8-22-2018

Of Typicality and Predictive Distributions in Discriminant Function Analysis

Lyle W. Konigsberg
Department of Anthropology, University of Illinois at Urbana–Champaign, lylek@illinois.edu

Susan R. Frankenberg
Department of Anthropology, University of Illinois at Urbana–Champaign

Of Typicality and Predictive Distributions in Discriminant Function Analysis

Lyle W. Konigsberg[1]* and Susan R. Frankenberg[1]

[1]Department of Anthropology, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA.

*Correspondence to: Lyle W. Konigsberg, Department of Anthropology, University of Illinois at Urbana–Champaign, 607 S. Mathews Ave, Urbana, IL 61801 USA. E-mail: lylek@illinois.edu.

Short Title: Typicality and Predictive Distributions in Discriminant Functions

KEY WORDS: ADMIXTURE, POSTERIOR PROBABILITY, BAYESIAN ANALYSIS, OUTLIERS, TUKEY DEPTH

Abstract

While discriminant function analysis is an inherently Bayesian method, researchers attempting to estimate ancestry in human skeletal samples often follow discriminant function analysis with the calculation of frequentist-based typicalities for assigning group membership. Such an approach is problematic in that it fails to account for admixture and for variation in why individuals may be classified as outliers, or non-members of particular groups. This paper presents an argument and methodology for employing a fully Bayesian approach in discriminant function analysis applied to cases of ancestry estimation. The approach requires adding the calculation, or estimation, of predictive distributions as the final step in ancestry-focused discriminant analyses. The methods for a fully Bayesian multivariate discriminant analysis are illustrated using craniometrics from identified population samples within the Howells published data. The paper also presents ways to visualize predictive distributions calculated in more than three dimensions, explains the limitations of typicality measures, and suggests an analytical route for future studies of ancestry and admixture based in discriminant function analysis.

Discriminant function analysis is often used within forensic anthropology and bioarchaeology to estimate the sex of an individual or the ancestral group or groups to which the individual might belong based on skeletal morphology (Bidmos and Asala 2003; Bidmos and Dayal 2004; Calcagno 1981; Dibennardo and Taylor 1983; Giles and Elliot 1963; Henke 1977; Johnson et al. 1989; Kajanoja 1966; Konigsberg et al. 2009; Long 1966; Ousley and Jantz 1996; Ousley and Jantz 2012; Ousley and Jantz 2013; Ousley et al. 2009; Rightmire 1970; Šlaus and Tomicic 2005; Snow et al. 1979; Ubelaker et al. 2002; Walker 2008; Zakrzewski 2007). We use the term "group" here to represent a category within a classification. These categories may be synonymous with populations, but generally, the sampling is too incomplete to make an equivalence with populations. At least in its initial step, discriminant function analysis is inherently a Bayesian method in that it uses likelihoods to update prior probabilities of group membership and thus generate posterior probabilities of group membership. In the sex estimation setting, these posterior probabilities can be interpreted only as posterior probabilities that the individual is male or female. But in the case of ancestry assessment, posterior probabilities can also be interpreted as mixing proportions or admixture estimates (Algee-Hewitt 2016; Algee-Hewitt 2017).

While discriminant function analysis is inherently Bayesian, and researchers use posterior probabilities to guide their choice of potential groups, they then go on to calculate typicalities (Elliott and Collard 2009; Geller and Stojanowksi 2017; Guyomarc'h and Bruzek 2011; Kallenberger and Pilbrow 2012; Urbanová et al. 2014), which are inherently frequentist. Typicalities are the probabilities of obtaining a Mahalanobis squared distance as large as, or larger than, an observed distance if the case in question was indeed randomly sampled from the multivariate normal distribution representing a given population. While the calculation of

typicalities following on the calculation of posterior probabilities may make sense in some settings, it can be counter-productive for assessing ancestry when admixture is a very real possibility. In the admixture setting, the posterior probabilities can be interpreted as admixture estimates rather than as posterior probabilities of group membership. Given that the reference populations in most discriminant function analyses are selected so that they are not admixed, it is quite possible for an individual to have high posterior probabilities of group membership with the two or more groups from which their ancestry is derived while at the same time being atypical for each of these groups.

We argue here that rather than abandoning the Bayesian paradigm by calculating frequentist typicalities as the final step in estimating ancestry, one should instead employ a fully Bayesian analysis. The full Bayesian analysis calculates not only the posterior probabilities of group membership, but also the predictive distributions for the skeletal metrics, which in our case are craniometrics from the W.W. Howells' dataset. If the skeletal metrics are not well characterized by the predictive distributions then the model used is suspect. Alternatively, if the skeletal metrics are well characterized, this lends credence to the model and suggests that the individual may be admixed from the groups with high posterior probabilities. We suggest adding the calculation of predictive distributions as the final step in ancestry-focused discriminant analyses, especially when the posterior probability of an individual belonging to one particular group is not high. This paper presents and illustrates methods for carrying out a fully Bayesian approach in multivariate discriminant analysis, and for visualizing predictive distributions calculated in more than three dimensions. It also details the limitations of typicality measures in estimating ancestry under conditions of admixture.

Materials and Methods

Samples.

This study uses eight craniometric variables (see Table 1) from 29 local samples measured by W.W. Howells (1973; 1989). We use the term "sample" to refer to a collection of individuals from a restricted geographic location. Howells used the term "population" to refer to such samples, but it is clear that these collections of individuals should not be considered as populations. Howells originally had 30 local samples, but two of these, North and South Maori, were each comprised of only ten individuals each. We have combined the Maori samples into a single sample of twenty, which is still the smallest local sample among the Howells data. The next larger sample, Chinese from Anyang, has 42 individuals. The eight craniometric variables were converted to Darroch and Mosimann (Darroch and Mosimann 1985; Jungers et al. 1995) shape variables on the raw scale. In other words, each variable was divided by the geometric mean of the eight variables for that individual. This "size correction" was used so that males and females could be considered together in all further analyses. As sex is only known for a portion of Howells' sample, converting the data to shape variables to reduce the effect of sexual dimorphism is more conservative then "sex correcting" (Relethford 1994) the data. We then calculated the Mahalanobis squared distance between each of the 29 local samples using the pooled within-group covariance matrix. All calculations were done in the statistics and graphics program "R" (R Core Team 2017).

Posterior Probabilities.

To simplify, we assume that only two samples are under consideration. We further assume that the covariance matrices $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$ are equal so that we can write $\mathbf{\Sigma}$ for both matrices. Letting

$\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$ be the vectors of means for group *A* and group *B* and **x** be the vector of shape variables for a case not contained in group *A* or group *B*, the Mahalanobis squared distances are $D_A^2 = (\mathbf{x}-\boldsymbol{\mu}_A)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_A)$ and $D_B^2 = (\mathbf{x}-\boldsymbol{\mu}_B)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_B)$. If p_A is the prior probability that the new case is from group *A*, then the posterior probability that the case is from group *A* is:

$$p(g = A|\mathbf{x}) = \frac{p_A \times \exp\left(-\frac{1}{2}D_A^2\right)}{p_A \times \exp\left(-\frac{1}{2}D_A^2\right) + (1-p_A) \times \exp\left(-\frac{1}{2}D_B^2\right)}. \quad (1)$$

If the covariance matrices are not equal, then $D_A^2 = (\mathbf{x}-\boldsymbol{\mu}_A)'\boldsymbol{\Sigma}_A^{-1}(\mathbf{x}-\boldsymbol{\mu}_A)$ and $D_B^2 = (\mathbf{x}-\boldsymbol{\mu}_B)'\boldsymbol{\Sigma}_B^{-1}(\mathbf{x}-\boldsymbol{\mu}_B)$ and the posterior probability that the case is from group *A* is:

$$p(g = A|\mathbf{x}) = \frac{p_A \times |\Sigma_A|^{-1/2} \times \exp\left(-\frac{1}{2}D_A^2\right)}{p_A \times |\Sigma_A|^{-1/2} \times \exp\left(-\frac{1}{2}D_A^2\right) + (1-p_A) \times |\Sigma_B|^{-1/2} \times \exp\left(-\frac{1}{2}D_B^2\right)}. \quad (2)$$

Typicality

The typicality for a new case **x** against group *A* is found from the *F* distribution as:

$$\frac{D_A^2\left(N_A\left(N_A - t\right)\right)}{t\left(N_A^2 - 1\right)} \sim F_{(t, N_A - t)}, \quad (3)$$

where *t* is the number of traits and N_A is the sample size for group *A* (Krzanowski 2000: p. 213; Mason and Young 2002: eqn. 2.14). Equation (3) gives the probability density function, while the "typicality probability" (Campbell 1984; Huberty 1984; McKay and Campbell 1982) is one minus the distribution function. The typicality for group *B* substitutes the Mahalanobis squared distance to group *B* and the sample size for group *B*. The Mahalanobis distances can be calculated using the group specific covariance matrices as we have done in equations (1) and (2), though some software uses the pooled covariance (Ousley and Jantz 2005; Ousley and Jantz

2013, and see the helpfile at

http://math.mercyhurst.edu/~sousley/Fordisc/Help/Fordisc3_Help.pdf). This typicality probability is specifically for an independent (i.e., new) case, as opposed to what could be considered an outlier probability. An outlier probability is the typicality probability for a case which is known to be from the sample that generated the summary statistics. The outlier probability is obtained from the following probability density function:

$$\frac{D_A^2\left(N_A - t + 1\right)}{N_A\, t} \sim F_{(t, N_A - t + 1)}, \tag{4}$$

as first given in Hotelling (1951: p. 25) . This F test has also been given as an equivalent incomplete Beta distribution test in Algee-Hewitt (2016: eqn. 8), Mason and Young (2002: eqn. 2.15), and Aitchison and Dunsmore (1975: eqn. 11.20).

To complicate matters, Hawkins (1981) gave an F test similar to that obtained from equation (4), but which deletes the case in question from calculation of the vector of means and the covariance matrix. For a single group *A*, Hawkins' F test is obtained from:

$$\frac{D_A^2\left(N_A - t - 1\right)}{t\left(\left(N_A - 1\right)^2 - N_A D_A^2\right)} \sim F_{(t, N_A - t - 1)}. \tag{5}$$

This alternative F test is incorporated in the R library MissMech (Jamshidian et al. 2014) as the object "Hawkins." Konigsberg and others (2009) used equation (5) when equation (3) would have been more appropriate. In this paper we use equation (3) as the appropriate F test for calculating a typicality probability value when the case in question is not from any of the reference groups. In Appendix 1 we use simulation to show that equation (3) does indeed provide the correct calculation of typicality.

Discriminant Models between the Linear and the Quadratic.

The methods so far have described discriminant analysis as being either linear, with equal covariance matrices, or as quadratic, with different covariance matrices. Bensmail and Celeux (1996) have described an additional six models that have more parameters than the linear model but fewer parameters than the quadratic model. These additional models contain various constraints on the eigenvalue decompositions of the covariance matrices. The covariance matrices are written as $\lambda\mathbf{DAD}'$, where λ is the t^{th} root of the determinant of a given covariance matrix, **D** contains the eigenvectors of the covariance matrix, and **A** is a diagonal matrix with a determinant equal to 1.0. The λ parameter is a volume parameter, the **A** matrix controls the shape of the ellipsoids, and the **D** matrix controls their orientations. The parameterization $\lambda\mathbf{DAD}'$ is the linear discriminant model where the covariance matrices are equal while $\lambda_K\mathbf{D}_K\mathbf{A}_K\mathbf{D}_K^{'}$ is the quadratic discriminant model where each of the *K* covariance matrices is unique. Table 2 contains the notation from Bensmail and Celeux's (1996) Table 1 for the eight models considered here. Bensmail and Celeux also consider six more models where the traits are uncorrelated, but we do not consider these models here as craniometrics typically are correlated. For ease of reference we number the models from one to eight. The second model is one where the covariance matrices are proportional. Manly and Rayner (1987) and Flury et al. (1994) have referred to this as the "proportional matrices" model. The fourth model has equal correlation matrices. Manly and Rayner (1987) referred to this as the "equal correlations" model while Flury et al. (1994) referred to this as the "common principal components" model, or the CPC model for short. Models three and five through seven do not have names in the literature.

The models in Table 2 can be fit using MclustDA in the library mclust (Scrucca et al. 2016) within R. This is specifically done using the model "EDDA." The acronym stands for "eigenvalue decomposition discriminant analysis," the name given by Bensmail and Celeux

(1996). Table 2 lists the three-letter model names given in mclust for each of the eight models; each model is named according to whether the **V**olume, **S**hape and **O**rientation covariance matrices are variable ("V") or equal ("E"). While Bensmail and Celeux used cross-validation with parsimony to choose among tied models, mclust picks the model with the highest Bayesian Information Criterion (BIC). The BIC is two times the log-likelihood for a given model minus the product of the number of parameters in the model and the log of the total sample size. Table 2 lists the number of parameters for the covariance matrices by individual volume, shape and orientation components, and for the total where *G* is the number of groups. All models also contain $G \times t$ parameters for the vector of means within each group.

Predictive Distributions.

If $p\left(g = A|\mathbf{x}\right)$ is the posterior probability that a given case belongs to group *A* and $1 - p\left(g = A|\mathbf{x}\right)$ is the posterior probability that the case belongs to group *B*, then the predictive distribution for the measurements is proportional to:

$$p\left(g = A|\mathbf{x}\right) \times MVN\left(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A\right) + \left(1 - p\left(g = A|\mathbf{x}\right)\right) \times MVN\left(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B\right), \quad (6)$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_A = \boldsymbol{\Sigma}_B$ if the covariance matrices for the two groups are equal. Equation (6) can be converted to a proper probability density function by integrating across the *t* traits and dividing equation (6) by the integral. Alternatively, and more simply, the probability density function can be approximated by simulating out of the two multivariate normal distributions in the proportions given by the posterior probabilities. Note that equation (6) is a simplification of what is known as the "posterior predictive check" (Gelman et al. 2014; Kruschke 2015). A complete posterior predictive check accounts for uncertainty in the measurements once the

parameters are known as well uncertainty in the parameters themselves. We ignore the uncertainty in the parameters in order to simplify the problem. If the case in question can reasonably be considered to have been formed by admixture from groups *A* and *B*, then the measurements for the individual should fall within the predictive distribution. It is not immediately clear how to measure the extent to which a t-dimensional point falls within a mixture of two or more t-dimensional multivariate normal distributions. There are a number of options available in the literature for examining the centrality of a point within a multivariate distribution (Nolan 1999; Small 1990; Zuo and Serfling 2000), but their effectiveness for the problem at hand has yet to be evaluated. We use Tukey's (1975) half-space depth to examine the centrality of a given cranium within the predictive distribution because this commonly used nonparametric method can be easily visualized in low dimensions.

Tukey's Half-Space Depth.

Tukey's half-space depth is easiest to understand in one and two dimensions, and there are graphical methods specifically intended for the bivariate case. Tukey's (1975:529) definition for the one dimensional case with the ordered variate *y* is:

> depth of $y_{i|n}$ = the lesser of *i* or $n+1-i$ = result of counting in from the nearest end.

For example, if $n=51$ then the depth for the leftmost (smallest) value is 1 while the depth for the rightmost (largest value) is also 1. The depth for the second value is 2 as is the depth for the 50th value. The depth for the third and 49th values is 3, and so on. The 26th point has a depth of 26 from both the leftmost and rightmost value and is the median for the 51 numbers.

One of the bivariate methods based on Tukey's half-space depth is known as a "bagplot" (Rousseeuw et al. 1999) or "sunburst" plot (Liu et al. 1999) and is a bivariate generalization of the familiar box and whiskers plot. The other method, a depth contour plot (Ruts and Rousseeuw 1996), is the one we will use here to illustrate the concept of half-space depth for two dimensional problems. Later we will use the "bagplot" as it does provide a view similar to the box and whiskers plot. Depth contours can be drawn in the R package "depth" (Genest et al. 2017) which also estimates the depth of multidimensional points as well as the "depth median." Figure 1 shows a small example of simulation from a bivariate mixture model. The axes are labelled BNL (basion-nasion length) and AUB (biauricular breadth) as this is a small example of a larger analysis we will present in the results section. The simulation for Figure 1 contains only 50 points. The outer convex hull has a Tukey half-space depth of one. As with the one-dimensional problem, the half-space depth is the minimum number of points to either side of a line struck through the point in question, where the point on the line is counted in the total. Each point at the vertices of the outer convex hull has a half-space depth of one with 49 points to the other side of the line, one point on the line, and no points to the other side of the line. The inner convex hull shown in Figure 1 has a Tukey half-space depth of five, as can be determined by counting the asterisked points on and above the heavy line struck through the point marking a vertex on the inner hull. The deepest point in Figure 1, marked with a circle, has a depth of 22 and the dashed line struck through this point has 21 points to the right plus the 22nd point on the line.

Figure 2 shows an outer convex hull for the Tukey half-space depth of two for the same set of points as in Figure 1. The inner hull here is the 22-depth hull. Figure 2 also demonstrates that vertices on convex hulls deeper than the one-deep contour need not contain observed points.

Such vertices can be constructed by the intersection of two lines each struck through two observed points. The dashed lines in Figure 2 show one such intersection that defines the 2-depth hull. The heavy line through that vertex shows the two points (marked with asterisks) above the heavy line. The 22-depth hull contains one observed point and is the deepest convex hull for these 50 points. This median region summarized by a single point is the center of gravity for the vertices of the deepest hull. Figure 3 shows in greater detail how the 22-depth (median) hull, filled with light gray, is constructed. The intersections of the lighter-weight lines are the vertices for the 22-depth hull. Note the 22 points below the heavy dashed line struck through one vertex of this hull.

Bagplots are created by finding the median (and therefore deepest) region and then drawing the deepest hull that contains 50% of the points. This particular hull constitutes the "bag," and is analogous to the quartiles that form the box of a box and whiskers plot. The "fence" of a bagplot is not drawn, but consists of a three-fold expansion of the bag relative to the median. The "loop" is drawn and consists of a convex hull containing the least deep points that are still within the fence. Outliers are then any points lying outside the loop. "Whiskers" can be drawn from the bag to each point between the bag and the loop. A "sunburst" plot is identical to a bagplot with whiskers but without the loop drawn. The bagplots we present later in this paper were drawn in R using the package "aplpack" (Wolf and Bielefeld 2014).

Pairwise comparisons of the eight craniometrics used here result in 28 unique bagplots, too many to visually inspect. As a better alternative, Tukey's half-space depth can be generalized to three and higher dimensions by multiplying the $n \times t$ matrix of points by $t \times 1$ unit vectors and finding the smallest half-space depth along the variously produced lines. Unlike the bivariate and trivariate cases which have exact solutions, the higher dimensional problems can

only be solved approximately by taking a large number of random direction vectors. It is then possible to compare the depth of a given cranium in the cloud of points to the depth of the median in that cloud. If these depths are similar, then the cranium has "centrality" in the cloud of points representing the predictive distribution. We can also plot the multidimensional median from the predictive distribution against the actual cranium (defined by eight measurements) to assess similarity.

Results

Our initial analysis of the 29 local population samples from the Howells data produced 406 unique Mahalanobis D-squared distances based on the pooled within-group covariance matrix and the 29 vectors of means. Rather than present this sizeable matrix, we instead present a plot of the first two principal coordinates of the distance matrix, accounting for 54.8% of the variation, in Figure 4, and a plot of the first three principal coordinates, accounting for 71.8%, in Figure 5. The largest Mahalanobis squared distance is between Easter Islanders and the Buriat. We consequently focus on these two groups for more detailed distance analyses.

Analysis of Buriat and Easter Islanders under the Assumption of Homoscedasticity.

We begin with the simplifying assumption that the within-group covariance matrices for the Buriat and Easter Islanders are equal, which as we will see in the next section, is not supportable. We start with this assumption in order to clarify the use of the terms "internal outlier" and "external outlier" (Huberty and Olejnik 2006: p. 405-406). To form an internal outlier we take the simple average of the Buriat and Easter Islander centroids. This produces an individual who, beginning from equal prior probabilities of group membership, will have equal posterior

probabilities of group membership. For an external outlier, we choose individual number 703 from the Howells data. This individual is a Zulu with a posterior probability of group membership from the Buriat of 0.32 and of group membership from the Easter Islanders of 0.68 (after having started with equal priors). Figure 6 shows a two-dimensional principal coordinates plot of the Buriat and Easter Islanders, of their average, and of the Zulu individual. To clarify the distributions, the Buriat and Easter Islanders are shown with their respective convex hull "peels" (Green 2006). This Figure clearly shows that the average of the two centroids is an internal outlier as it "is located 'in-between' two groups" (Huberty and Olejnik 2006:405) while the Zulu individual is an external outlier as the individual is "located 'outside' the groups" (Huberty and Olejnik 2006:406).

Analysis of Buriat and Easter Islanders by Eigenvalue Decomposition Discriminant Analysis (EDDA).

The discriminant model that yielded the highest BIC from MclustDA within the package mclust was model 6 (see Table 2). This model has equal distributional shapes for the Buriat and Easter Islanders but different orientations and volumes. Table 3 lists the λ parameters, the diagonal of the **A** matrix, the **D** matrices for both groups, and the vectors of means for both groups. From these parameters the covariance matrices can be calculated and equation (2) can be used to find the posterior probabilities of group membership for each of the 195 individuals. We made these calculations under equal priors for group-membership. This resulted in 100% correct classification. As an example of an external outlier, we will use the Zulu individual from the previous analysis. Under the current model with equal priors, this individual has a posterior probability of group-membership with the Buriat of 0.676 and a posterior probability of group-

membership with the Easter Islanders of $1-0.676=0.324$. The typicality probabilities of this individual with both groups are far less than 0.0001.

The internal outlier, or more exactly an individual with equal posterior probabilities of group-membership with the Buriat and Easter Islanders, is now not a simple average of the two group centroids. Instead, we must solve the equation below for *m*:

$$\begin{aligned}&\mathbf{x}=m\times\boldsymbol{\mu}_A+(1-m)\times\boldsymbol{\mu}_B\\&-0.5\ln|\boldsymbol{\Sigma}_A|-0.5D_A^2=-0.5\ln|\boldsymbol{\Sigma}_B|-0.5D_B^2,\end{aligned}\tag{7}$$

where *m* is the admixture rate from the Buriat, *A* and *B* refer respectively to the Buriat and Easter Islanders, and the Mahalanobis squared distances are for the vector **x** against the two groups. Solving for *m* gives an admixture rate from the Buriat of 0.442 and from Easter Islanders of $1-m=0.558$. Given that we solved for equal posterior probabilities, this hypothetical individual (or "synthetic skull") has equal posterior probabilities of being from both groups. This individual's typicality probabilities against both groups are, like the Zulu individual, far below 0.0001.

While equation (7) gives the admixture rates for a synthesized internal outlier with equal posterior probabilities of group-membership from the Buriat and Easter Islanders, we need to solve a different equation to find a "synthetic skull" that yields posterior probabilities equal to those from the Zulu skull but that remains an internal outlier. Solving the following equation:

$$\begin{aligned}&\mathbf{x}=m\times\boldsymbol{\mu}_A+(1-m)\times\boldsymbol{\mu}_B\\&\frac{\exp\left(-0.5\ln|\boldsymbol{\Sigma}_A|-0.5D_A^2\right)}{\exp\left(-0.5\ln|\boldsymbol{\Sigma}_A|-0.5D_A^2\right)+\exp\left(-0.5\ln|\boldsymbol{\Sigma}_B|-0.5D_B^2\right)}-0.6762=0,\end{aligned}\tag{8}$$

for *m* gives admixture proportions of 0.444 from the Buriat and $1\text{-}0.444=0.556$ from Easter Islanders. These admixture proportions yield a synthetic skull with the same posterior

probabilities as for the Zulu skull (0.676 and 0.324 from the Buriat and Easter Islanders, respectively) and will be used to form predictive distributions.

Predictive Distributions.

Figure 7 shows the basion-nasion length and biauricular breadth "bagplot" for 10,000 simulations with mixing proportions of 0.444 and 0.556 from the Buriat and Easter Islanders, respectively. These mixing proportions give the posterior probabilities of group-membership for the Zulu individual (0.676 and 0.324 from the Buriat and Easter Islanders, respectively) mentioned above. Note that the Zulu cranium is an outlier falling beyond the loop of the bagplot. If we consider all of the bagplots generated using the 28 unique pairs of variables, the Zulu cranium falls within the bag hull only seven times, falls outside the bag but within the loop 20 times, and falls outside the loop the one time shown in Figure 7. That the Zulu cranium is not centrally located can be found by assessing the multivariate depth for this individual. The cranium has an estimated eight-dimensional depth of 0, indicating that it falls entirely outside of the cloud of points. Figure 8 compares the estimated depth median measurements with the actual measurements of the Zulu cranium. The median has an estimated depth of 3,616 out of 10,000 simulated crania. Note that the median and the actual cranium are quite dissimilar.

Figure 9 shows the same bagplot as in Figure 7, but this time with mixing proportions of 0.442 from the Buriat and 0.558 from Easter Islanders, and a "synthetic skull" formed using these mixing proportions as admixture proportions. These are the mixing proportions calculated from equation (8), the equation for an internal outlier with equal posterior probabilities of group-membership. Note that the plotted "synthetic skull" is well within the bag and is very close to the depth median. In fact, for all 28 bagplots it is within the bag and is also close to the "center"

(depth median). The "synthetic skull" has an estimated eight-dimensional depth of 3,777 out of the 10,000 simulated cranium. This is deeper than the estimated median depth of 3,481, a consequence of the fact that Tukey depths can only be estimated and not solved for numerically beyond three dimensions. Figure 10 shows the similarity between the median depth and this particular synthesized case.

Discussion

This paper has shown that the simple procedure of following calculations of posterior probabilities in discriminant function analyses with calculations of typicalities has potential problems. The use of typicalities to decide on ancestral group assignment can fail on two fronts. First, typicality values may not meet the first assumption of what Aitchison and Dunsmore (1975) refer to in their chapter 11 as "diagnosis." In discussing the statistical basis for differential diagnoses of diseases, their (Aitchison and Dunsmore 1975:215) first assumption is that "Each case belongs to one and only one of a finite set $T = \{1, ..., r\}$ of possible *types*." In admixture analyses, including the calculation of typicalities, this exclusivity does not necessarily hold, so that the ultimate goal of assigning a case to a single group is unrealistic. Second, the calculation of typicalities does not differentiate between "internal outliers" and "external outliers" (Huberty and Olejnik 2006: p. 405-406). Given that admixture analyses are typically carried out on descendant individuals in comparison to two or more ancestral populations with little to no admixture, it is quite possible to have admixed individuals who are "internal outliers" and are consequently not well represented by any of the non-admixed ancestral populations. It is in this context that the calculation of predictive distributions is particularly useful. The method maintains the Bayesian approach first used in calculating posterior probabilities of group-

membership and allows us to treat these parameters as estimates of admixture rather than as probabilities of exclusive membership in a group.

For future analyses we suggest the following analytical route. First, posterior probabilities of group membership and typicalities should be calculated. In some analyses this may lead to unambiguous results. For example, if a case is found to have a high posterior probability of group membership for one group and is also shown to be "typical" for that group, then it is reasonable to deduce that the case in question derives from the given group. But in more ambiguous cases, and especially if the analysis began with a linear discriminant function, then the first few principal coordinates of the distance matrix or the R matrix should be plotted (as in our Figure 8). Alternatively, the position of groups and the case in question can be plotted on the first few canonical variates. If the posterior probabilities of group membership are high for some number of groups and the individual represents someone with admixture from those given groups, then the individual should show "centrality" for those groups in the plot of principal coordinates or canonical variates. This graphical approach can be found in Corruccini et al.'s (1982) Figure 10, Hallgrímsson et al.'s (2004) Figure 2, Koehl and Long's (2018) Figure 3, Stull et al.'s (2014) Figures 2, 3, and 6, and Wijsman and Neves' (1986) Figure 2.

When the analysis has been based on a non-linear discriminant function then plotting an ordination of the distances is not a possibility. In this setting when the posterior probabilities and typicalities do not unambiguously assign an individual to one group, we recommend using simulation to build the predictive distributions for the measurements. This simulation should be done using the mixing proportions that recreate the posterior probabilities of the observed case. Bagplots can then be used to see if the individual is an "internal outlier" who falls within the bag of the bagplot. For a full multivariate analysis the Tukey depth of the actual individual within

the cloud of simulated points can be used to determine whether the individual is an internal outlier or an external outlier. If the individual is an internal outlier, then it is reasonable to deduce that said individual is the result of admixture between the groups with high posterior probabilities of group membership.

Acknowledgements We thank Drs. Bridget F.B. Algee-Hewitt and Jieun Kim for their invitation to participate in the AAPA poster symposium: "Thinking Computationally about Forensics" that led to the current paper. We also thank the anonymous reviewers for their comments that contributed substantially to the revisions of our original manuscript.

Received 29 January 2018; revision accepted for publication 24 April 2018.

Literature Cited

Algee-Hewitt, B. F. 2016. Population inference from contemporary American craniometrics. *Am. J. Phys. Anthropol.* 160:604–624.

Algee-Hewitt, B. F. 2017. Geographic substructure in craniometric estimates of admixture for contemporary American populations. *Am. J. Phys. Anthropol.* 164:260–280.

Bensmail, H., and G. Celeux. 1996. Regularized Gaussian discriminant analysis through eigenvalue decomposition. *J. Am. Stat. Assoc.* 91:1,743–1,748.

Bidmos, M. A., and S. A. Asala. 2003. Discriminant function sexing of the calcaneus of the South African whites. *J. Forensic Sci.* 48:1,213–1,218.

Bidmos, M. A., and M. R. Dayal. 2004. Further evidence to show population specificity of discriminant function equations for sex determination using the talus of South African Blacks. *J. Forensic Sci.* 49:1,165–1,170.

Calcagno, J. M. 1981. On the applicability of sexing human skeletal material by discriminant function analysis. *J. Hum. Evol.* 10:189–198.

Campbell, N. 1984. Some aspects of allocation and discrimination. In *Multivariate Statistical Methods in Physical Anthropology*, G. N. Van Vark and W. W. Howells, eds. Hingham, MA: Kluwer Academic Publishers, 177–192.

Corruccini, R. S., J. S. Handler, R. J. Mutaw et al. 1982. Osteology of a slave burial population from Barbados, West Indies. *Am. J. Phys. Anthropol.* 59:443–459.

Darroch, J. N., and J. E. Mosimann. 1985. Canonical and principal components of shape. *Biometrika* 72:241–252.

Dibennardo, R., and J. V. Taylor. 1983. Multiple discriminant function analysis of sex and race in the postcranial skeleton. *Am. J. Phys. Anthropol.* 61:305–314.

Elliott, M., and M. Collard. 2009. FORDISC and the determination of ancestry from cranial measurements. *Biol. Lett.* 5:849–852.

Flury, B. W., M. J. Schmid, and A. Narayanan. 1994. Error rates in quadratic discrimination with constraints on the covariance matrices. *J. Classif.* 11:101–120.

Geller, P. L., and C. M. Stojanowksi. 2017. The vanishing Black Indian: Revisiting craniometry and historic collections. *Am. J. Phys. Anthropol.* 162:267–284.

Genest, M., J.-C. Masse, and J.-F. Plante. 2017. *Depth: Nonparametric depth functions for multivariate analysis. R package version 2.1-1.* https://CRAN.R-project.org/package=depth.

Giles, E., and O. Elliot. 1963. Sex determination by discriminant function analysis of crania. *Am. J. Phys. Anthropol.* 21:53–68.

Green, P. J. 2006. Peeling data. In *Encyclopedia of Statistical Sciences* (e-book version) (2nd ed.), S. Kotz, C. B. Read, N. Balakrishnan, and B. Vidakovic, eds. Hoboken, NJ: Wiley-Interscience. Retrieved from Wiley Online Library: http://www.library.illinois.edu/proxy/go.php?url=http://dx.doi.org/10.1002/0471667196.

Guyomarc'h, P., and J. Bruzek. 2011. Accuracy and reliability in sex determination from skulls: A comparison of Fordisc® 3.0 and the discriminant function analysis. *Forensic Sci. Int.* 208:180.e1–180.e6.

Hallgrímsson, B., B. Ó Donnabháin, G. B. Walters et al. 2004. Composition of the founding population of Iceland: Biological distance and morphological variation in early historic Atlantic Europe. *Am. J. Phys. Anthropol.* 124:257–274.

Hawkins, D. M. 1981. A new test for multivariate normality and homoscedasticity. *Technometrics* 23:105–110.

Henke, W. 1977. On the method of discriminant function analysis for sex determination of the skull. *J. Hum. Evol.* 6:95–100.

Hotelling, H. 1951. A generalized T test and measure of multivariate dispersion. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, ed. Berkeley, CA: University of California Press, 23–41.

Huberty, C. J. 1984. Issues in the use and interpretation of discriminant analysis. *Psychol. Bull.* 95:156–174.

Jamshidian, M., S. J. Jalal, and C. Jansen. 2014. Missmech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (mcar). *J. Stat. Softw.* 56:1–31.

Johnson, D., P. O'Higgins, W. Moore et al. 1989. Determination of race and sex of the human skull by discriminant function analysis of linear and angular dimensions. *Forensic Sci. Int.* 41:41–53.

Jungers, W. L., A. B. Falsetti, and C. E. Wall. 1995. Shape, relative size, and size-adjustments in morphometrics. *Yearb. Phys. Anthropol.* 38:137–161.

Kajanoja, P. 1966. Sex determination of Finnish crania by discriminant function analysis. *Am. J. Phys. Anthropol.* 24:29–33.

Kallenberger, L., and V. Pilbrow. 2012. Using CRANID to test the population affinity of known crania. *J. Anat.* 221:459–464.

Kaplan, E. L., and P. Meier. 1958. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53:457–481.

Koehl, A. J., and J. C. Long. 2018. The contributions of admixture and genetic drift to diversity among post-contact populations in the Americas. *Am. J. Phys. Anthropol.* 165:256–268.

Konigsberg, L. W., B. F. Algee-Hewitt, and D. W. Steadman. 2009. Estimation and evidence in forensic anthropology: Sex and race. *Am. J. Phys. Anthropol.* 139:77–90.

Krzanowski, W. J. 2000. *Principles of Multivariate Analysis: A User's Perspective* (Rev. ed.). New York: Oxford University Press.

Liu, R. Y., J. M. Parelius, and K. Singh. 1999. Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion and a rejoinder by Liu and Singh). *Ann. Stat.* 27:783–858.

Long, J. K. 1966. A test of multiple-discriminant analysis as a means of determining evolutionary changes and intergroup relationships in physical anthropology. *Am. Anthropol.* 68:444–464.

Manly, B. F., and J. Rayner. 1987. The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika* 74:841–847.

Mason, R. L., and Young, J. C. 2002. *Multivariate Statistical Process Control with Industrial Applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

McKay, R., and N. Campbell. 1982. Variable selection techniques in discriminant analysis: II. Allocation. *Br. J. Math. Stat. Psychol.* 35:30–41.

Nolan, D. 1999. On min–max majority and deepest points. *Stat. Probab. Lett.* 43:325–333.

Ousley, S. D., and R. L. Jantz. 1996. *FORDISC 2.0*. Knoxville, TN: Department of Anthropology.

Ousley, S. D., and R. L. Jantz. 2005. *FORDISC 3.1: Computerized Forensic Discriminant Functions*. Knoxville, TN: University of Tennessee.

Ousley, S. D., and R. L. Jantz. 2012. Fordisc 3 and statistical methods for estimating sex and ancestry. In *A Companion to Forensic Anthropology,* D. C. Dirkmatt, ed. London: Wiley-Blackwell, 311–329.

Ousley, S. D., and R. L. Jantz. 2013. Fordisc 3: The third generation of computer-aided forensic anthropology. *Rechtsmedizin* 23:97–99.

Ousley, S. D., R. L. Jantz, and D. Freid. 2009. Understanding race and human variation: Why forensic anthropologists are good at identifying race. *Am. J. Phys. Anthropol.* 139:68–76.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Relethford, J. H. 1994. Craniometric variation among modern human populations. *Am. J. Phys. Anthropol.* 95:53–62.

Rightmire, G. P. 1970. Bushman, Hottentot and South African negro crania studied by distances and discrimination. *Am. J. Phys. Anthropol.* 33:169–196.

Rousseeuw, P. J., I. Ruts, and J. W. Tukey. 1999. The bagplot: A bivariate boxplot. *Am. Stat.* 53:382–387.

Ruts, I., and P. J. Rousseeuw. 1996. Computing depth contours of bivariate point clouds. *Comput. Stat. Data Anal.* 23:153–168.

Scrucca, L., M. Fop, T. B. Murphy et al. 2016. mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *R J.* 8:289–317.

Šlaus, M., and Ž. Tomicic. 2005. Discriminant function sexing of fragmentary and complete tibiae from medieval Croatian sites. *Forensic Sci. Int.* 147:147–152.

Small, C. G. 1990. A survey of multidimensional medians. *Int. Stat. Rev.* 58:263–277.

Snow, C. C., S. Hartman, E. Giles et al. 1979. Sex and race determination of crania by calipers and computer: A test of the Giles and Elliot discriminant functions in 52 forensic science cases. *J. Forensic Sci.* 24:448–460.

Stull, K. E., M. W. Kenyhercz, and E. N. L'Abbé. 2014. Ancestry estimation in South Africa using craniometrics and geometric morphometrics. *Forensic Sci. Int.* 245:206.e1–206.e7.

Tukey, J. W. 1975. Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians* 2:523–531.

Ubelaker, D. H., A. H. Ross, and S. M. Graver. 2002. Application of forensic discriminant functions to a Spanish cranial sample. *Forensic Science Communications* 4:1–6.

Urbanová, P., A. H. Ross, M. Jurda et al. 2014. Testing the reliability of software tools in sex and ancestry estimation in a multi-ancestral Brazilian sample. *Leg. Med. (Tokyo)* 16:264–273.

Walker, P. L. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. *Am. J. Phys. Anthropol.* 136:39–50.

Wijsman, E. M., and W. A. Neves. 1986. The use of nonmetric variation in estimating human population admixture: A test case with Brazilian blacks, whites, and mulattos. *Am. J. Phys. Anthropol.* 70:395–405.

Wolf, H. P., and U. Bielefeld. 2014. *Aplpack: Another Plot PACKage: Stem.leaf, bagplot, faces, spin3R, plotsummary, plothulls, and some slider functions. R package version 1.3.0.* https://CRAN.R-project.org/package=aplpack.

Zakrzewski, S. R. 2007. Population continuity or population change: Formation of the ancient Egyptian state. *Am. J. Phys. Anthropol.* 132:501–509.

Zuo, Y., and R. Serfling. 2000. General notions of statistical depth function. *Ann. Stat.* 28:461–482.

Table 1. Measurements in the Current Study

Abbreviation	Measurement
GOL	Glabello-occipital length
NOL	Nasio-occipital length
BNL	Basion-nasion length
BBH	Basion-bregma height
XCB	Maximum cranial breadth
XFB	Maximum frontal breadth
ZYB	Bizygomatic breadth
AUB	Biauricular breadth

Table 2. Discriminant Models Ranging between, and Including, the Linear and the Quadratic

No.	Model	Common Names	**V**	**S**	**O**	# V	# S	# O	Total Covariance Parmeters
1	$\lambda DAD'$	Linear	E	E	E	1	t-1	t(t-1)/2	t(t+1)/2
2	$\lambda_K DAD'$	Proportional	V	E	E	G	t-1	t(t-1)/2	t(t+1)/2 + G - 1
3	$\lambda DA_K D'$	NA	E	V	E	1	G(t-1)	t(t-1)/2	$(t^2+t(2G-1))/2-G+1$
4	$\lambda_K DA_K D'$	CPC	V	V	E	G	G(t-1)	t(t-1)/2	$(t^2+t(2G-1))/2$
5	$\lambda D_K A D_K'$	NA	E	E	V	1	t-1	Gt(t-1)/2	$(Gt^2+t(2-G))/2$
6	$\lambda_K D_K A D_K'$	NA	V	E	V	G	t-1	Gt(t-1)/2	$(Gt^2+t(2-G))/2+G-1$
7	$\lambda D_K A_K D_K'$	NA	E	V	V	1	G(t-1)	Gt(t-1)/2	Gt(t+1)/2-G+1
8	$\lambda_K D_K A_K D_K'$	Quadratic	V	V	V	G	G(t-1)	Gt(t-1)/2	Gt(t+1)/2

The column "Model" uses the notation from Bensmail and Celeux (1996). "**V**," "**S**," and "**O**" refer to volume, shape, and orientation. "E" and "V" within those columns refer to equal or variable. The columns "# V," "# S," and "# O" list the number of parameters given *t* traits and *G* groups and the last column gives the total number of covariance matrix parameters.

Table 3. Multivariate Parameters for Analysis of the Buriat and Easter Islanders by EDDA

Means

	GOL	NOL	BNL	BBH	XCB	XFB	ZYB	AUB
Buriat	1.27269	1.25932	0.71496	0.93529	1.09189	0.89405	1.00321	0.95579
Easter Isl.	1.37107	1.34149	0.79326	1.03613	0.96844	0.80416	0.96527	0.88192

Buriat covariance parameters

$\lambda = 1.985711 \times 10^{-4}$

Diagonal of **A** (Buriat and Easter Islander) =

$$\begin{bmatrix} 12.857152 & 7.253005 & 5.264252 & 2.325695 & 1.905311 & 1.039380 & 0.206931 & 0.002137 \end{bmatrix}$$

D =

$$\begin{bmatrix} 0.512589 & -0.328968 & 0.119033 & -0.253622 & 0.008962 & 0.009092 & 0.697538 & 0.252622 \\ 0.519989 & -0.331487 & 0.051839 & -0.162720 & -0.059483 & -0.051821 & -0.708742 & 0.286397 \\ 0.214975 & 0.216393 & 0.029701 & 0.810998 & -0.021315 & 0.121611 & 0.059027 & 0.479200 \\ 0.129719 & 0.670416 & -0.420684 & -0.428215 & -0.191824 & 0.068255 & 0.007004 & 0.363136 \\ -0.426095 & -0.391868 & -0.095417 & -0.044236 & -0.608702 & 0.432423 & 0.018496 & 0.309793 \\ -0.294750 & -0.328171 & -0.573918 & 0.015549 & 0.498697 & -0.284998 & 0.024691 & 0.381144 \\ -0.234374 & 0.122073 & 0.494855 & -0.241060 & 0.531469 & 0.471389 & -0.078774 & 0.341031 \\ -0.283650 & 0.097235 & 0.471082 & -0.088336 & -0.239698 & -0.698114 & 0.020848 & 0.367578 \end{bmatrix}$$

Easter Island covariance parameters

$\lambda = 1.484856 \times 10^{-4}$

D =

$$\begin{bmatrix}
-0.582997 & -0.020503 & 0.353972 & -0.043896 & 0.174563 & -0.002355 & 0.661410 & -0.254026 \\
-0.531404 & -0.077456 & 0.283155 & -0.008830 & 0.068099 & -0.149454 & -0.736747 & -0.248165 \\
-0.200891 & 0.156354 & -0.300402 & 0.068107 & -0.714024 & 0.376426 & 0.016088 & -0.434201 \\
-0.128398 & 0.205731 & -0.649463 & 0.332282 & 0.544229 & -0.076827 & -0.006617 & -0.326870 \\
0.305133 & -0.596073 & 0.242838 & 0.530289 & 0.094055 & 0.284558 & -0.002022 & -0.348698 \\
0.157556 & -0.470403 & -0.251839 & -0.677655 & 0.041990 & -0.213552 & 0.038215 & -0.427118 \\
0.324827 & 0.503041 & 0.322524 & -0.304138 & 0.309569 & 0.462425 & -0.106168 & -0.352100 \\
0.312152 & 0.311730 & 0.236787 & 0.223658 & -0.228980 & -0.699746 & 0.081866 & -0.387965
\end{bmatrix}$$

Appendix 1.

Given the potential confusion between equations (3) through (5), we use simulation to demonstrate that equation (3) is the correct F test when a case is a new one not contained in the reference sample. The simulation uses the following distributions:

$$\begin{aligned} \mathbf{x} &\sim MVN\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \\ \mathbf{V} &\sim W\left(N-1, \boldsymbol{\Sigma}\right)/\left(N-1\right), \end{aligned} \tag{9}$$

where the **x** vector is one realization of the simulation, *MVN* is a multivariate normal distribution with a vector of means and covariance matrix equal to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, and *W* is a Wishart distribution with degrees of freedom equal to $N-1$ and a covariance matrix equal to $\boldsymbol{\Sigma}$. The simulated Mahalanobis squared-distance is then $D^2 = \left(\mathbf{x}-\boldsymbol{\mu}\right)'\mathbf{V}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)$. One million simulated D^2 values were then plotted as a Kaplan-Meier (Kaplan and Meier 1958) survivorship function. For comparison, the "inverse" of equation (3) was found by plotting $1-\left(1,...,9999\right)/10000$ against $G\left[\left(\left(1,...,9999\right)/10000\right), t, N-t\right]\times\left(t\left(N^2-1\right)/\left(N\left(N-t\right)\right)\right)$, where $G\left[\alpha, t, N-t\right]$ is the value in the F distribution with t and $N-t$ degrees of freedom that gives a lower tail area of α. For comparison, we also plotted the χ^2 distribution with *t* degrees of freedom as the distribution of Mahalanobis squared-distances is asymptotically distributed as a χ^2.

We use the above described simulation in the case of both large and small sample sizes. We use the Buriat data, with a sample size of 109, as the large sample. Only the Norse, Tolai, and Peruvians (N = 110 each) and Egyptians (N = 111) have larger sample sizes within the Howells dataset. We use the Maori data as an example of small sample size (N = 20). Figures A.1 and A.2 show the results of 1,000,000 simulations represented as Kaplan-Meier survivorship curves for the Buriat and Maori samples, respectively. Within each of these figures, the Kaplan-

Meier survivorship curve is compared to the F distributions from equation (3) and the χ^2 distribution with eight degrees of freedom. Both Figures show that equation (3) provides the appropriate F-test for typicality with a new case. A comparison between Figures A.1 and A.2 also demonstrates the asymptotic behavior of the χ^2 distribution. The χ^2 distribution provides a better fit with the larger sample (N = 109) in Figure A.1 than with the smaller sample (N = 20) in Figure A.2.

Appendix 2. Labeling for Figure 5

Population	Symbol	Population	Symbol
Ainu	A	Maori	P
Andaman	B	Mokapu	Q
Anyang	C	Moriori	R
Arikara	D	N. Japan	S
Atayal	E	Norse	T
Australia	F	Peru	U
Berg	G	Phillipines	V
Buriat	H	S. Japan	W
Bushman	I	Santa Cruz Isl.	X
Dogon	J	Tasmania	Y
Easter Isl.	K	Teita	Z
Egypt	L	Tolai	1
Eskimo	M	Zalavar	2
Guam	N	Zulu	3
Hainan	O		

Figure 1.

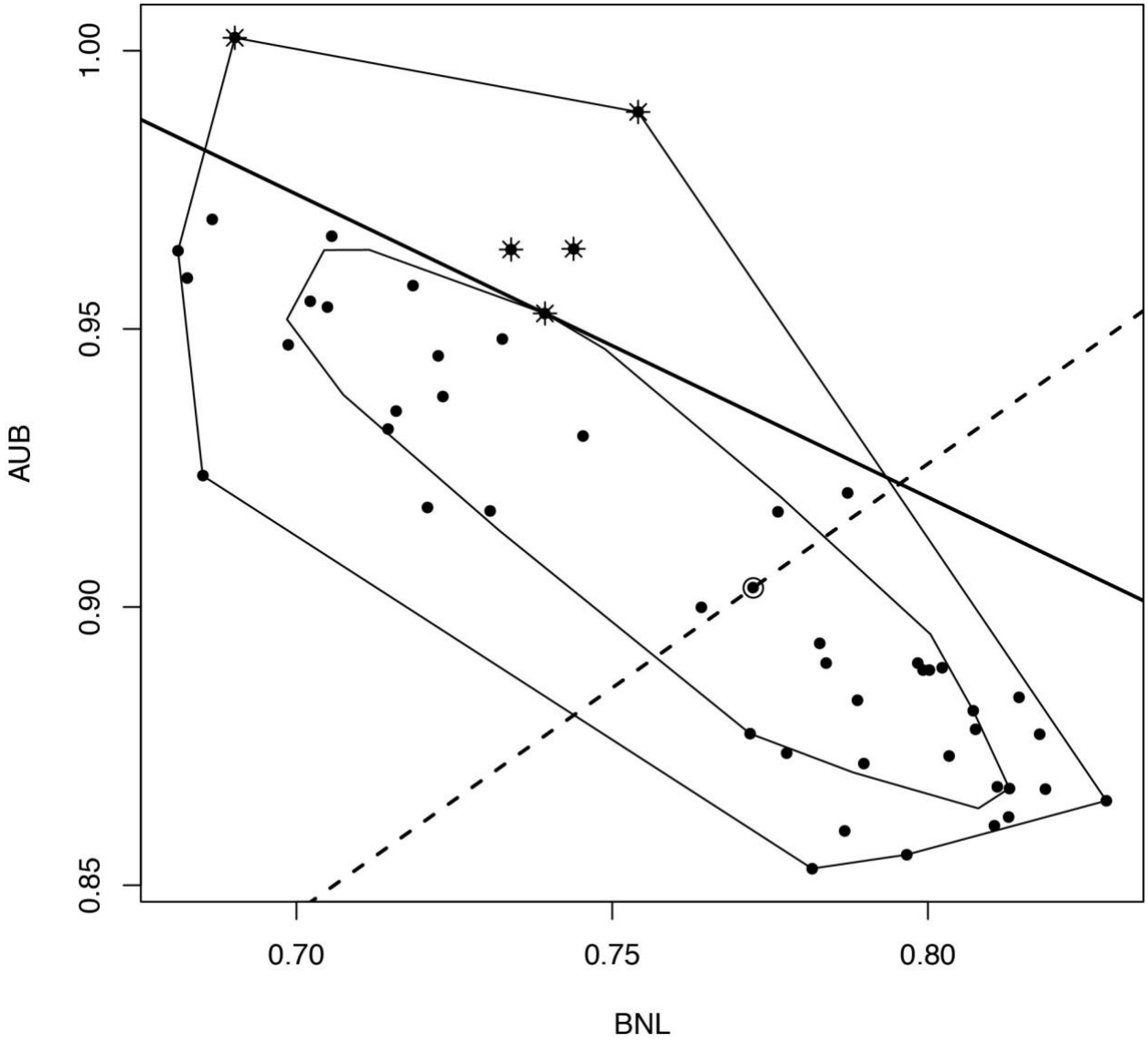

Figure 2.

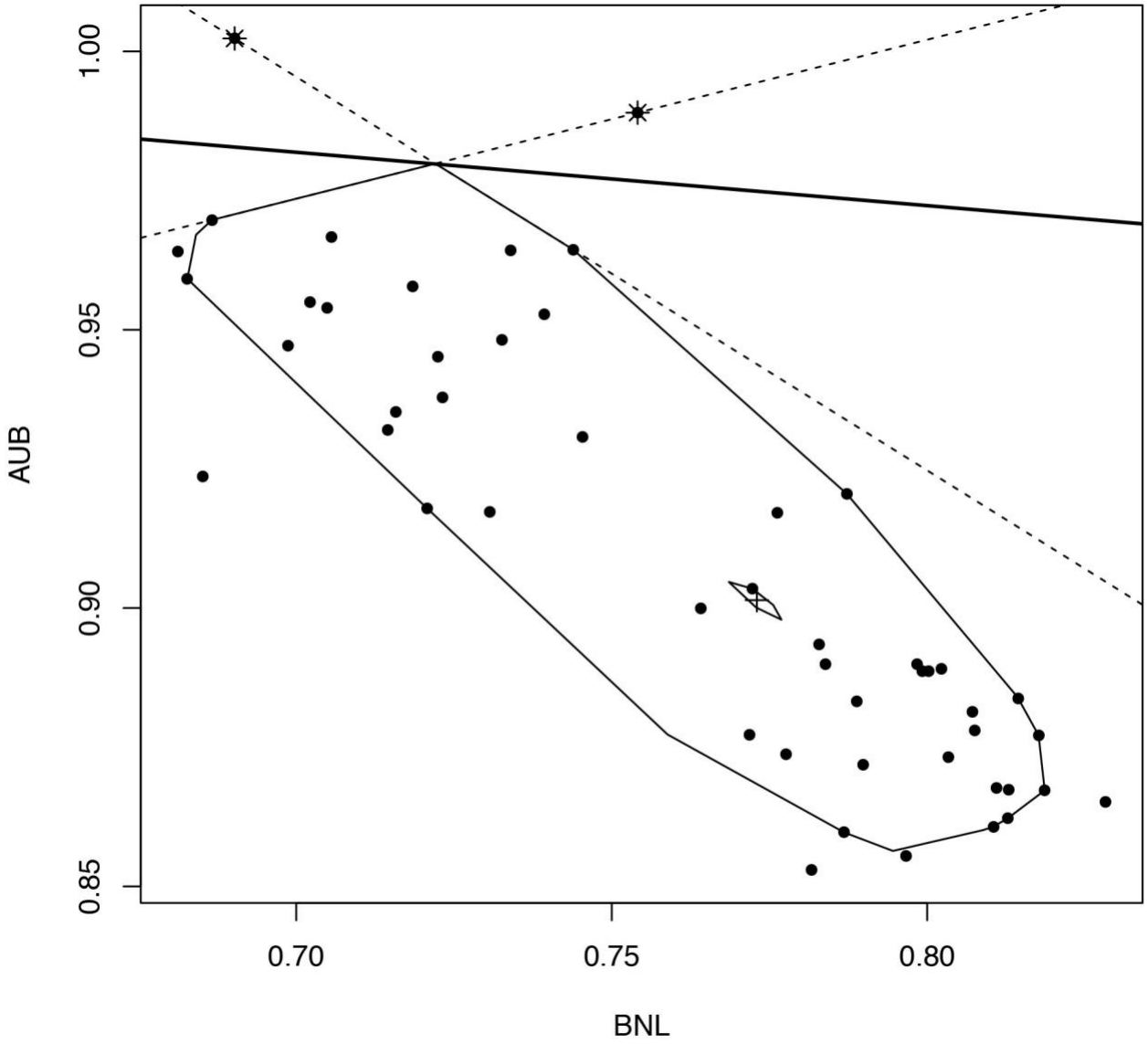

Figure 3.

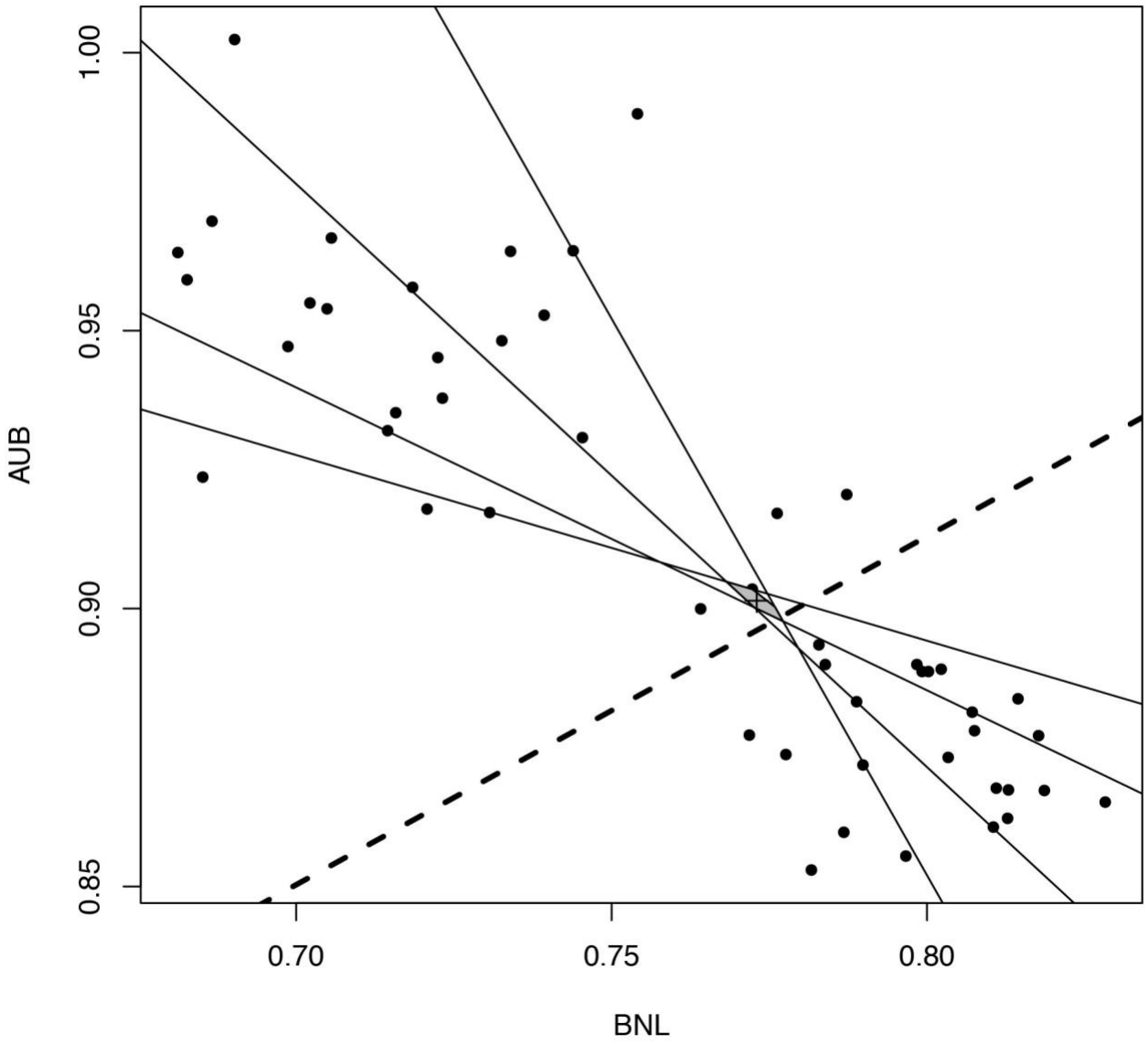

Figure 4.

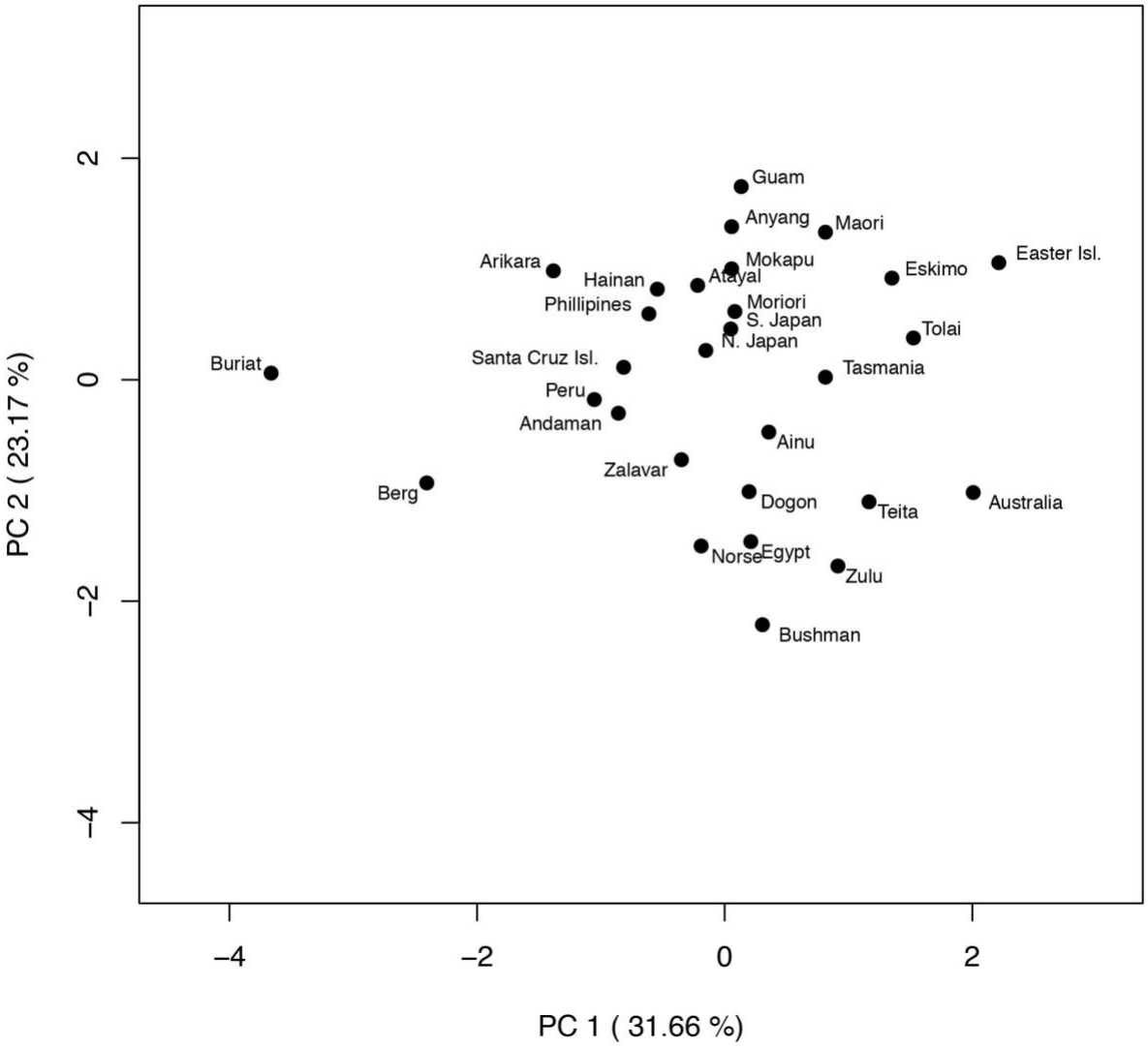

Figure 5.

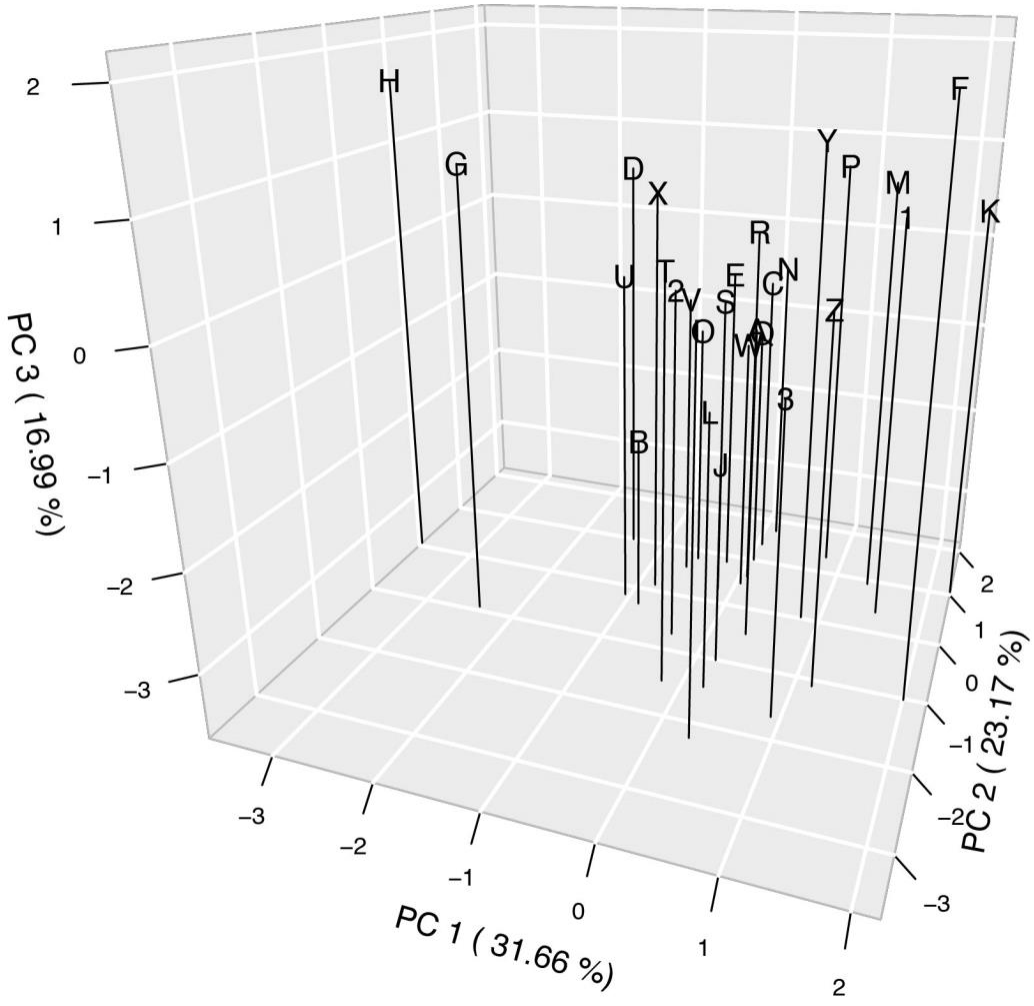
PC 3 (16.99 %)
PC 1 (31.66 %)
PC 2 (23.17 %)
2
1
0
−1
−2
−3
H
G
D
X
Y
P
M
1
F
K
R
U
T
2
V
S
E
C
N
Z
O
W
A
3
B
F
J

Figure 6.

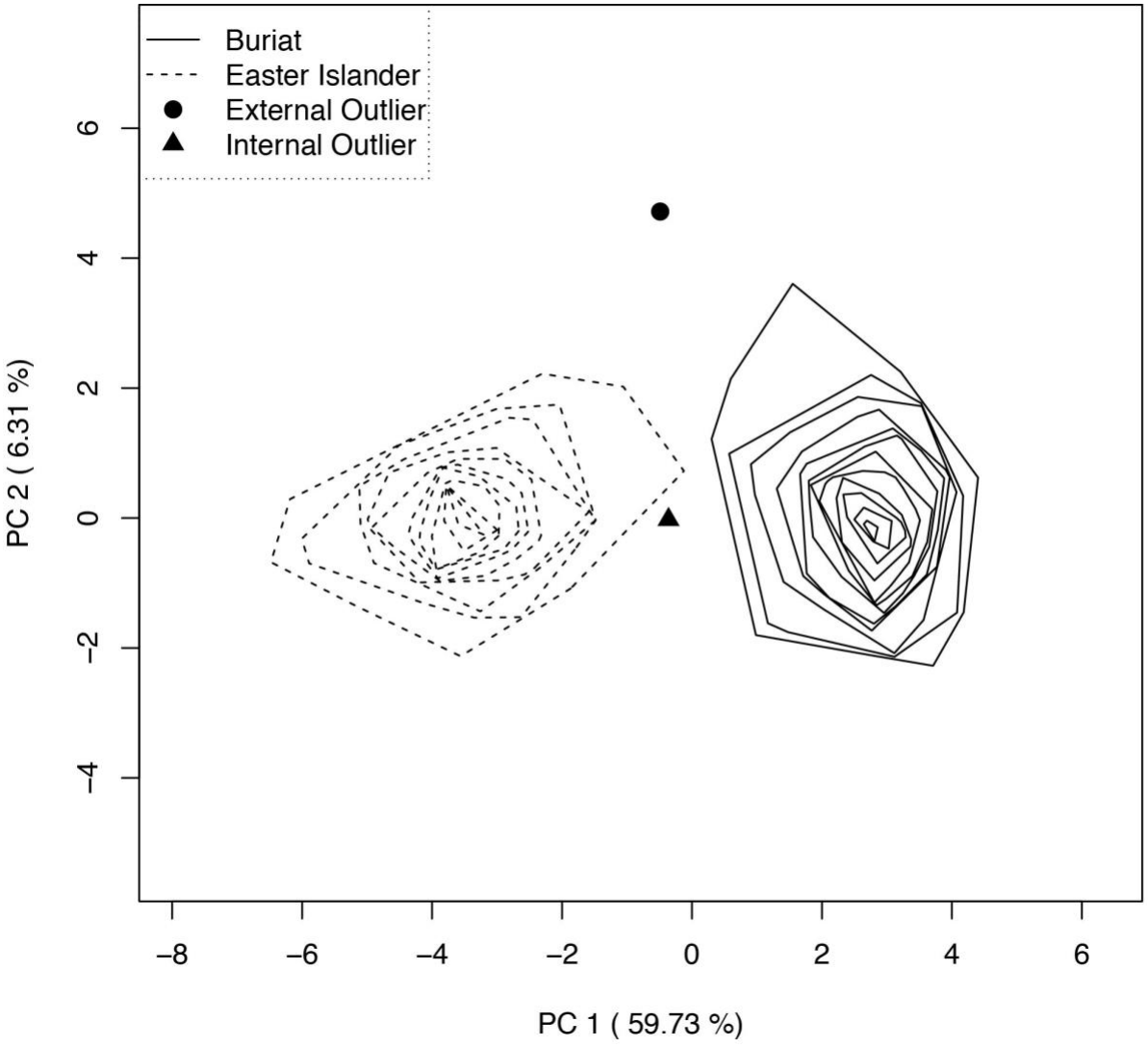

Figure 7.

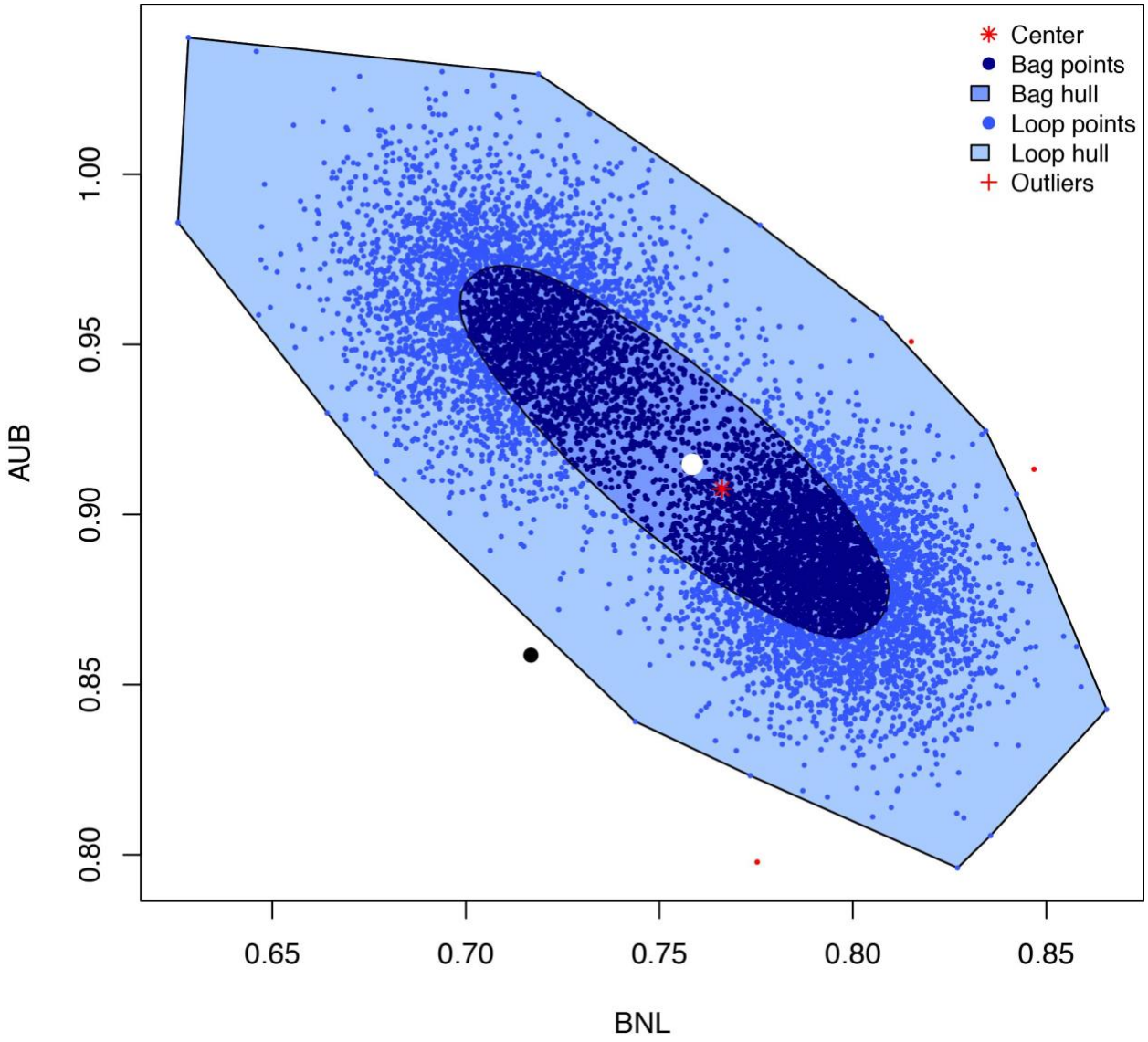

Figure 8.

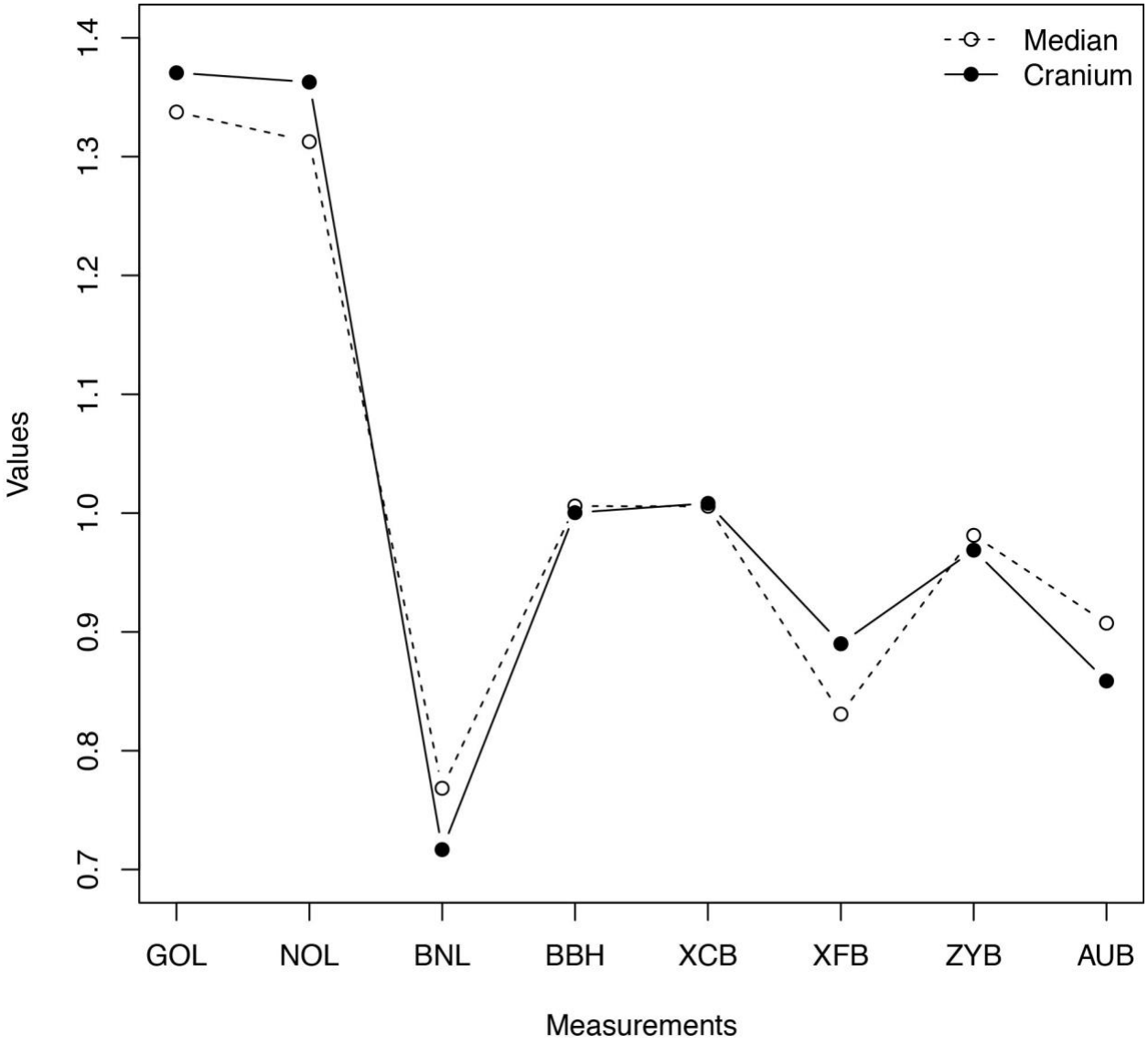

Figure 9.

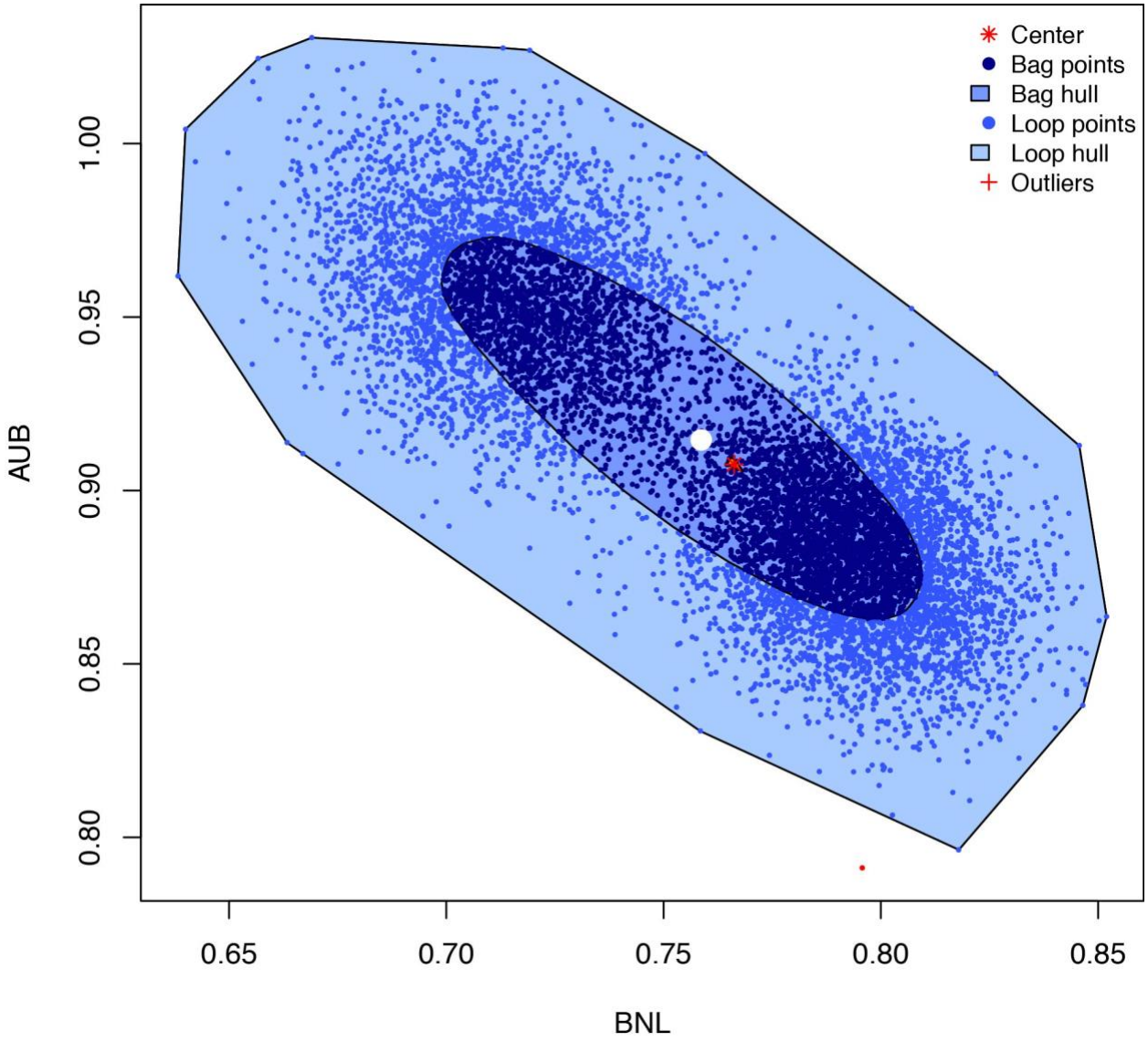
Center
Bag points
Bag hull
Loop points
Loop hull
Outliers
AUB
1.00
0.95
0.90
0.85
0.80
0.65
0.70
0.75
0.80
0.85
BNL

Figure 10.

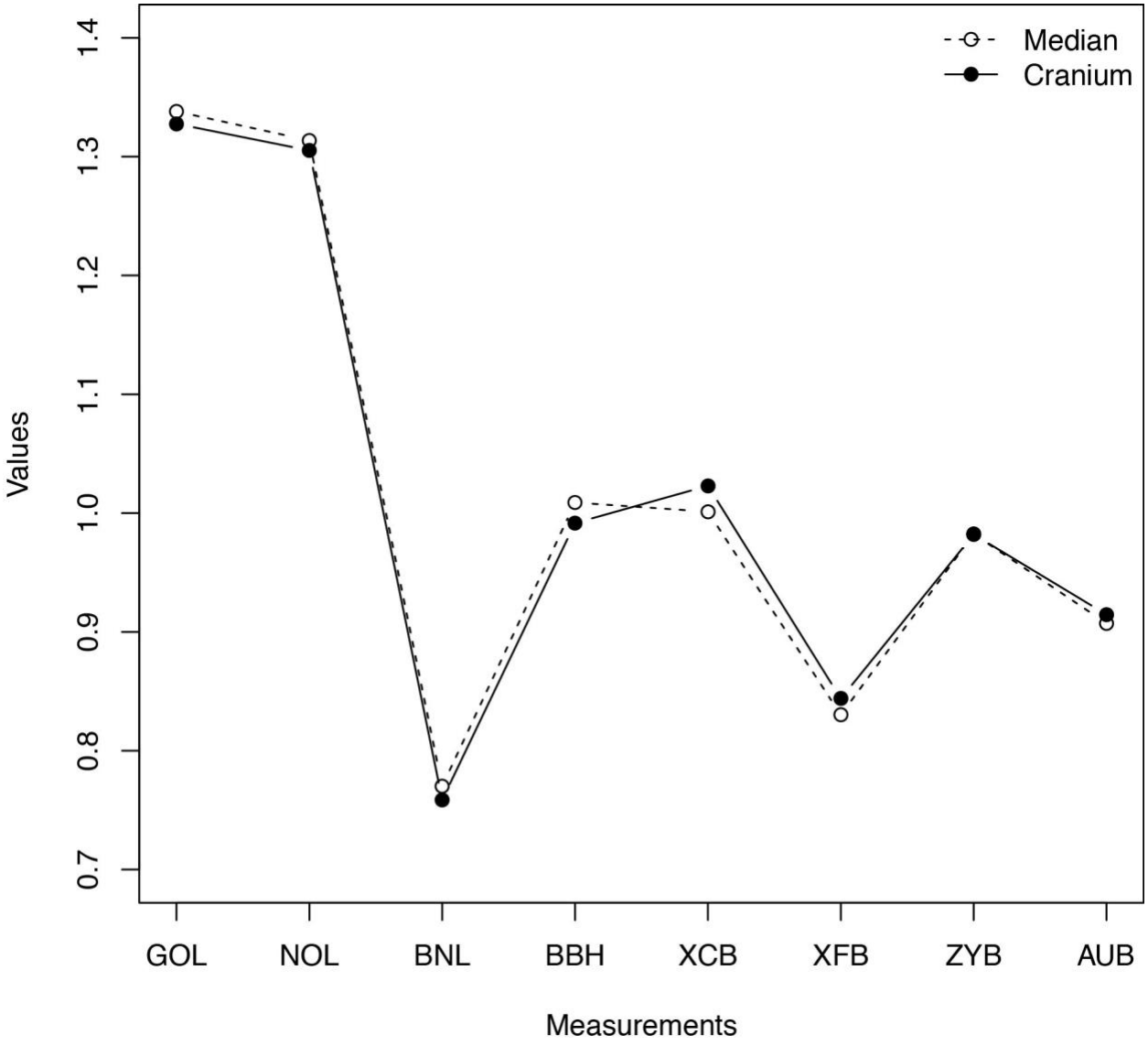

Figure A.1.

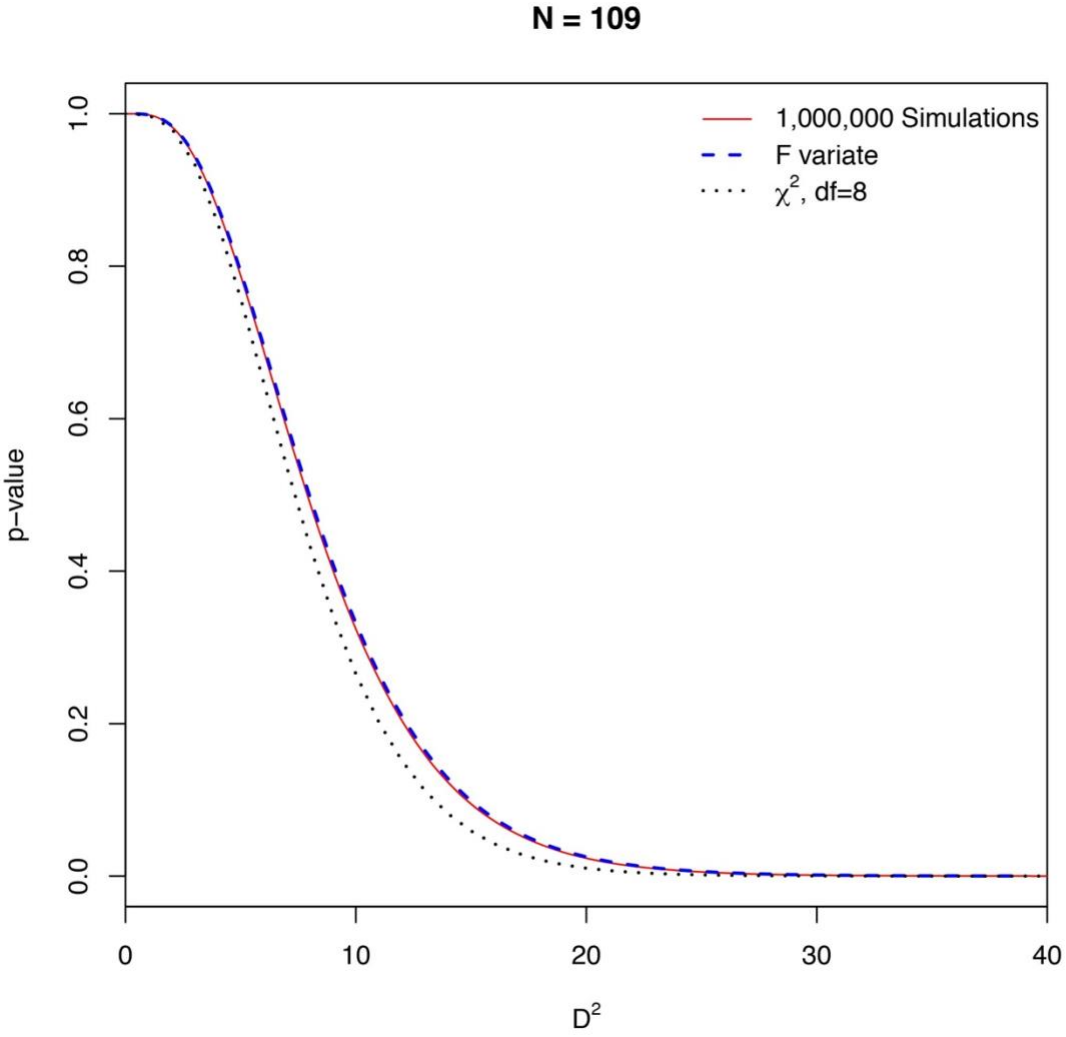
N = 109
1,000,000 Simulations
F variate
χ², df=8
p–value
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
D²

Figure A.2.

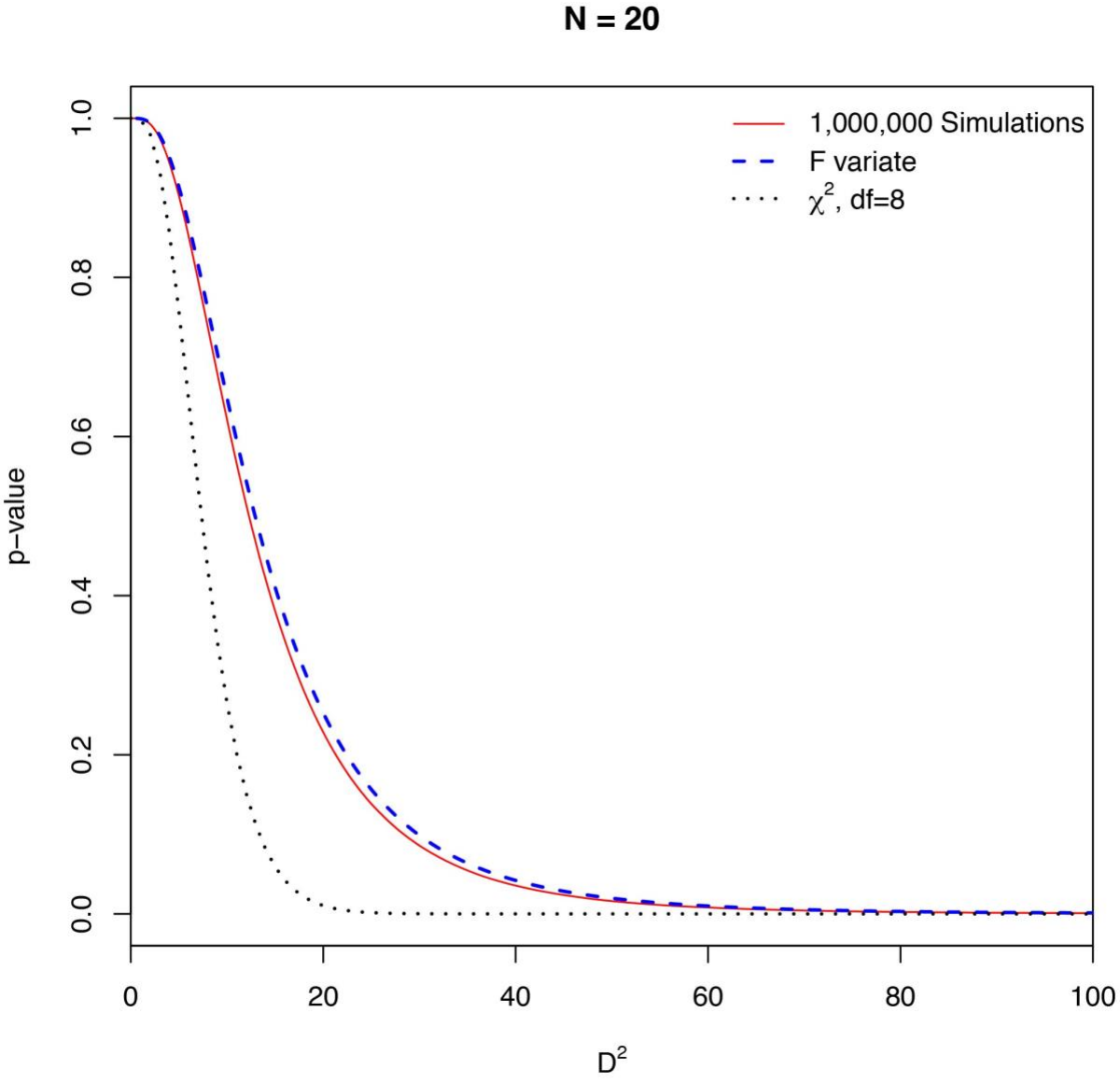
N = 20
1,000,000 Simulations
F variate
χ², df=8
p−value
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
D²

Figure Captions

Figure 1. Example of a bivariate "depth contour" (Ruts and Rousseeuw 1996) plot with 50 points showing the 1-depth and 5-depth contours. The five points marked with asterisks lying on or above the heavy line define the depth at that particular vertex of the 5-depth contour. The circled point is the deepest point with a depth of 22. This depth is found by counting up the 21 points to the lower right of the dashed line plus the circle point lying on this line.

Figure 2. Continuation of the example from Figure 1. The dashed intersecting lines show one of the vertices of the 2-depth contour. A heavy line is struck through this vertex and the two asterisked points lay above the line. This shows that the line drawn through the constructed point on the 2-depth contour does indeed separate two points from the remaining 48 points below the heavy line. Also shown is the 22-depth contour with its center of gravity marked by a "plus" symbol.

Figure 3. Continuation of Figure 2. The light lines show the boundaries of the deepest (22-depth) contour. The heavy dashed line shows the depth of 22 with 22 points below and to the right of the dashed line. The gray polygon is the 22-depth contour with its center of gavity marked by a "plus" symbol. This is the bivariate generalization of the univariate median.

Figure 4. Two-dimensional principal coordinates plot of the Howells 29 local population samples. Note how the Buriat and Easter Islanders have the greatest overall distance in two-dimensions.

Figure 5. Three-dimensional principal coordinates plot of the Howells 29 local population samples. "H" marks the Buriat and "K" the Easter Islanders. The key for all of the labels is provided in Appendix 2.

Figure 6. Two-dimensional principal coordinates plot of the Easter Islander and Buriat data (shown as convex hull "peels"), the average of the two centroids (labelled "internal outlier"), and individual number 703 (a Zulu) from the Howells craniometric data (labelled "external outlier").

Figure 7. Bagplot of 10,000 simulated values from the predictive distribution given admixture proportions of 0.444from the Buriat and 0.556 from Easter Islanders. These are the admixture proportions that give the same posterior probabilities as those for a Zulu cranium (Howells individual number 703). The legend in the upper right corner gives the symbols and colored regions from "deepest" to "shallowest." The actual cranium (shown as a filled black point) is slightly below the BNL median, well below the AUB median, and outside of the bagplot's "loop." The white point is the internal outlier "synthetic skull" with the given mixing proportions.

Figure 8. Plot of the estimated depth median with mixing proportions of 0.444 from the Buriat and 0.556 from Easter Islanders and the measurements from the Zulu cranium.

Figure 9. Bagplot of 10,000 simulated values from the predictive distribution given admixture proportions of 0.442 from the Buriat and 0.558 from Easter Islanders. These are the admixture proportions that result in a cranium with equal posterior probabilities for group-membership with

the Buriat and Easter Islanders. The legend in the upper right corner gives the symbols and colored regions from "deepest" to "shallowest." The "synthetic skull" (shown as a filled white point) is well within the bag and is near the center or depth median.

Figure 10. Plot of the estimated depth median with mixing proportions of 0.442 from the Buriat and 0.558 from Easter Islanders. These are the admixture proportions that result in a cranium with equal posterior probabilities for group-membership with the Buriat and Easter Islanders. The other line is the "synthetic skull" formed with the given admixture proportions.

Figure A.1. Comparison of 1,000,000 simulations from the Buriat parameters (represented as a Kaplan-Meier survivorship curve) to the F distribution from equation (3) and a χ^2 distribution with eight degrees of freedom.

Figure A.2. Comparison of 1,000,000 simulations from the Maori parameters (represented as a Kaplan-Meier survivorship curve) to the F distribution from equation (3) and a χ^2 distribution with eight degrees of freedom.