This is a pre-print of an article published in the International Journal of Legal Medicine. The final authenticated version is available online at: https://doi.org/10.1007/s00414-019-02148-4

Sex estimation: A comparison of techniques based on binary logistic, probit and cumulative probit regression, linear and quadratic discriminant analysis and neural networks using ordinal variables

Efthymia Nikita, PhD[a]; Panos Nikitas, PhD[b]

[a] Science and Technology in Archaeology and Culture Research Center, The Cyprus Institute, 2121 Aglantzia, Nicosia, Cyprus; efi.nikita@gmail.com

[b] Department of Chemistry, Aristotle University of Thessaloniki, University Campus, 54124 Thessaloniki, Greece; nikitas@chem.auth.gr

Abbreviated title: Sex estimation using regressions, discriminant analyses and neural networks

Declarations of interest: none

Corresponding author contact details:

Efthymia Nikita, PhD

Science and Technology in Archaeology and Culture Research Center, The Cyprus Institute, 20 Konstantinou Kavafi Street, 2121 Aglantzia, Nicosia, Cyprus

tel. +357 22 208 669

efi.nikita@gmail.com

ORCID: 0000-0003-2094-5047

Acknowledgments

Efthymia Nikita's contribution to this research was supported by the H2020 Promised project (Grant Agreement 811068), and the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation (People in Motion project: EXCELLENCE/1216/0023).

Abstract

The performance of six classification methods, binary logistic (BLR), probit (PR) and cumulative probit (CPR) regression, linear (LDA) and quadratic (QDA) discriminant analysis, and artificial neural networks (ANN), is examined in skeletal sex estimation. These methods were tested using cranial and pelvic sexually dimorphic traits recorded on a modern documented collection, the Athens Collection. For their implementation, an R package has been written to perform cross-validated (CV) sex classification and give the discriminant function of each of the methods studied. A simple algorithm that combines two discriminant functions is also proposed. It was found that the differences in the classification performance between BLR, PR, CPR, LDA, QDA, and ANN are overall small. However, LDA is simpler and more flexible than CPR, QDA and ANN and has a small but clear advantage over BLR and PR. Consequently, LDA may be preferred in skeletal sex estimation. Finally, it is striking that the combination of pelvic and cranial traits via their discriminant functions determined either by BLR or LDA removes practically any population-specificity and yields much better predictions than the individual functions; in fact, the prediction accuracy increases above 97%.

Keywords: Forensic Anthropology; sex estimation; cranium; pelvis; statistical methods

Introduction

Sex estimation is a key aspect of any forensic anthropological analysis as sex is one of the three key parameters used in the identification of an unknown individual, along with age and stature [1-3]. Even though sexual dimorphism in humans is small, there are morphological and metric differences between male and female skeletons, which have been used in sex estimation. Morphological methods focus primarily on the pelvis and secondarily on the cranium, and they assess the degree of expression of specific traits. Based on the qualitative assessment of this degree, they assign the individual to the male or female sex [4-6]. More recently, statistical analysis has started being employed whereby the degree of expression of selected pelvic and cranial traits, recorded in an ordinal scale, is used as input variable in sex prediction using binary logistic regression analysis and/or discriminant analysis [7-8]. This statistical approach allows both the prediction of sex but also the estimation of the probability of an individual being male or female.

Discriminant analysis and, more recently, artificial neural networks have also been used in metric sex estimation whereby the variables are measurements that capture both size and shape differences between males and females [9-14]. A limitation of metric methods compared to morphological ones is that they are of limited applicability to fragmented partially preserved remains.

A comparison of different techniques used for sex classification based on ordinal variables was carried out by Walker [8]. Walker [8] used cranial traits and examined the performance of four multivariate techniques: k-nearest neighbor, binary logistic regression (BLR), linear (LDA) and quadratic (QDA) discriminant analysis. He found that QDA, and especially k-nearest neighbor analysis, performed poorly in terms of the sex bias criterion since they produced discriminant functions the classification error rates of which were much greater for one sex than the other. In contrast, BLR and LDA had both low misclassification and low sex bias rates. From these two techniques, Walker [8] chose BLR as the best technique, mainly because it relies on fewer assumptions than LDA. Comparisons between artificial neural networks (ANN) and the classical methods LDA, QDA and BLR have been recently carried out and found that ANN perform better than the other methods. However, all these comparisons concern metric and not ordinal data [15-17].

The current paper examines the performance of the above classification methods, BLR, LDA, QDA, and ANN, in sex estimation using ordinal predictors as well as two alternatives of BLR, the probit (PR) and the cumulative probit (CPR) regression. From these methods the most popular is the first one, since it allows the direct development of mathematical relationships between sex traits and sex. These relationships can then be easily applied for sex estimation in different assemblages. In contrast, discriminant analyses and neural networks require that the training sample be used each time they are employed in sex estimation and this is a major disadvantage. For this reason, in the present paper, apart from the evaluation of these six methods, we develop simple mathematical relationships for each of them that can be used for sex estimation, as is already the case for binary logistic regression. In addition, we explore the performance of a new approach that combines two different discriminant functions for sex prediction.

Classification methods

Binary logistic and probit regression

Logistic and probit models are common statistical tools for the analysis of dichotomous data. In the context of sex prediction, binary logistic regression (BLR) presumes that the sex traits (predictors) have been properly coded, usually in an ordinal scale. Similarly, the sex of the individual may also be coded as a binary variable, say 0 for male and 1 for female. Consider the sex traits X_1, X_2, … Binary logistic regression establishes a relationship among these variables and the probability P(sex=1) that the individual is female (sex=1) based on the following expression [18]:

$$ln\frac{P(sex=1)}{1-P(sex=1)} = c_0 + c_1X_1 + c_2X_2 + \cdots \quad (1)$$

where c_0, c_1, c_2, … are (adjustable) parameters usually determined using maximum likelihood estimation provided that a training sample with documented sex vs. X_1, X_2, … data is available. When c_0, c_1, c_2, … have been calculated, the above equation can be used to calculate P(sex=1). Thus, if

$$y = c_0 + c_1X_1 + c_2X_2 + \cdots \quad (2)$$

then

$$P(sex = 1) = \frac{1}{1+e^{-y}} \quad (3)$$

Therefore, if P(sex=1) > 0.5, the individual is female, otherwise the individual is male with a probability equal to P(sex=0) = 1 - P(sex=1).

Tables with c_0, c_1, c_2, … values or the expression $y = c_0 + c_1X_1 + c_2X_2 + \cdots$ can be found in the literature and they can be used to compute P(sex=1) by means of Equation (3) and, therefore, estimate sex from specific traits. Such a well-known table is Walker's [8] table with y equations for estimating sex from cranial traits based on American/English and Native American assemblages. In this table, as well as in all relevant tables, the percentage of correctly classified males/females is also presented. This is an important piece of information because it evaluates the performance of the regression equations when applied to a certain sample. This information is obtained by means of leave-one-out cross validation (LOOCV). According to this technique, we remove from the dataset the first case, apply binary logistic regression to the remaining cases and, based on the obtained regression equation, we estimate the sex of the held-out case. This procedure is repeated for the second case and so on until the last case. Then we count the correct predictions and express them as a percentage of correctly classified cases.

In probit regression (PR), the inverse of the standard normal cumulative distribution Φ^{-1} of the probability P(sex=1) is modeled as a linear combination of the predictors [18, 19]:

$$\Phi^{-1}\big(P(sex = 1)\big) = c_0 + c_1X_1 + c_2X_2 + \cdots \tag{4}$$

where the adjustable parameters $c_0, c_1, c_2, \ldots$ may also be determined using the maximum likelihood method as in BLR. When $c_0, c_1, c_2, \ldots$ are known, the above equation can be used to calculate P(sex=1) since:

$$P(sex = 1) = \Phi(y) = \Phi(c_0 + c_1X_1 + c_2X_2 + \cdots) \tag{5}$$

where Φ is the standard normal cumulative distribution function. This function may be easily calculated using for example the NORMSDIST function in Excel or the pnorm function of R.

BLR and PR are not subject to strict assumptions. BLR does not require a linear relationship between the dependent and independent variables, the error terms (residuals) or the predictors do not need to be normally distributed, it can handle both categorical and continuous variables, whereas homoscedasticity is not required [18]. The same properties hold for PR but for this method the error terms should be normally distributed. In addition, both methods are sensitive to outliers but this is not a serious issue when using ordinal data.

Cumulative probit regression

The use of binary logistic and probit models in sex assessment and in particular the validity of the estimated probabilities P(sex=1) was criticized by Konigsberg and Hens [19] within the frames of Bayes' Theorem. These authors proposed as a better alternative the cumulative probit model/regression (CPR). Here, we applied the cumulative probit model as follows. First, consider the simple case of just one sex indicator, X_1, which can be coded using a J-point ordinal scale. That is, it can take the integer values i = 1, 2, ..., J. If we model the ordinal variable X_1 as a function of the binary variable *sex* (independent variable), the fitting probit model is expressed not by just one equation but from (J-1) equations of the form:

$$\Phi^{-1}(P_{ci}) = a_i + b * sex,\ i = 1, 2, \ldots, J - 1 \tag{6}$$

where P_{ci} is the cumulative probability of category i, that is, the probability of X_1 falling in category i or below. Therefore,

$$P_{ci} = P_1 + P_2 + \cdots + P_i \tag{7}$$

where P_1, P_2, …, P_i is the probability of X_1 falling exactly in category 1, 2, …, i, respectively. In fact, P_1, P_2, …, P_i are conditional probabilities and, in particular, P_i is the probability conditional on *sex* that an individual is in the i-th category of the sex indicator X_1. When we fit the probit model to the ordinal response X_1 using *sex* as independent variable, we obtain the adjustable parameters $a_1, a_2, \ldots, a_{J-1}, b$ and then the conditional probabilities may be calculated from the following system of equations:

$$P_{c1} = P_1 = \Phi(a_1 + b * sex)$$
$$P_{c2} = P_1 + P_2 = \Phi(a_2 + b * sex)$$
$$\cdots \qquad (8)$$
$$P_{c(J-1)} = P_1 + P_2 \cdots + P_{J-1} = \Phi(a_J + b * sex)$$

Now, if the observed score of X_I for an individual is i, the probability that the individual is a female may be estimated by applying Bayes' Theorem [19]

$$P(sex = 1) = P_i P_M / (P_i P_M + P_i P_F) \qquad (9)$$

where P_M is the prior probability that an individual is male and P_F is the prior probability that an individual is female. These quantities may be estimated, as in Discriminant Analysis, from $P_M = n_M/n$, $P_F = n_F/n$ where n_M, n_F is the number of males and females, respectively, and n is the total number of cases. Alternatively, they may be estimated via an optimization procedure that searches for the P_M, P_F values that yield the optimum prediction performance in the training sample.

When there are many sex traits, X_I, X_2, …, Equation (9) is still valid but now $P_i = P_I$ is the probability conditional on *sex* that an individual is in the state I composed of the i_1-th category of X_I, i_2-th category of X_2 and so on. In this case we may calculate the conditional probabilities P_I, P_2, …, P_J for each sex indicator X_I, X_2, … and then P_I may be estimated under the assumption of conditional independence. This is the main assumption of this method.

Discriminant analysis

An alternative approach for sex classification and prediction is Discriminant Analysis (DA). DA is based on the Mahalanobis distance between $x = (X_1, X_2, \ldots, X_m)$ and the mean vectors μ_0 and μ_1, where x is a vector of sex traits (variables) and μ_0 and μ_1 are the mean vectors of the male and female subsamples.

There are two main variants of DA: quadratic (QDA) and linear (LDA) discriminant analysis. In QDA the probability P(sex=1) is estimated via the following relationships [20-21]:

$$P(sex = 1) = \frac{e^{d_1(x)}}{e^{d_1(x)} + e^{d_0(x)}} \qquad (10)$$

$$d_k(x) = -\frac{1}{2}(x - \mu_k)^t \Sigma_k^{-1}(x - \mu_k) - \frac{1}{2}\ln|\Sigma_k| + lnP_k \qquad (11)$$

where k = 0 or 1 are the classes of male and female, respectively, Σ_k is the covariance matrix for class k, and P_k are the prior probabilities used. The last quantity is estimated from $P_k =$

n_k/n, where n_k is the number of cases k and n is the total number of cases. Note that the first term in Equation (11) is the so-called Mahalanobis distance between x and μ_k. It can be shown that Equations (10) and (11) may be expressed as Equations (2) and (3) where y is given by:

$$y = a_0 + \sum_{i=1}^{m} a_i X_i + \sum_{i=1}^{m} b_i X_i^2 + \sum_{i<j}^{m} c_{ij} X_i X_j \tag{12}$$

QDA is reduced to LDA if we assume $\Sigma_k = \Sigma$, where Σ is the pooled sample covariance matrix. In this case Equation (12) is simplified to

$$y = a_0 + \sum_{i=1}^{m} a_i X_i \tag{13}$$

Equations (12) and (13) can be used like Equation (2) for sex estimation. That is, based on a training dataset, the probabilities P = P(sex=1) are estimated for each case using linear or/and quadratic discriminant analysis. Then the adjustable parameters a_0, a_i, b_i, c_{ij} may be calculated by means of linear or multilinear regression using log(P/(1-P)) as dependent variable and X_i or X_i, X_i^2, $X_i X_j$ as independent variables. Sex prediction in the target sample is carried out via Equation (3) and Equation (13) or/and Equation (12).

Discriminant analysis is a parametric multivariate technique and, therefore, it is subject to several assumptions: The size of the smallest group should exceed the number of sex traits (independent variables), the variables should follow the multivariate normal distribution, the variance/covariance matrices of variables should be homogenous across groups, whereas multi-collinearity among variables should be excluded. Note that the equality of covariances assumption is not required in QDA and this is the basic reason that LDA is a much less flexible classifier than QDA. Nonetheless, as Tabachnick and Fidell [18] point out, if classification is the primary goal, then most of the above requirements do not affect the classification performance, provided that there are no outliers. Deviation from these assumptions may distort only tests of statistical significance, just as in MANOVA, if such tests have been carried out.

Neural networks

An (artificial) neural network (ANN) is a system of interconnected neurons, which is used to identify clusters in a data set via processes that mimic the way the human brain works [22-23]. From this point of view, it can be used for sex classification and prediction. A simple neural network is shown in Figure 1. When the neural network is used for sex classification, the leftmost layer of the network (input layer) consists of m input units that represent the traits $X_1, X_2, \dots, X_m$, whereas the output layer has only one node, which is related to the sex variable. Note that this variable, i.e. the output of the network, takes values between 0 and 1 and, therefore, it is related to the probability P(sex=1). In the example of Figure 1, there is just one hidden layer with three nodes. It is called hidden layer because its values are not observed in the training set. The circles labeled “1” are called bias units.

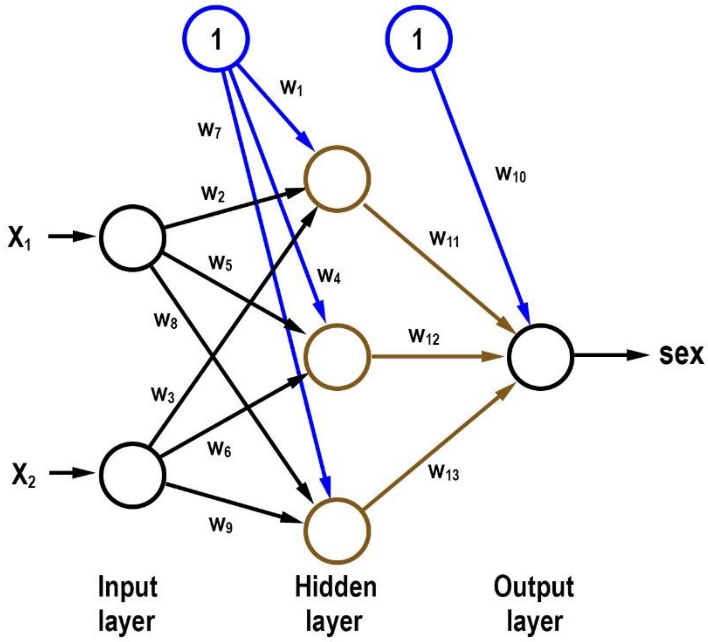

Fig. 1 Simple neural network for sex classification with two input nodes and one hidden layer with three nodes

In every neural network the input values $X_1, X_2, \ldots, X_m$ are transformed to an output value via proper weights, w_i. Thus, a neural network with m input nodes and one hidden layer with p nodes represents the following function provided that the numbering of the weights is as in Figure 1 [23, p. 143]:

$$sex = f\left(w_{(m+1)p+1} + \sum_{i=1}^{p} w_{(m+1)p+i} H_i \right) \tag{14}$$

where

$$H_i = f\left(w_{(i-1)(m+1)+1} + \sum_{j=1}^{m} w_{(i-1)(m+1)+1+j} X_j \right) \tag{15}$$

Here, f is the activation function of the network and $sex \approx$ P(sex=1), i.e. we may at least as a first approximation adopt that the ANN output is the probability P(sex=1). In what concerns the function f, we have adopted the logistic function $f(x)=e^x/(1+e^x)$.

It is seen that if we know the number of inputs and the hidden nodes as well as the weights, the estimation of sex is straightforward via Equations (14) and (15). There are several algorithms to train a network and compute weights. The most interesting point is that all these algorithms do not assume any underlying pattern for the data.

Statistical methods

An R package has been written to perform cross validated sex estimation based on the above six classification methods. This package includes eleven R functions: CVBLR, CVPR, CVCPR, CVLDA, CVQDA, CVNN, sdiscr, sdiscr2, discrNN, discrQDA, and discrCPR. The first five perform cross-validated binary logistic and probit regression, cumulative probit regression, and linear and quadratic discriminant analysis, whereas the sixth function is used to implement artificial neural networks. The next function, sdiscr, may be used to estimate the sex and the corresponding probability for one or more individuals using an equation of the form of Equations (2), (3), (5), (12), (13), i.e. equations that may be determined from the CVBLR, CVPR, CVLDA, and CVQDA functions. The sdiscr2 function can combine two discriminant functions that arise either from the BLR or LDA methods. Finally, the discrNN function estimates the sex and the corresponding (approximate) probability by means of weighting factors obtained from the CVNN function using Equation (14), the discrQDA function uses Equations (10) and (11) instead of Equation (12) for predictions based on QDA, and the last function, discrCPR, can be used for predictions based on the CPR method.

In this R package, binary logistic and probit regression are implemented using the basic glm function in combination to the CVbinary function of the DAAG library to perform leave-one-out cross-validation (LOOCV). In addition, the stepAIC function of the MASS library is used for backward stepwise model selection, i.e. for variable reduction, which is a crucial step for building a simpler model without losing the predictive power of the data. Linear and quadratic discriminant analyses are performed via the lda and qda functions of the MASS library. These functions can also perform LOOCV. For variable reduction, the stepclass function of the klaR library is used adopting the backward technique and the accuracy (AC) as performance measure. For the CPR method, we used the polr function of the MASS library. The CVCPR function may perform LOOCV but it does not have the option for variable reduction.

For the implementation of ANN, the R function nnet of the homonymous library was adopted. This function is based on the feed-forward algorithm, which is a rather fast algorithm. However, we should point out that, irrespective of the algorithm used for the estimation of the weights, different runs of an ANN algorithm may give different weights, especially if their number is large enough. For this reason, for the neural networks training we used 20 repetitions and the networks with the optimum performance were selected for sex prediction. It is evident that the ANN performance depends upon the number of the hidden nodes. This number was varied from 1 to 40. An important consideration when using ANN are overfitting issues and, for this reason, it is important to perform LOOCV. Note that nnet does not have the option to perform LOOCV, and for this reason in the present study, we have written the relevant code. In what concerns variable reduction, this may be achieved qualitatively via Garson's plot, which shows the relative importance of the input variables. In the present study, we instead chose to use in the input layer the variables obtained from BLR/LDA/QDA.

Finally, we combined two discriminant functions based on different skeletal traits to test if this would increase the performance of the individual functions. An algorithm that can be used to combine two such functions, say one based on cranial and the other on pelvic traits, is the following. Consider that the prediction for an individual using the first discriminant function is male and using the second function is female. Then, we examine the inequality P1(sex=0)>P2(sex=1), where P1(sex=0) is the probability of the individual being male according to the first function and P2(sex=1) the probability of the individual being female according to the second function. If the inequality is true, the individual is assumed to be male, otherwise the individual is female. The case where the first discriminant function predicts female and the second function predicts male is treated similarly. This algorithm is implemented in the sdiscr2 function.

More details and the code of the R functions are given as Online Resources 1 and 2.

Materials

The Athens Collection, housed at the Department of Animal and Human Physiology at the National and Kapodistrian University of Athens, Greece, consists of 225 skeletons. Most of this material has documented age, sex, occupation, and cause of death and represents individuals who lived in the second half of the twentieth century and were buried in cemeteries in the area of Athens [24]. Of this collection, 191 skeletons (106 males and 85 females) were studied in the context of this paper, excluding subadults, individuals with insufficient documentation and individuals with pathological lesions or taphonomic damage that could inhibit the correct recording of the sex traits under examination. Note that the Athens Collection is divided in two parts; the skeletons that formed the original core of this collection have the coding WLH, while the skeletons that were added in the collection subsequently received a code of ABH. From these data, the 132 skeletons with ABH code were used for training and the 59 skeletons with WLH code were used for prediction.

The pelvic traits recorded included the ventral arc (VA), subpubic concavity (SPC), and medial aspect of the ischiopubic ramus (MAR) and their scoring followed the 5-grade scheme proposed by Klales et al. [7]. The cranial traits included the mental eminence (ME), supraorbital margin (ORB), supraorbital ridge/glabella (GL), nuchal crest (NC), and mastoid process (MA), recorded as described by Walker [8]. Part of this dataset (120 cases) is provided as Online Resource 3 in order to practice using the provided R code.

Results and Discussion

The results obtained are presented in Tables 1 to 7 as well as in the Tables and Figures given as Online Resource 4. Tables 1 to 3 show the percentage of correctly classified individuals by means of BLR, PR, LDA and QDA, whereas Tables 4 and 5 present selected results concerning the classification performance of the neural networks when the number p of the nodes of the hidden layer is equal to 1 and when p gets its optimum value, i.e. the value that yields optimum sex classification. Table 6 shows the discriminant functions of LDA that result in the optimum sex classification for the Athens Collection and, finally, Table 7 shows the percentage of correctly classified individuals when two discriminant functions, one using pelvic and the other cranial traits, are combined. The results concerning the CPR method are presented in Tables S1 and S2 and Figure S1. In addition, Online Resource 4 contains the complete results obtained from the neural networks (Tables S3 to S5 and Figure S2). Note that in all ANN results, for the weights decay we used the value 0.5.

From Tables 1 to 3 we obtain that the differences between BLR, PR, LDA, QDA are overall small; LDA gives slightly better results than BLR in 36% of the cases given in these tables, the opposite is valid in 26% of the cases, and the two methods give identical results in the remaining 38% of the cases. However, if we examine the sex prediction on the target sample, we observe that LDA generally gives better and more balanced results between sexes. For the differences between LDA and QDA, the above figures are 24%, 18%, and 58%, respectively. It is seen that although QDA includes many more terms than LDA in its discriminant function, Equations (12), (13), it does not give better results. In fact, given this large number of parameters, we readily conclude that QDA does not offer a better alternative to either LDA or BLR.

From these tables we also observe that the logit model produces results very similar to probit regression. Therefore, the choice of probit versus logit depends largely on individual preferences. The same holds for the cumulative probit model. Although there are differences between the results obtained from BLR and CPR, these differences are random and do not indicate that either of these two models provides better predictions (Figure S1). Note that the results of the cumulative probit model are not improved, at least in the majority of the cases examined in this study, if we optimize the prior probabilities, P_M, P_F (Table S2).

In what concerns the ANN method, its performance strongly depends upon the number p of the nodes in the hidden layer. Best results are usually obtained for p ranging from 10 to 40. This means that the discriminant functions expressed via Equation (14) include a great number of terms, which makes their use difficult, especially if we have to publish such equations for use by other scholars. If, in addition, we compare the prediction performance of ANN with that of LDA using the combination of traits for which LDA gives the best results, we readily conclude that LDA should be preferred over ANN.

Therefore, from the classification methods studied, LDA has a small but clear advantage over the other approaches and it may be preferred in sex estimation. However, one critical question concerning the applicability of LDA is the fulfilment of the various assumptions that should be met for the method to give reliable results. The most important of these assumptions are the normality, equality of variances/covariances, and the lack of multicollinearity and outliers. Note that for BLR the only assumptions that should be taken into account are the lack of multicollinearity of the sex traits and the lack of outliers. The use of ordinal variables minimizes the existence of outliers in all methods. In R multicollinearity may be tested using the vif function of the car library. The application of this function showed that there is no multicollinearity among the sex traits except for ORB when it is examined in combination with pelvic traits. Therefore, only the results concerning the eight variables in Table 3 should be treated cautiously. In what concerns normality and equality of variances/covariances across groups, both are not valid when the predictors (sex traits) are ordinal variables. The reason is that ordinal data cannot be drawn from a multivariate Gaussian distribution and the Box's M test that is usually adopted to test the homogeneity of variance-covariance matrices, is very sensitive to violations of normality, leading to the rejection of this assumption. Note that in R the Box's M test may be performed by means of the boxM function of the biotools library and its application to the datasets under study showed that, indeed, the homogeneity assumption is violated. Thus, two basic assumptions underlying LDA application are violated. However, the improved prediction performance of LDA is a strong indication that the method is fairly robust to violations of these assumptions when classification is the primary goal. This is in line with Tabachnick and Fidell [18], who also point out that if a 95% accuracy in classification is achieved, "you hardly worry about the shape of distributions" (18; p. 381). Thus, the fact that the accuracy of LDA is comparable to that of BLR shows that this method can be applied for sex classification without tests for the underlying pattern for the data, as in the BLR and ANN methods.

Table 6 summarizes the discriminant functions of LDA that result in optimum sex classification for the Athens Collection. As expected, pelvic traits result in higher correct sex classification percentages in relation to cranial traits. It is also interesting to observe that the combination of pelvic and cranial traits improves the predictions, especially the total sex prediction, but it gives rather unbalanced results per sex as it overestimates males and underestimates females.

The combination of different sex traits is straightforward by means of the methods under study. The combination of discriminant functions already published in the literature may be achieved by means of the algorithm described in the *Statistical methods* section. Some of the results obtained are presented in Table 7. In particular, this table shows the percentage of correctly classified individuals by means of the Klales et al. [7] BLR discriminant function, various Walker's [8] BLR discriminant functions, their combination and the corresponding LDA discriminant functions of the present work. We observe that Walker's [8] equations have a very poor prediction performance, which in turn shows that cranial sex traits exhibit very strong population-specificity, as supported by previous studies [25-26]. The Klales et al. [7] discriminant function gives better predictions; the total sex classification rate is rather high, 93.2%, but sex bias is also high, -10.35%, indicating unbalanced results per sex, suggesting some population-specificity. Similar results have been obtained from Oikonomopoulou et al.

[27]. However, the most interesting finding of Table 7 is the results obtained from the combination of discriminant functions. It is seen that in most cases the total sex classification rate is greater than 98%, whereas the sex bias is decreased below 4% and these figures are practically the same with those obtained from the LDA discriminant functions derived from the Athens Collection. That is, the combination of pelvic and cranial traits via their discriminant functions removes practically any population-specificity and yields much better predictions than the individual functions, especially those obtained from the literature. This is very clearly shown in case 14, Table 7, whereby we combined the Klales et al. [7] equation with Walker's [8] equation using only the glabella and mental eminence. It can be seen that this combination achieved a correct sex classification of 98.9% in total, while the application of the individual equations achieved a correct classification in 93.2% for the pelvic traits and 75% for the cranial ones. In addition, the results of the combined function are identical to those obtained when we derive equations for sex prediction based on the Athens Collection itself (case 5, Table 7).

To summarize the above results, the differences in the classification performance between BLR, PR, CPR, LDA, QDA, ANN are overall small. However, LDA is more simple and flexible than CPR, QDA and ANN and has a small but clear advantage over BLR/PR. Despite being a parametric multivariate technique, LDA is fairly robust to violations of relevant assumptions, and may be preferred in sex classification problems. These results concern cases where pelvic and cranial traits are examined independently. The most striking result of the current study is that the proposed method to combine pelvic and cranial traits via their discriminant functions, either LDA or BLR, yields better predictions than the individual functions (correct classification rates over 98% for pooled sexes and sex bias below 3), free from population-specificity issues.

References

1. Dirkmaat D (2014) A companion to forensic anthropology. John Wiley & Sons, New York.

2. İşcan MY (2005) Forensic anthropology of sex and body size. For Sci Int 147:107-112. https://doi.org/10.1016/j.forsciint.2004.09.069

3. Spradley MK, Jantz RL (2011) Sex estimation in forensic anthropology: skull versus postcranial elements. J Forensic Sci 56:289-296. https://doi.org/10.1111/j.1556-4029.2010.01635.x

4. Buikstra JE, Ubelaker DH (1994) Standards for data collection from human skeletal remains. Arkansas Archaeological Survey Research Series 44, Fayetteville.

5. Phenice TW (1969) A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 30:297-301. https://doi.org/10.1002/ajpa.1330300214

6. Williams BA, Rogers TL (2006) Evaluating the accuracy and precision of cranial morphological traits for sex determination. J Forensic Sci 51:729-735. https://doi.org/10.1111/j.1556-4029.2006.00177.x

7. Klales AR, Ousley SD, Vollner JM (2012) A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. Am J Phys Anthropol 149:104-114. https://doi.org/10.1002/ajpa.22102

8. Walker PL (2008) Sexing skulls using discriminant function analysis of visually assessed traits. Am J Phys Anthropol 136:39–50. https://doi.org/10.1002/ajpa.20776

9. Ali Z, Cox C, Stock MK, Zandee vanRilland EE, Rubio A, Fowler DR (2018) Estimating sex using metric analysis of the scapula by postmortem computed tomography. J Forensic Sci 63:1346-1349. https://doi.org/10.1111/1556-4029.13751

10. Alunni-Perret V, du Jardin P, Nogueira L, Buchet L, Quatrehomme G (2015) Comparing discriminant analysis and neural network for the determination of sex using femur head measurements. Forensic Sci Int 253:81–87. https://doi.org/10.1016/j.forsciint.2015.05.023

11. Bidmos MA, Steinberg N, Kuykendall KL (2005) Patella measurements of South African whites as sex assessors. Homo 56:69–74. https://doi.org/10.1016/j.jchb.2004.10.002

12. Blake KA, Hartnett-McCann K (2018) Metric assessment of the pubic bone using known and novel data points for sex estimation. J Forensic Sci 63:1472-1478. https://doi.org/10.1111/1556-4029.13732

13. Clavero A, Salicrú M, Turbón D (2015) Sex prediction from the femur and hip bone using a sample of CT images from a Spanish population. Int J Legal Med 129:373-383. https://doi.org/10.1007/s00414-014-1069-y

14. Nikita E (2017) Osteoarchaeology: A guide to the macroscopic study of human skeletal remains. Academic Press, San Diego.

15. Darmawan MF, Yusuf SM, Abdul Kadir MR, Haron H (2015) Comparison on three classification techniques for sex estimation from the bone length of Asian children below 19 years old: An analysis using different group of ages. Forensic Sci Int 247:130.e1–130.e11. https://doi.org/10.1016/j.forsciint.2014.11.007

16. Du Jardin P, Ponsaillé J, Alunni-Perret V, Quatrehomme G (2009) A comparison between neural network and other metric methods to determine sex from the upper femur in a modern French population. Forensic Sci Int 192:e1–e6. https://doi.org/10.1016/j.forsciint.2009.07.014

17. Mahfouz M, Badawi A, Merkl B, Fatah EEA, Pritchard E, Kesler K, Moore M, Jantz R, Jantz L (2007) Patella sex determination by 3D statistical shape models and nonlinear classifiers. Forensic Sci Int 173:161–170. https://doi.org/10.1016/j.forsciint.2007.02.024

18. Tabachnick B, Fidell L (2012) Using multivariate statistics, 6th edition. Pearson Education Limited, Boston.

19. Konigsberg LW, Hens SM (1998) Use of ordinal categorical variables in skeletal assessment of sex from the cranium. Am J Phys Anthropol 107:97-112.

20. Johnson RA, Wichern DW (1988) Applied Multivariate Statistical Analysis, 2nd edition. Prentice Hall International, London.

21. Narsky I, Porter FC (2014) Statistical Analysis Techniques in Particle Physics: Fits, Density Estimation and Supervised Learning, 1st edition. Wiley, Hoboken, NJ.

22. Haykin S (2009) Neural networks and learning machines, 3rd edition. Prentice Hall, London.

23. Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge.

24. Eliopoulos C, Lagia A, Manolis S (2007) A modern, documented human skeletal collection from Greece. Homo 58:221-228. https://doi.org/10.1016/j.jchb.2006.10.003

25. Klales AR, Cole SJ (2017) Improving nonmetric sex classification for Hispanic individuals. J Forensic Sci 62:975-980. https://doi.org/10.1111/1556-4029.13391

26. Krüger GC, L'Abbé EN, Stull KE, Kenyhercz MW (2015) Sexual dimorphism in cranial morphology among modern South Africans. Int J Legal Med 129:869–875. https://doi.org/10.1007/s00414-014-1111-0

27. Oikonomopoulou EK, Valakos E Nikita E (2017) Population-specificity of sexual dimorphism in cranial and pelvic traits: evaluation of existing and proposal of new functions for sex assessment in a Greek assemblage. Int J Legal Med 131:1731-1738. https://doi.org/10.1007/s00414-017-1655-x

Table 1. Percentage of correctly classified individuals by means of BLR, PR, LDA and QDA based on pelvic traits

Dataset	Accuracy	BLR / PR		LDA		QDA	
		VA,SPC,MAR	VA,SPC	VA,SPC,MAR	VA,SPC	VA,SPC,MAR	VA,SPC
Original	CV (T)	94.74	96.32	96.84	96.32	96.84	95.79
Original	CV (M)	96.19	97.14	99.05	97.14	99.05	97.14
Original	CV (F)	92.94	95.29	94.12	95.29	94.12	94.12
Original	Prediction (T)	96.32	96.32	96.84	96.32	96.84	95.79
Original	Prediction (M)	96.19	97.14	99.05	97.14	99.05	97.14
Original	Prediction (F)	96.47	95.29	94.12	95.29	94.12	94.12
Training	CV (T)	96.18	96.95	96.95	96.95	96.95	96.95
Training	CV (M)	96.97	98.48	98.48	98.48	98.48	98.48
Training	CV (F)	95.38	95.38	95.38	95.38	95.38	95.38
Training	Prediction (T)	96.95	96.95	96.95	96.95	96.95	96.95
Training	Prediction (M)	96.97	98.48	98.48	98.48	98.48	98.48
Training	Prediction (F)	96.92	95.38	95.38	95.38	95.38	95.38
Target	Prediction (T)	91.53	91.53	94.92	94.92	94.92	94.92
Target	Prediction (M)	92.31	92.31	94.87	94.87	94.87	94.87
Target	Prediction (F)	90.00	90.00	95.00	95.00	95.00	95.00
Target	Sex bias	2.31	2.31	-0.13	-0.13	-0.13	-0.13

Key: T = total, M = male, F = female; sex bias is the difference between male and female percentage correct predictions.

Table 2. Percentage of correctly classified individuals by means of BLR, PR, LDA and QDA based on cranial traits

Dataset	Accuracy	BLR / PR			LDA			QDA		
		All 5 vars	GL,MA,NU	GL,MA,ORB	All 5 vars	GL,MA,NU	GL,MA,ORB	All 5 vars	GL,MA,NU	GL,MA,ORB
Original	CV (T)	88.70/89.83	89.27	89.53	89.83	89.27	88.48	89.27	88.70	89.01
Original	CV (M)	87.50/88.54	87.50	86.79	87.50	87.50	84.91	87.50	87.50	85.85
Original	CV (F)	90.12/91.36	91.36	92.94	92.59	91.36	92.94	91.36	90.12	92.94
Original	Prediction (T)	90.40	89.27	89.53	89.83	89.83	88.48	90.96	90.40	89.53
Original	Prediction (M)	89.58	87.50	86.79	87.50	87.50	84.91	88.54	87.50	86.79
Original	Prediction (F)	91.36	91.36	92.94	92.59	92.59	92.94	93.83	93.83	92.94
Training	CV (T)	90.76/91.6	90.15/91.67	90.15	92.44	92.44	91.67	90.76	91.60	92.42
Training	CV (M)	89.66	86.57/88.06	88.06	87.93	87.93	86.57	89.66	89.66	88.06
Training	CV (F)	91.80/93.44	93.85/95.38	92.31	96.72	96.72	96.92	91.80	93.44	96.92
Training	Prediction (T)	94.12	91.67	91.67/90.15	92.44	92.44	92.42	92.44	93.28	92.42
Training	Prediction (M)	91.38	88.06	88.06	87.93	87.93	88.06	89.66	89.66	88.06
Training	Prediction (F)	96.72	95.38	95.38/92.31	96.72	96.72	96.92	95.08	96.72	96.92
Target	Prediction (T)	84.48	84.75	83.05/84.75	86.21	84.48	81.36	86.21	84.48	83.05
Target	Prediction (M)	89.47	87.18	87.18/89.74	86.84	84.21	82.05	89.47	86.84	84.62
Target	Prediction (F)	75.00	80.00	75.00	85.00	85.00	80.00	80.00	80.00	80.00
Target	Sex bias	14.47	7.18	12.18/14.74	1.84	-0.79	2.05	9.47	6.84	4.62

Table 3. Percentage of correctly classified individuals by means of BLR, PR, LDA and QDA based on combinations of cranial and pelvic traits

Dataset	Accuracy	BLR / PR				LDA				QDA			
		All 8 vars	GL, VA, SPC, MAR	GL, VA, SPC	MA, VA, SPC	All 8 vars	GL, VA, SPC, MAR	GL, VA, SPC	MA, VA, SPC	All 8 vars	GL, VA, SPC, MAR	GL, VA, SPC	MA, VA, SPC
Original	CV (T)	97.16	98.95	97.89	96.8/96.3	98.86	98.95	97.89	96.84	98.30	99.47	97.37	96.84
Original	CV (M)	97.89	99.05	98.10	97.1/96.2	100.0	100.0	98.10	97.14	98.95	100.0	98.10	97.14
Original	CV (F)	96.30	98.82	97.65	96.47	97.53	97.65	97.65	96.47	97.53	98.82	96.47	96.47
Original	Prediction (T)	100.0	100.0	97.89	96.84	98.86	98.95	97.89	96.84	100.0	99.47	97.37	96.84
Original	Prediction (M)	100.0	100.0	98.10	97.14	100.0	100.0	98.10	97.14	100.0	100.0	98.10	97.14
Original	Prediction (F)	100.0	100.0	97.65	96.47	97.53	97.65	97.65	96.47	100.0	98.82	96.47	96.47
Training	CV (T)	98.3/97.5	99.24	100/99.2	100.0	100.0	98.47	98.47	99.24	98.31	98.47	98.47	99.24
Training	CV (M)	98.25	98.48	100/98.5	100.0	100.0	96.97	96.97	100.0	96.49	96.97	96.97	98.48
Training	CV (F)	98.4/96.7	100.0	100.0	100.0	100.0	100.0	100.0	98.46	100.0	100.0	100.0	100.0
Training	Prediction (T)	100.0	100.0	100.0	100.0	100.0	99.24	98.47	99.24	100.0	100.0	98.47	100.0
Training	Prediction (M)	100.0	100.0	100.0	100.0	100.0	98.48	96.97	100.0	100.0	100.0	96.97	100.0
Training	Prediction (F)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.46	100.0	100.0	100.0	100.0
Target	Prediction (T)	87.93	94.92	94.92	91.53	96.55	96.61	96.61	91.53	96.55	98.31	96.61	91.53
Target	Prediction (M)	89.47	100.0	100.0	92.31	100.0	100.0	100.0	92.31	100.0	100.0	100.0	92.31
Target	Prediction (F)	85.00	85.00	85.00	90.00	90.00	90.00	90.00	90.00	90.00	95.00	90.00	90.00
Target	Sex bias	4.47	15	15	2.31	10	10	10	2.31	10	5	10	2.31

Table 4. Percentage of correctly classified individuals when using pelvic and cranial variables by means of ANN when p = 1 and p gets its optimum value.

Dataset	Accuracy	Pelvic variables		Cranial variables		
		VA,SPC,MAR (p=1 / p=10)	VA,SPC (p=1 / p=20)	GL,MA,NU,ME,ORB (p=1 / p=15)	GL,MA,NU (p=1/ p=30)	GL,MA,ORB (p=1 / p=10)
Original	CV (T)	96.8 / 97.4	93.7 / 95.8	87.8 / 90	88.1 / 89.2	89.2 / 89.2
Original	CV (M)	100 / 100	97.1 / 97.1	86.5 / 88.5	87.7 / 87.7	87.7 / 87.7
Original	CV (F)	92.9 / 94.1	89.4 / 94.1	89.3 / 91.7	88.6 / 90.9	90.9 / 90.9
Original	Prediction (T)	97.4 / 97.4	94.7 / 96.3	90 / 91.1	89.2 / 90.2	89.2 / 89.2
Original	Prediction (M)	100 / 100	99.1 / 97.1	88.5 / 89.6	87.7 / 88.7	87.7 / 87.7
Original	Prediction (F)	94.1 / 94.1	89.4 / 95.3	91.7 / 92.9	90.9 / 92.1	90.9 / 90.9
Training	CV (T)	96.2 / 97	97 / 97	90.1 / 91.7	89.6 / 91	88.8 / 91.8
Training	CV (M)	98.5 / 98.5	98.5 / 98.5	86.2 / 86.2	85.1 / 86.6	86.6 / 88.1
Training	CV (F)	93.9 / 95.4	95.4 / 95.4	93.7 / 96.8	94 / 95.5	91 / 95.5
Training	Prediction (T)	97 / 97	97 / 97	90.9 / 93.4	91 / 91.8	91 / 91.8
Training	Prediction (M)	98.5 / 98.5	98.5 / 98.5	87.9 / 89.7	86.6 / 88.1	86.6 / 88.1
Training	Prediction (F)	95.4 / 95.4	95.4 / 95.4	93.7 / 96.8	95.5 / 95.5	95.5 / 95.5
Target	Prediction (T)	94.9 / 94.9	94.9 / 94.9	74.6 / 84.8	83.3 / 86.7	73.3 / 81.7
Target	Prediction (M)	94.9 / 94.9	94.9 / 94.9	65.8 / 84.2	87.2 / 89.7	69.2 / 84.6
Target	Prediction (F)	95 / 95	95 / 95	90.5 / 85.7	76.2 / 81	81 / 76.2
Target	Sex bias	-0.1	-0.1	-1.5	8.7	8.4

Table 5. As in Table 4 when using combinations of pelvic and cranial traits.

Dataset	Accuracy	Pelvic + cranial variables			
		All 8 vars (p=1 / p=40)	GL,MA,VA,SPC,MAR (p=1 / p=30)	GL,VA,SPC (p=1 / p=10)	MA, VA, SPC (p=1 / p=30)
Original	CV (T)	91.5 / 98.3	95.8 / 98.4	97.9 / 98.4	94.7 / 96.3
Original	CV (M)	95.8 / 100	99.1 / 100	100 / 100	97.1 / 99.1
Original	CV (F)	86.4 / 96.3	91.8 / 96.5	95.3 / 96.5	91.8 / 92.9
Original	Prediction (T)	93.8 / 98.9	96.3 / 99	98.4 / 98.4	94.7 / 96.8
Original	Prediction (M)	96.8 / 100	99.1 / 100	100 / 100	97.1 / 99.1
Original	Prediction (F)	90.1 / 97.5	92.9 / 97.7	96.5 / 96.5	91.8 / 94.1
Training	CV (T)	92.4 / 97.5	96.2 / 99.2	99.2 / 100	95.4 / 97
Training	CV (M)	94.7 / 96.5	97 / 98.5	98.5 / 100	98.5 / 98.5
Training	CV (F)	90.2 / 98.4	95.4 / 100	100 / 100	92.3 / 95.4
Training	Prediction (T)	95.8 / 98.3	96.2 / 100	100 / 100	96.2 / 97
Training	Prediction (M)	96.5 / 98.3	97 / 100	100 / 100	98.5 / 98.5
Training	Prediction (F)	95.1 / 98.4	95.4 / 100	100 / 100	93.9 / 95.4
Target	Prediction (T)	81 / 96.6	93.2 / 96.6	96.6 / 96.6	93.2 / 94.9
Target	Prediction (M)	81.6 / 100	94.9 / 100	100 / 100	94.9 / 94.9
Target	Prediction (F)	80 / 90	90 / 90	90 / 90	90 / 95
Target	Sex bias	10	10	10	-0.1

Table 6. Linear discriminant equations and percentage of correctly classified individuals.

Dataset	LDA discriminant equations	Accuracy	Total	Males	Females	Sex bias
Pelvis	y= 22.1513-3.4643*VA-2.6502*SPC-1.7547*MAR	Original	96.84	99.05	94.12	4.93
		Cross-validated	96.84	99.05	94.12	4.93
	y= 18.5549-3.6464*VA-3.1649*SPC	Original	96.32	97.14	95.29	1.85
		Cross-validated	96.32	97.14	95.29	1.85
Cranium	y= 11.7529 -2.3944*GL-0.7448*MA-0.1265*ORB-	Original	89.83	87.3	92.59	-5.29
	0.0656*ME-0.6657*NU	Cross-validated	89.83	87.3	92.59	-5.29
	y= 11.3849-2.4138*GL-0.7479*MA-0.6933*NU	Original	89.83	87.5	92.59	-5.09
		Cross-validated	89.27	87.5	91.36	-3.86
	y= 10.0779-2.414*GL-0.893*MA-0.1568*ORB	Original	88.48	84.91	92.94	-8.03
		Cross-validated	88.48	84.91	92.94	-8.03
Pelvis-cranium	y= 34.0444-2.0697*GL-1.3527*MA-	Original	98.86	100	97.53	2.47
	0.2024*ORB+0.1032*ME-0.5395*NU-3.7313*VA-	Cross-validated	98.86	100	97.53	2.47
	2.4763*SPC-1.4778*MAR	Original	98.95	100	97.65	2.35
	y= 34.3576-2.08*GL-1.6251*MA-3.8325*VA-2.4818*SPC-	Cross-validated	98.42	100	96.47	3.53
	1.7434*MAR	Original	97.89	98.1	97.65	0.45
	y= 24.8873-2.4469*GL-3.7121*VA-2.9414*SPC	Cross-validated	97.89	98.1	97.65	0.45
		Original	96.84	97.14	96.47	0.67
	y= 26.766-1.9489*MA-4.0197*VA-3.1869*SPC	Cross-validated	96.84	97.14	96.47	0.67

Table 7. Percentage of correctly classified individuals when two discriminant functions, one based on pelvic and the other on cranial traits, are combined.

No	Function	Total	Males	Females	Sex bias
	LDA functions from present work				
1	y=22.1513-3.4643*VA-2.6502*SPC-1.7547*MAR	96.59	98.95	93.83	5.12
2	y=10.3637-2.5315*GL-0.9168*MA-0.0445*ME	89.77	87.37	92.59	-5.22
3	1+2	98.86	100	97.53	2.47
4	y=7.3023-2.7024*GL+0.0104*ME	89.77	88.42	91.36	-2.94
5	1+4	98.86	100	97.53	2.47
6	y=9.8342-2.4391*GL-0.9168*MA	88.95	85.71	92.94	-7.23
7	1+6	98.95	100	97.65	2.35
8	y=5.566-1.3719*MA-0.2379*ME	74.43	87.37	59.26	28.11
9	1+8	97.73	100	95.06	4.94
	BLR functions by Klales et al. [7] and Walker [8]				
10	y=16.312-2.726*VA-1.073*SPC-1.214*MAR	93.18	88.42	98.77	-10.35
11	y= 9.128-1.375*GL-1.185*MA-1.151*ME	77.27	95.79	55.56	40.23
12	10+11	98.3	100	96.3	3.7
13	y= 7.372-1.525*GL-1.485*ME	75	94.74	51.85	42.89
14	10+13	98.86	100	97.53	2.47
15	y= 7.434-1.568*GL-1.459*MA	81.58	97.14	62.35	34.79
16	10+15	98.42	100	96.47	3.53
17	y= 7.382-1.629*MA-1.415*ME	58.52	88.42	23.46	64.96
18	10+17	97.73	100	95.06	4.94