Forensic Science International 245 (2014) 204.e1–204.e8

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Forensic Anthropology Population Data

Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines

Frédéric Santos [a,*], Pierre Guyomarc'h [a], Jaroslav Bruzek [a,b,c]

[a] University Bordeaux—CNRS—MCC, PACEA, UMR 5199, F-33615 Pessac, France
[b] Charles University, Faculty of Science, Department of Anthropology and Human Genetics, Prague, Czech Republic
[c] West Bohemia University, Faculty of Humanities, Department of Anthropology, Plzen, Czech Republic

ARTICLE INFO

Article history:
Received 12 February 2014
Received in revised form 21 July 2014
Accepted 5 October 2014
Available online 13 October 2014

Keywords:
Forensic anthropology population data
Statistics
Sex estimation
Population
Accuracy
Reliability

ABSTRACT

Accuracy of identification tools in forensic anthropology primarily rely upon the variations inherent in the data upon which they are built. Sex determination methods based on craniometrics are widely used and known to be specific to several factors (e.g. sample distribution, population, age, secular trends, measurement technique, etc.). The goal of this study is to discuss the potential variations linked to the statistical treatment of the data. Traditional craniometrics of four samples extracted from documented osteological collections (from Portugal, France, the U.S.A., and Thailand) were used to test three different classification methods: linear discriminant analysis (LDA), logistic regression (LR), and support vector machines (SVM). The Portuguese sample was set as a training model on which the other samples were applied in order to assess the validity and reliability of the different models. The tests were performed using different parameters: some included the selection of the best predictors; some included a strict decision threshold (sex assessed only if the related posterior probability was high, including the notion of indeterminate result); and some used an unbalanced sex-ratio. Results indicated that LR tends to perform slightly better than the other techniques and offers a better selection of predictors. Also, the use of a decision threshold (i.e. $p > 0.95$) is essential to ensure an acceptable reliability of sex determination methods based on craniometrics. Although the Portuguese, French, and American samples share a similar sexual dimorphism, application of Western models on the Thai sample (that displayed a lower degree of dimorphism) was unsuccessful.

1. Introduction

Sex determination of skeletal remains is a major step in the identification process in forensic anthropology. Depending on the experience of the forensic anthropologist and the choice of practice during the evaluation of the osteobiographic profile, the ancestry of the subject may remain unknown [1]. Due to the general social globalization (people and their remains move across borders and continents), it is essential to aim at the development of robust techniques that are as independent as possible from the population sample on which they are built. Sex determination methods are indeed known to be population-specific, explained by variations in sexual dimorphism [2,3]. Moreover, secular trends influence skeletal morphology at a significant rate with modern social changes and mobility, jeopardizing the forensic application of traditional methods that are not based on contemporaneous samples [4–6]. Consequently, the data used to create identification methods reflects the influence of complex genetic and epigenetic factors. Above all, the reliability of an identification method also depends on the statistical treatment of this data; our study specifically investigates this aspect.

Sex determination methods based on different skeletal parts have been increasingly published, and most rely on linear discriminant analysis (LDA). According to a recent review [7] made on 24 specialized journals between 2000 and 2010, LDA is used nine times more often than logistic regression (LR). Other classification techniques are rarely chosen in this context; however, the results obtained with LDA commonly indicate that the relation between a group of measurements and the probability of being male or female is not linear [8,9].

A decade ago, Feldesman stated that anthropologists widely use LDA for classification, but he did not discuss the limitations that have been well documented in multivariate statistics manuals [10]. Since then, the general approach for the development of sex

* Corresponding author at: University Bordeaux, UMR 5199 PACEA, A3P, bat B8, allée Geoffroy St Hilaire, CS 50023, 33615 Pessac Cedex, France.
Tel.: +33 540008932.
E-mail address: frederic.santos@u-bordeaux.fr (F. Santos).

http://dx.doi.org/10.1016/j.forsciint.2014.10.010
0379-0738/

determination methods has not drastically changed. Of note, LDA is usually available in traditional software and easy to implement (i.e. SPSS, SAS, Statistica). The use of discriminant functions largely depends on a multivariate normal distribution of the variables and an equality of the matrices' covariance between the two compared groups. Such factors are rarely taken into account in the majority of the publications with the exception of a few recent studies [11,12].

In order to free the variables from LDA constraints [8], researchers have proposed sex determination models based on non-linear classifications, such as LR [13–17] or neural networks (NN) [8]. However, similar to the LDA, NN is sensitive to the non-equality of the groups' variances. When the variances are high, it appears that LDA, LR, and especially NN, are limited in terms of correct classification [18].

Some techniques, such as those that support vector machines (SVM) hold an advantage compared to the above-mentioned methods. During the last few years, SVM has been introduced in several disciplines, including medical sciences, where it showed successful applications [19–22]. This method is particularly efficient in a two-group classification [23], such as male versus female. New investigations are still necessary to establish promote alternative solutions to LDA in practice [22].

The new standards required in forensic sciences stress the importance of scientific methodology. This study addresses two relevant factors inherent to the Daubert ruling: the empirical testing of methods and the related potential error rate [24,25]. The choice of the model in sex determination methods is directly linked to the portability and applicability of the model. In practice, LDA is particularly used in forensic anthropology (e.g. [1,26–28]) and it daily influences the experts' decisions, because it proposes a simple application of formulae published in peer-reviewed journals. The goal of this study is to test the classification power of three methods on skeletal measurements: LDA, LR and SVM. Increased attention brought to the methodology in forensic anthropology also includes the correct use of statistical tools. We chose to test the models with craniometric data, not only because if offers wider variations within and among populations, but also because of its extensive use in forensics.

2. Materials and methods

2.1. Data

The sample of the study is composed of four different subsamples of skulls (individuals of known age and sex), extracted from osteological reference collections of different geographical origins:

Table 1
Sex distribution and mean age of the sample by population (POR = Portugal; FRA = France; USA = United States; THA = Thailand; m = mean age in years; sd = standard deviation).

Population	Male		Female		Total (*n*)
	n	*m* (SD)	*n*	*m* (SD)	
POR	53	51 (19)	54	57 (20)	107
FRA	25	50 (10)	25	56 (13)	50
USA	33	54 (13)	33	69 (18)	66
THA	47	63 (14)	45	63 (16)	92
Total (*n*)	158		157		315

1. The Coimbra Identified Skeletons Collection (University of Coimbra, Portugal), composed of Portuguese individuals deceased during the first half of the twentieth century [29,30].
2. The Olivier Documented Collection (Musée de l'Homme, Museum National d'Histoire Naturelle, Paris, France), composed of French individuals deceased during the mid-twentieth century [31–33].
3. The Maxwell Museum's Documented Skeletal Collection (Albuquerque, New Mexico, USA), composed of American residents deceased at the end of the twentieth century [34,35].
4. The Thai Skeletal Collection hosted by the Department of Anatomy of the University of Chiang-Mai (Thailand), composed of Thai individuals deceased at the end of the twentieth century [36–38].

The sex and age repartition of each sample is summarized in Table 1.

Craniometric variables of the Portuguese, American and Thai samples were measured in millimeters using digital calipers. The French sample was measured with a digitizer, and linear measurements were extracted from the 3D coordinate data of the corresponding landmarks (see [32,39] for description and validation of the technique). Sixteen traditional measures, conventionally applied in sexing techniques, are used in this study; codes and definitions are presented in Table 2. Subsamples of individuals are also used depending on the measurements' availability in the different collections.

Table 2
List of the craniometric variables available for each population (POR, Portugal; FRA, France; USA, United States; THA, Thailand): The measurements that are missing for a whole sample are indicated by an X. Sexual dimorphism is evaluated by the *p*-values of *t*-tests, rounded to the second decimal place. Craniometric measurements codes [66,67] and definitions of the variables are also given.

Code		Definition	POR	FRA	USA	THA
Martin [66]	Howells [67]					
M1	GOL	Glabello-occipital length	<0.01	0.03	<0.01	0.29
M5	BNL	Basion-nasion length	<0.01	0.62	0.12	0.24
M8	XCB	Maximum cranial breadth	<0.01	0.14	0.08	0.35
M9	WFB	Minimum frontal breadth	<0.01	<0.01	<0.01	0.02
M12	ASB	Biasterionic breadth	<0.01	0.36	0.01	X
M17	BBH	Basion-Bregma height	<0.01	0.01	<0.01	0.02
M11	AUB	Biauricular breadth	<0.01	0.02	<0.01	0.54
M29	FRC	Nasion-Bregma chord	<0.01	<0.01	X	0.86
M30	PAC	Bregma-Lambda chord	<0.01	<0.01	<0.01	0.1
M31	OCC	Lambda-Opisthion chord	<0.01	0.01	X	0.43
M40	BPL	Basion-Prosthion length	0.14	X	X	0.06
M45	ZYB	Bizygomatic breadth	0.02	0.59	X	0.93
M48	NPH	Nasion-Prosthion height	<0.01	0.04	X	0.7
M52	OBH	Orbit height (left)	0.01	0.08	X	0.79
M54	NLB	Nasal breadth	<0.01	0.61	X	0.05
M61	MAB	Palate breadth	<0.01	0.01	X	0.1