationships (Albrecht, 1980; Affifi and Clark, 1997). These graphical methods are especially illuminating in biodistance studies.

There are important assumptions of LDFA–in other words, requirements–in order for the Mahalanobis distances to best represent intergroup relationships. LDFA requires multivariate normality of each group's data, which is usually limited to continuous measurements. As it turns out, interpreting multivariate normality tests is not straightforward, because they depend on sample sizes, the number of measurements analyzed, and other factors. Multivariate normality is simply assumed when each measurement is normally distributed; a more basic requirement is that the discriminant scores themselves are normally distributed within groups. Of course, the normal distribution is also a model, and in fact is based on an infinite sample size, and small deviations from multivariate normality do not affect LDFA accuracy very much (Huberty, 1994). Outliers often affect LDFA models more than normality per se, and there are methods to detect and quantify outliers using the LDFA model and typicality probabilities (Ousley and Jantz, 2012). Additionally, osteometric data appear to be normally distributed as long as sample sizes are large enough and outliers are removed, which is true for nearly any other metric data from animals or plants. Another requirement of LDFA is that variation in each group is more or less the same, and has been described as being “in the same ballpark” (Huberty and Olejnik, 2006, p. 83). As with normality, testing for the homogeneity of VCMs is not especially straightforward. For example, Box's *M* test for homogeneity is notoriously oversensitive to slight heterogeneity and to small departures from normality (Huberty and Olejnik, 2006; Tabachnik and Fidel, 2007). Importantly, small deviations from homogeneity do not seem to affect the estimated relationships or classification accuracy. Classification accuracy is central to evaluating any classification function, because with a potentially large number of measurements available, many unique combinations of measurements could be analyzed, and some combinations will be more accurate than others. As it turns out, estimating classification accuracy is not as straightforward as it may seem.

Descriptive Statistics and Analysis in Biochemical Ecotoxicology

François Gagné, in Biochemical Ecotoxicology, 2014

12.4.2 Discriminant Function Analysis

Discriminant function analysis is similar to multivariate ANOVA but indicates how well the treatment groups or study sites differ with each other. Discriminant analysis seeks out a linear combination of biomarker data for each treatment group that maximizes the difference between treatment groups or study sites for proper classification. The analysis provides a classification function that determines to which groups an individual belongs:

$$s_i = c_i + w_{i1}{}^*x_1 + w_{i2}{}^*x_2 + \cdots + w_{im}{}^*x_m$$

where *i* represents the respective groups or sites and numbers 1,2, ..., *m* represent the *m* biomarkers (variables). c_i is the constant for group *i* and w_{ij} is the ponderation factor of variable (biomarkers) *j* for group *i*. S_i is the classification

value. Different sites will give classification efficiency greater than 70% while similar sites would give undistinguishable results (<70% classification). The analysis provides a classification matrix that shows how well each individual or treatment group replicate is associated to a given site or treatment group. The results for discriminant function analysis are usually expressed by the Wilks' lambda statistic and displayed by a 2D plot of the two best discriminant functions in respect to site classification (Figure 12.8). The analysis also provides factorial analysis to identify which biomarkers are strongly correlated with the discriminant functions 1 and 2. In the example, 11 biomarkers were examined in six sites (groups 1–6) in local mussel populations. In this analysis, each data point represents the root function of the combined 11 biomarkers for each individual. The data points are also displayed for each treatment group or site of collection. In the legend, the classification results are presented in percentage correctness. For example, group or site 2 was correctly classified at 100% but sites 3 and 5 were misclassified as site 2. This indicates that sites 2, 3, and 5 could be considered a similar "composite site," or these sites could be considered similar in their ecotoxicological properties. Group 6 was correctly classified at 100% although close to site/group 1. The biomarkers that had the highest correlation with the root function are listed in parentheses on the *x*- and *y*-axes. This type of analysis is often used in ecotoxicology because it provides a great deal of information for data mining and analysis. Indeed, it provides information on how sites differ with each other (if sites are similar in nature or not) and which biomarkers contribute to site classification/discrimination.

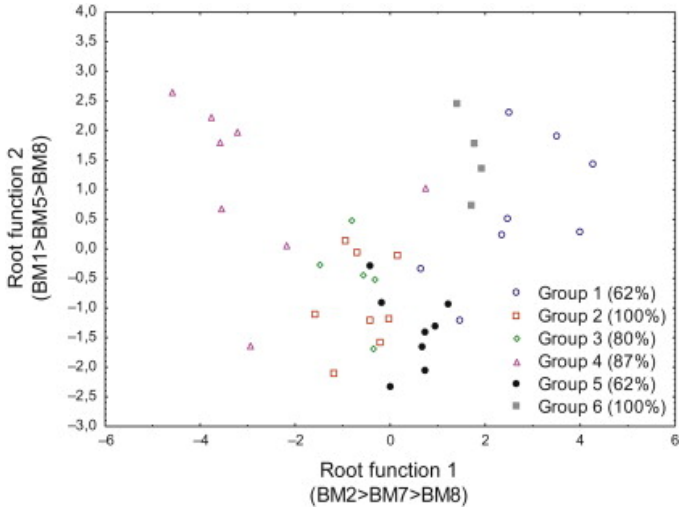

Figure 12.8. Example of discriminant function analysis for site classification. Eleven biomarkers (BM) were determined in six groups (sites or treatments) and analyzed by discriminant function analysis. Each data point corresponds to each replicate individual in a group. The percentage values of groups 1–6 represent the classification correctness. Groups that overlap with each other are considered similar, while distinct groups are considered different. BMs that were the most strongly correlated with the root functions 1 and 2 are mentioned on the *x*- and *y*-axis.

In discriminant function analysis, the factorial scores could be calculated for each observation, which is a linear combination of the standardized biomarker measured for each individual (Figure 12.9). It provides an integrated metric that reflects the changes in the 11 biomarkers tested. The larger the factorial weights (absolute value) the greater the contribution of the biomarker to the root functions 1 and 2.

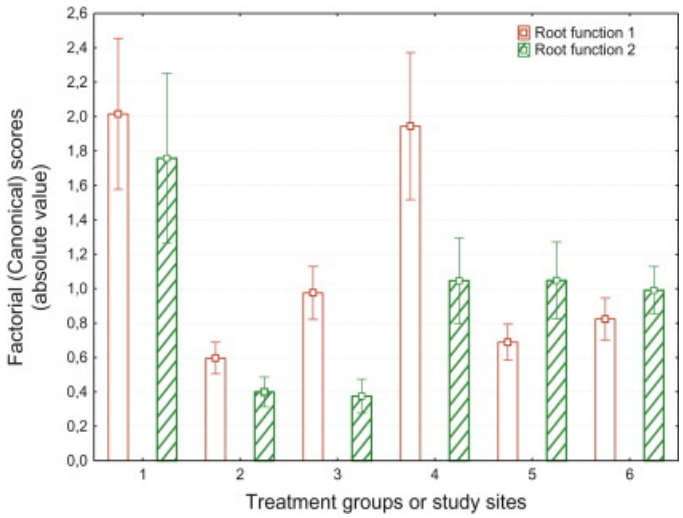

Figure 12.9. Factorial scores of the biomarker data. These data are expressed as the mean of the absolute scores with the SE.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B978012411604700012X

A Biodistance Analysis of Mandibles From Taiwan, Asia, and the Pacific

M. Pietrusewsky, ... M.T. Douglas, in Biological Distance Analysis, 2016

Stepwise Discriminant Function Analysis

Applying stepwise discriminant function analysis to nine measurements recorded in 267 male mandibles indicates that mandibular length, bigonial breadth, and bicanine breadth contribute most to the discrimination in this analysis.

Eigenvalues (values not shown), which represent the amount of variance accounted for by each function or variate, expressed as the percentage of total dispersion, provide an indication of the proportion of dispersion accounted for by each canonical variate. In this analysis, the first three canonical variates account for 70% of the total variation. The first five eigenvalues are significant at the 1% level, indicating significant heterogeneity for these canonical variates.

The jackknifed classification results (not shown) indicate that the Namu, Watom Island, New Zealand, and Sigatoka are among the series with the best classification results, with more than 45% of the cases correctly assigned to their original group. None of the seven mandibles from NKLE is correctly classified; two are reclassified as Korea and one each is classified as Pohnpei, Namu, Ryukyu, Kanto, and Siberia. The majority of the misclassifications of NKLE are to Asian mandibular series. Overall, six mandibles from Asia and two from Micronesia are reclassified as NKLE. Surprisingly, four of the SSH mandibles are reclassified to Hawaii, two are assigned to New Zealand, and one each to Northern Mariana and Namu series representing the Pacific Islands. Three of the Late Lapita mandibles from Watom are correctly classified to Watom and the remaining two are reclassified to the Northern Marianas.

Some general associations emerge when the group means are plotted on the first two canonical variates (Fig. 24.2). With the exception of New Zealand Maori, the Polynesian series form a loose cluster with the Northern Marianas mandibular series in this diagram. NKLE clusters with the series from East and Southeast Asia while Shihsanhang is closest to Pohnpei, a series from Micronesia. NKLE is closest to Mongolia, Siberia, and the Ryukyu Island mandibular series. Shihsanhang is moderately close to Pohnpei. The Late Lapita mandibular series from Watom Island (New Britain), Namu (in eastern Solomon Islands), and New Zealand occupy marginal positions in this plot.

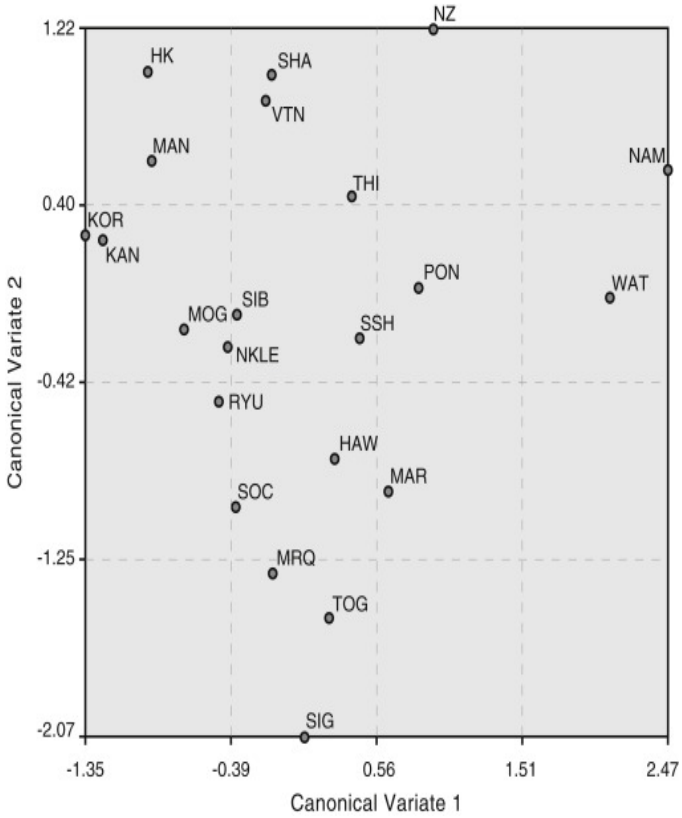

Figure 24.2. Plot of 22 group means on the first two canonical variates resulting from the application of stepwise discriminant function analysis to nine mandibular measurements. Abbreviations of the cranial samples are explained in Table 24.1.

When the group means are plotted on the first three canonical variates (Fig. 24.3), NKLE again is closest to the mandibular series from Siberia, Manchuria, Hong Kong, and Mongolia. Shihsanhang, Shanghai, and the Northern Marianas form a relatively tight association. Again, Watom, Namu, and New Zealand occupy marginal positions in this plot.

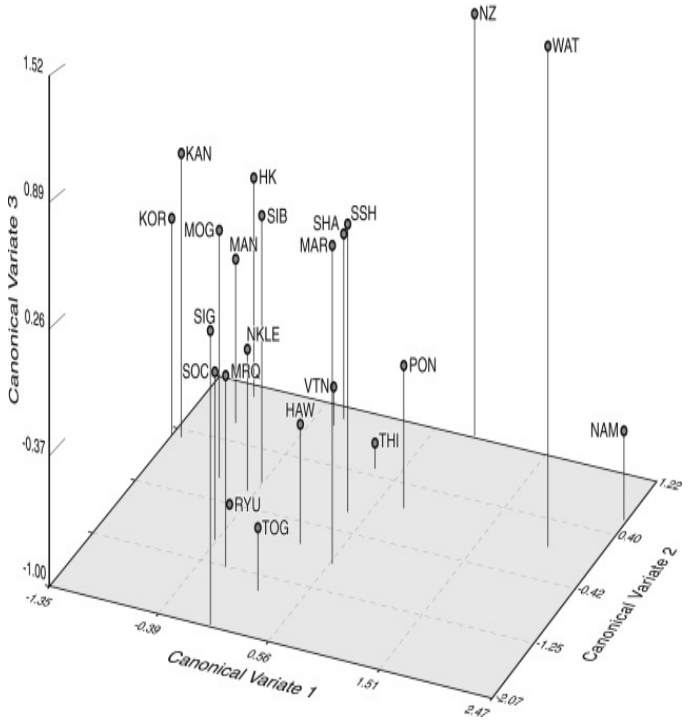

Figure 24.3. Plot of 22 group means on the first three canonical variates resulting from the application of stepwise discriminant function analysis to nine mandibular measurements. Abbreviations of the cranial samples are explained in Table 24.1.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B978012801966500024X

CHEMOMETRICS AND STATISTICS | Multivariate Classification Techniques

M.J. Adams, in Encyclopedia of Analytical Science (Second Edition), 2005

Discriminant Function Analysis

In LDA, or discriminant function analysis, we are seeking to create new synthetic features (variables) that are linear combinations of the original variables and that best indicate the differences between the known groups in contrast to the variable variances within the groups.

The process of performing LDA aims to derive and construct a boundary between the known classes of the training objects using statistical parameters. This boundary is derived using a discriminant function that provides a value or score when applied to a test object that indicates the group to which the new object should be assigned.

If $f(x_i, k)$ is some measure of likelihood of object x_i belonging to group or class k, then the discriminant score D_i, for assigning x_i to one of two classes is given by

$$\boldsymbol{D_i = f(x_i, k_1) - f(x_i, k_2)} \quad [2]$$

which may be interpreted as saying we classify test object x_i into class 1 if D_i is positive, otherwise x_i is considered as belonging to class 2.

The value of the discriminant score is calculated from a linear combination of the recorded values of the variables describing the objects, each suitably weighted to provide optimum discriminatory power. For two variables

$$\boldsymbol{D_i = w_0 + w_1 x_{i,1} + w_2 x_{i,2}} \quad [3]$$

The weights, or variables coefficients used in calculating D (eqn [3]), are determined by

$$w = (\overline{x}_j(1) - \overline{x}_j(2))S^{-1}$$
$$w_0 = -\tfrac{1}{2}\left[(\overline{x}_j(1) - \overline{x}_j(2))\mathbf{S}^{-1}(\overline{x}_j(1) + \overline{x}_j(2))\right] \quad [4]$$

This represents the ratio of the separation of the means of the two groups to the within-group variance for the groups as given by the pooled covariance matrix, **S**.

$\overline{x}_j(1)$ and $\overline{x}_j(2)$ are the vectors of the mean values for variable *j* for groups (1) and (2), respectively, and easily obtained from

$$\overline{x}_j(k) = \frac{\sum_{i=1}^{n(k)} x_{i,j}(k)}{n(k)} \quad [5]$$

where $n(k)$ is the number of objects in group (k), and $x_{i,j}(k)$ is the value for object i of variable j in group (k).

The pooled covariance matrix, **S**, of two training classes is given by

$$\mathbf{S} = \frac{\mathbf{S}_{(1)} - \mathbf{S}_{(2)}}{n(1)+n(2)-2} \quad [6]$$

$\mathbf{S}_{(k)}$ represents the covariance matrix of group (k).

The vector of weights coefficients can be thus be calculated from the classified training data and the discriminant score computed for each new, unclassified sample.

The discriminant function is linear; all the terms are added together to give a single number, the discriminant score.

For higher-dimensional pattern space the boundary is a hyperplane of m–1 dimensionality, where m is the number of variables. The partition boundary between two classes is defined at D_f=0 and in the two-dimensional case it is given by

$$w_0 + w_1 x_{i,1} + w_2 x_{i,2} = 0 \quad [7]$$

The classification boundary bisects a line between the centroids of the two clusters (Figure 8).

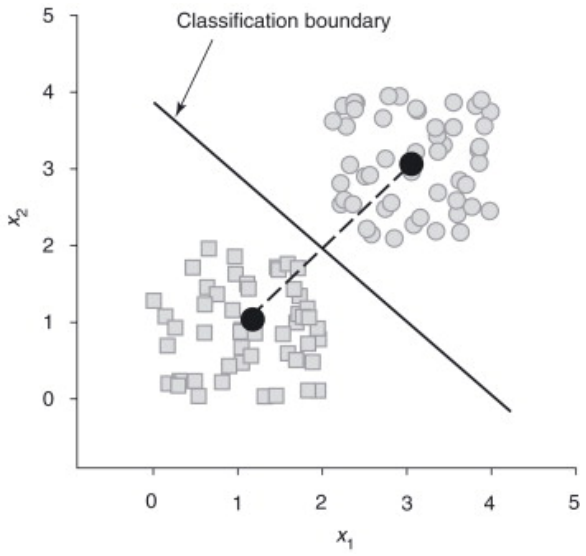

Figure 8. Linear discriminant analysis provides a linear partition boundary between the two known groups, bisecting the line between the cetroids of the two groups.

A useful result of performing LDA is the production of what is termed a discriminant plot, where every data point (from the training set or the new test sample set) is projected onto the discriminant function displayed as one-dimensional axis (Figure 9).

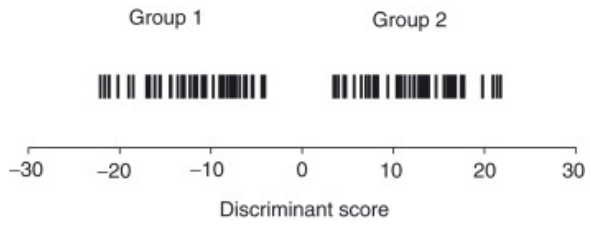

Figure 9. A discriminant plot projects the data on to a single axis (defined by the discriminant function).

The concept of a linear discriminant axis reduces the multidimensional classification problem to a single dimension, with the projection achieved so that discrimination between classes is preserved as well as possible.

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B0123693977000765

A Cognitive Portrait of Grade School Students with Mild Mental Retardation

Marcia Strong Scott, ... Lois-Lynn Stoyko Deuel, in International Review of Research in Mental Retardation, 1997

2 EARLY IDENTIFICATION OF MILD MENTAL RETARDATION

We first computed a stepwise discriminant function analysis using the following three dependent measures: object-class recognition score, associative recognition score, and number of errors. For the second stepwise discriminant function analysis, penalty score replaced the errors measure. In both analyses, the error (penalty) measure entered into the discriminant equation first and the object-class score entered second. The third variable, associative recognition score, did not enter into the equations. With both combinations, 82% of the students with mild retardation (sensitivity) and 88% of the students without mental retardation (specificity) were correctly identified.

When we examined the frequency distributions of the individual measures, the groups were not separable on the associative recognition measure. The object-class measure was associated with adequate specificity (85%) but inadequate sensitivity (39%). Number of errors yielded excellent sensitivity (91%) but poor specificity (68%). The penalty measure had better psychometric properties yielding a specificity of 91% and a sensitivity of 79%. Deriving a total score summed over the object-class and penalty measures was associated with the best outcome. Applying a cutoff at 7, and calling all children with a score of 7 or less mildly mentally retarded, we observed a sensitivity of 82% and a specificity of 88%. This is the same set of values achieved with both discriminant function analyses, except that using absolute scores, we only achieved this level with the combination of object-class and penalty scores. It seems obvious that the object-class recognition measure and penalty measure are the ones to include in the screening test, if the recognition task is presented in the same format. Given the difficulty of the task, however, even for 6- through 8-year-old children with mild mental retardation, we will probably simplify the task for further evaluation with a preschool-aged sample by requiring only identical recognition responses (i.e., “Point to all the pictures that you saw on the memory card.”).

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/S0074775008601792

Dental Metrics in Biodistance Analysis

M.A. Pilloud, M.W. Kenyhercz, in Biological Distance Analysis, 2016

Methods to Negate Multicollinearity

Many multivariate analyses, such as discriminant function analysis, assume that variables submitted to the analysis are not highly correlated (Klecka, 1980). Given the strong correlations among teeth within the dental arcade, data must be treated to remove correlations. Such treatment could include the identification and removal of correlated variables before multivariate analyses are conducted. However, there are also several techniques that can transform the raw data into something more useful for investigating variation within and among populations without leading to a reduction in the number of variables used in analyses.

Dental Indices. Populations differ in tooth size in one of two ways: (1) the overall average size of the majority of the variables (ie, a "size difference"); and (2) the relative distribution of size differences within the arcade, or "shape" differences. Wolpoff (1971) put forth the dental robusticity index (RI), which multiplies mesiodistal and buccolingual measures. Wolpoff believed that RI would essentially approximate the total occlusal area. Lukacs (1985) demonstrated the utility of several other dental indices including incisor breadth index, the molarization index, and the step-index. The incisor breadth index examines the relative proportion of incisor breadths by dividing the mesiodistal diameter of the lateral incisor by the mesiodistal diameter of the central incisor. The molarization index divides the buccolingual diameter of the lower second premolar by the buccolingual diameter of the lower first molar, essentially examining the relative "molar-ness" of the second premolar based on relative size. The step-index divides the buccolingual diameter of the second molar by the buccolingual diameter of the first molar, both maxillary and mandibular, to examine the relative size discrepancies. While the use of dental indices improved the utility of odontometric variables, Schmidt et al. (2011) suggested that dental indices, particularly RI, overestimate occlusal area.

PCA and Tooth Size Apportionment (TSA). PCA is a data reduction technique that takes a combination of correlated variables (in this case buccolingual and mesiodistal measures) and transforms them into a smaller subset of uncorrelated variables known as principal components (PCs) (Jolliffe, 2002). In PCA, each of the original variables influences each PC with different weights known as loadings. The relative influence of each variable is judged by the greatest loading, either positive or negative. Typically, the first PC comprises only positive loadings and is then considered to represent the overall size component (Jolliffe, 2002), and subsequent PCs are then thought to represent "shape" data. The number of PCs generated will equal the number of variables in the analysis; however, the first few PCs typically account for the majority of the variability in the sample as judged by eigenvalues. While there is no standard for selecting a particular eigenvalue cutoff for deciding how many PCs to retain, many studies only use PCs with eigenvalues greater than 1 (Kaiser, 1960). However, Jolliffe (2002) noted that PCs with eigenvalues over 0.7 should be retained.

PCA will have the option to rotate the data during the analysis. Rotation has traditionally been used to more easily interpret the PCs because the variance is distributed across the derived components (Kachigan, 1991). Harris (1997) suggests that varimax rotation should be utilized in odontometrics because the loadings for each PC are pushed to either one or zero, which simplifies interpretation. However, Hemphill (1991) argued that no rotation should be employed. Harris (1997) detailed three advantages to using PCA with odontometrics: (1) data are reduced into a compound variable that can be used to show shared variation; (2) developmental and statistical fields controlling tooth size are disclosed when PCA is used; and (3) each PC is statistically independent or uncorrelated. In some instances, PC1 is simply removed from the analysis, however, as noted by Harris (1997), there is still useful "shape" information in PC1.

The relative size of different dental fields, or how size is apportioned throughout the dental arcade, is an important consideration (Irish and Kenyhercz, 2013). TSA analysis examines the relative size of different teeth in the dental arcade within and among populations. TSA analysis has several advantages: (1) it removes the effect of overall size that often obscures odontometric relationships; (2) it is methodologically straightforward; and (3) with only population means necessary for analysis, many more meta-analyses can be conducted from published results.

To conduct TSA, a covariance matrix of mean buccolingual and mesiodistal diameters is submitted to PCA (Irish and Hemphill, 2004). Next, each of the diameters is multiplied by the loading for each tooth dimension. Then the PC

scores are regressed on some size measure. To examine the relative effect of size within the dentition, Harris and Rathbun (1989) regressed the PC scores with the sum of mesiodistal measures within the arcade. The resulting regression line then showed the expected PC scores given size. Deviations from this regression line (observed scores minus expected scores, otherwise known as residuals), demonstrate how tooth size has been apportioned. For example, if a group was shown to have positive residuals, or have a relatively greater PC score than estimated through regression, teeth with greater loadings on that PC score are proportionally larger than in other groups with negative residuals (Harris, 1997). However, Harris (1997) later advocated using three size measures to regress the PC scores: the sum of each of the mesiodistal X buccolingual measures, the sum of each of the mesiodistal measures, and the sum of each of the buccolingual measures. Harris (1997) argued that different PCs will be more in line with lengths or widths than simply with area.

Ancestry Estimation

Elizabeth A. DiGangi, Joseph T. Hefner, in Research Methods in Human Skeletal Biology, 2013

FORDISC and Discriminant Function Analysis

Once the data have been appropriately collected the next step is finding and using an appropriate known reference sample. In the United States, this is most often the **Forensic Anthropology Databank** (Jantz and Moore-Jansen, 1988) and the computer program ***FORDISC 3.0*** (FD3) (Jantz and Ousley, 2005). One part of properly utilizing FD3 is appreciating what, exactly, FD3 is doing. Fordisc uses **discriminant function analysis** (DFA) to classify an unknown individual into one of several reference populations and is, by and large, the most widely used classification statistic in forensic anthropology, particularly when the data are continuous.

Giles and Elliot (1962, 1963) first used a DFA on crania to determine sex and race for American White, American Black, and Amerindian[18] crania. Linear discriminant function analysis was developed as a means to classify a target individual (e.g., unknown crania) into one of several reference groups by incorporating a similar mathematical approach to regression analysis (Krzanowski, 2002). Whereas regression analysis uses a weighted combination of predictor variables to calculate some object's value (e.g., stature from measurements of the postcranial skeleton), DFA uses a weighted combination of those predictor variables to classify an unknown object into a reference group based on a distance statistic. The discriminant function score is a derived variable (Krzanowski, 2002), which is equal to the weighted sum of values for each variable.

The most common distance statistic employed in forensic anthropological research and classification is **Mahalanobis distance** (D^2), which is a distance measure similar in practice to Euclidean distance (the "ordinary" distance between two points as one would measure with a ruler), but that is not affected by scale or correlation (Krzanowski, 2002). Unlike Euclidean distance, D^2 is based on the covariance between variables and is used to measure the similarity (as the distance from a group **centroid**[19]) between unknown and known individuals. When interpreting the D^2 value, smaller distances equate to more similar individuals.

The statistical assumptions associated with DFA include *multivariate normality* and *homogeneity of variances/covariances*. Multivariate normality is one of the most common assumptions in statistics, as many tests and statistics are related to the normal distribution (think bell curve here). Generally, testing for multivariate normality is testing for univariate and bivariate normality, that is, testing to see that each variable is normally distributed and, likewise, that all pairs of variables are bivariate normal using one- and two-dimensional plots (i.e., histograms and scatterplots). In practice, this is generally sufficient for testing for multivariate normality, especially when using DFA as that method is relatively robust against deviations from multivariate normality. Other more robust methods to test for

multivariate normality exist, but are beyond the scope of this work (cf., Mardia's statistic of multivariate skewness/kurtosis [Mardia, 1970] or the Doornik-Hansen multivariate normality test [Doornik-Hansen, 2008]).

The second assumption involves whether there is homogeneity of variances/covariances (or, testing that the level of variation in each group is relatively similar) and testing for this is also relatively straightforward. There are a variety of tests for homogeneity. In FD3, homogeneity among samples is tested using the Kullback (1959) test for homogeneity. If the level of heterogeneity within groups is high the analyst is encouraged to explore other statistical procedures, such as logistic regression (Jantz and Ousley, 2005).

Two additional considerations in DFA are **outliers** and *multicollinearity*. Discriminant function analysis is sensitive to the inclusion of outliers (individuals or measurements falling far outside the collective distribution of all other individuals or measurements). The researcher should carefully consider the data through graphs (plots) and descriptive statistics to identify potential outliers. If outliers are found, the cause for each should be identified, when possible. Remember, transcription errors (e.g., 24 entered as 42), incorrect data entry (entering maximum cranial breadth (XCB) for maximum cranial length (GOL)), and measurements that are just wrong (XCB measured as 145 when it is in fact 120) may lead to outliers. When these types of errors are identified the data should be corrected. If no explanation can be found, the individual may be dropped from the analysis unless there is good reason to suspect he or she is just an expression of the variation seen in that population.

Multicollinearity is the same as trait interdependence (correlation). When two variables are highly correlated (or one is the sum of other dependents) the parameter estimates behave erratically when the model (or the variables) undergoes even minute changes. While this does not affect the overall model, it does affect classifications based on that model. In other words, collinearity also means the standardized discriminant function coefficients cannot reliably assess the relative importance of the predictor variable(s), decreasing the overall strength of the final discriminant function for classification purposes. As with outliers, graphs (two-dimensional plots) of the variables will assist in identifying highly correlated variables.

Two additional statistics that can be obtained from the discriminant function analysis provide further information about the classification. The FORDISC 3.0 help file (Jantz and Ousley, 2005) goes into great detail about posterior and typicality probabilities, but a brief explanation will help the reader better understand some of the analyses described below. **Posterior probability** is the probability that the unknown belongs to any one of the populations selected for in the analysis and is based on the relative distances the unknown has (calculated using Mahalanobis distance, or D^2) to each population. Because it is the probability of belonging to any one of the populations used in the analysis, the posterior probability will always sum to 1. A major assumption (of classification statistics in general) is that the unknown individual truly belongs to one of the reference groups (hence the need for strict guidelines when selecting reference samples), because a DFA will always "force" a classification.

We can use another statistic, **typicality probability**, as a measure of how *likely* it is that the unknown does, in fact, belong to any one of those populations. Typicality probability is based on the absolute distances of the unknown from all groups, rather than the relative distances. Please note that the typicality probability is essentially equivalent to a univariate t-test. In other words, it is a measure of how many other individuals in a population would be expected to be as far or farther from that population's centroid than the unknown individual. As Jantz and Ousley (2005:np) point out "[typicality probabilities] below 0.05 (5%), or certainly 0.01 (1%) for a group … indicate questionable probability of membership in that group or the possibility of measurement error." This means that the typicality probability can essentially be ignored if the value is greater than 0.05, since such values do not indicate a statistically significant difference in the suite of measurements. When the value is less than 0.05, carefully consider the measurements entered and the populations (reference samples) included in the analysis.

Read full chapter