Statistical analysis of shape through triangulation of landmarks: A study of sexual dimorphism in hominids

(compositional data/Hotelling's T² test/Mahalanobis distance/shape analysis)

Calyampudi R. Rao*† and Shailaja Suryawanshi‡

*Department of Statistics, 326 Thomas Building, Pennsylvania State University, University Park, PA 16802; and ‡Merck Research Laboratories, Rahway, NJ 07065

Contributed by Calyampudi R. Rao, December 18, 1997

ABSTRACT Two objects with homologous landmarks are said to be of the same shape if the configuration of landmarks of one object can be exactly matched with that of the other by translation, rotation/reflection, and scaling. In an earlier paper, the authors proposed statistical analysis of shape by considering logarithmic differences of all possible Euclidean distances between landmarks. Tests of significance for differences in the shape of objects and methods of discrimination between populations were developed with such data. In the present paper, the corresponding statistical methodology is developed by triangulation of the landmarks and by considering the angles as natural measurements of shape. This method is applied to the study of sexual dimorphism in hominids.

1. Introduction

In a previous paper (1), the authors presented a statistical analysis of the shape of objects considering the ratios of Euclidean distances between landmarks as basic data. As observed by Lele (2), such ratios, which are invariant to translation, rotation, and scaling of the configuration of landmarks, provide measurements on shape. If there are k landmarks, we have a set of k(k-1)/2 Euclidean distances, all of which may not be necessary to specify the configuration of landmarks on an object. We suggested the choice of a minimal set of distances that uniquely specify the configuration of landmarks for purposes of statistical inference. For a twodimensional object with k landmarks, this number lies between (2k - 3) and 3(k - 2). In general, when the relative positions of landmarks are known, (2k - 3) distances will do, as in the case of the human profile illustrated in our earlier paper and reproduced below (Fig. 1). There are 8 landmarks indicated by a, b, c, d, e, f, g, and h, with 28 distances between landmarks. However, 13 distances, as indicated in Fig. 1, specify the entire configuration.

It may be seen from the above diagram that the configuration of landmarks can also be specified by 6 triangles. Each triangle provides two angles that are invariant to translation, rotation, and scaling. There are altogether 12 angles arising out of 6 triangles, which appear to be natural shape measurements. We explore the possibility of studying differences in shape through angular coordinates resulting from a suitable triangulation of the landmarks. Such an approach was also indicated by Bookstein (3), although he did not develop the appropriate statistical methodology.

We shall first discuss the simple case of three landmarks and suggest a general approach in the case of many landmarks.



FIG. 1. Minimum number of distances required to fix the landmarks on human facial profile. (Modified from ref. 1.)

We also discuss the possibility of augmenting the angular data provided by a triangulation of the landmarks with sets of angles characterizing the shape of the edges (or profiles), if available, between landmarks. Such data may provide additional information in problems of discrimination and identification of objects by shape.

2. Objects Specified by Three Landmarks

First, we consider objects specified by three landmarks, say 1, 2, and 3, and denote the angles at the corresponding vertices by θ_1 , θ_2 , and θ_3 , which add up to 180 degrees (or π radians). These angles, which are natural measurements of shape, are referred to in statistical literature as compositional data. For purposes of statistical inference, one may choose a suitable stochastic model and apply the appropriate methodology for estimation and tests of significance. For a discussion of some models for compositional data, the reader is referred to a book by Aitchison (4) and a paper by Pukkila and Rao (5). It may be noted that because θ_1 , θ_2 , and θ_3 are non-negative, they can be transformed to directional data by considering $l_i =$ $\sqrt{\theta_i}/180$, i = 1, 2, 3, in which case an appropriate stochastic model for directional data may be used. See Mardia (6) and Pukkila and Rao (5) for a discussion of models for directional data and statistical inference based on them. Alternatively, one can use nonparametric methods.

For purpose of illustration, we use the angular data (in degrees) given in Aitchison (ref. 4, pp. 385–386) relating to three landmarks, nasion (N), alveolar (A), and basion (B), on seventeenth century English and Naqada skulls. The mean values of the angles designated as N, A, and B and the sample sizes (in parentheses) for male and female skulls are given in Table 1. The pooled sums of squares and products are given in Table 2.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

[@] 1998 by The National Academy of Sciences 0027-8424/98/954121-5\$2.00/0 PNAS is available online at http://www.pnas.org.

[†]To whom reprint requests should be addressed. e-mail: crr1@ psu.edu.

 Table 1.
 Mean values of the three angles for English and Naqada skulls

		Englis	h		Naqada			
θ	Male (29)	Female (22)	Combined (51)	Male (29)	Female (22)	Combined (51)		
N	65.241	64.750	65.029	69.579	69.389	69.497		
А	73.707	73.705	73.706	74.452	75.361	74.844		
В	41.052	41.591	41.284	35.976	35.250	35.663		

Because the sum of the angles is a constant, we need consider only two angles. We choose N and A. A new test for multivariate normality (bivariate in the present problem) developed by Rao and Ali (7) had *p*-values of the order of 0.75 for English and 0.29 for Naqada data (with possibly one or two outliers in the latter case) showing no significant departure from normality. In case non-normality is indicated, we may try transformations such as

$$x = \log \theta_1 - \log \theta_3$$
$$y = \log \theta_2 - \log \theta_3$$

or

$$x = \log \sin \theta_1 - \log \sin \theta_3$$
$$y = \log \sin \theta_2 - \log \sin \theta_3,$$

whose distributions may be close to bivariate normality than θ_1 and θ_2 . Table 3 gives the Hotelling T^2 values for testing differences between male and female skulls in English and Naqada skulls and also differences between groups ignoring sex. The formula for T^2 is

$$\frac{n-p+1}{p}\frac{n_1n_2}{n_1+n_2}\,\mathbf{d}'S^{-1}\mathbf{d},$$

where n_1 and n_2 are sample sizes for two groups under comparison, n is the degrees of freedom (98 in our case) for S as in Table 2, p is the number of variables (2 in our case), **d** is the vector of differences in mean values, and S^{-1} is the inverse of S. Under the normality assumption, T^2 has an Fdistribution with p and (n - p + 1) degrees of freedom.

It is seen that there are no differences in the shapes of male and female skulls within a group. However, the shapes of English and Naqada skulls are different. The mean shapes of triangles formed by N, A, B for the English and Naqada skulls are represented in Fig. 2.

The angles for an individual can be represented with aerial coordinates within an equilateral triangle. The shapes of triangles represented by points in different positions within the equilateral triangle are shown in Fig. 3.

3. More than Three Landmarks

3.1. Tests for Differences in Shape. When there are more than three landmarks, there is no unique way of triangulation that characterizes the configuration of the landmarks. Some possible triangulations with five landmarks, prosthion (1),

Table 2. Pooled sums of squares and products matrices (degrees of freedom = 98)

Ν	А	В
1490.213	-1321.989	-168.949
-1321.989	1636.930	-316.016
-168.949	-316.016	487.759

Table 3. Tests of hypotheses based on the angles N and A

	Hotelling's	d.f.	
Skulls	T^2	for F	<i>p</i> -values
English (male – female)	0.349	2, 97	0.706
Naqada (male – female)	0.731	2, 97	0.483
English - Naqada (ignoring sex)	85.908	2, 97	0.000

nasion (2), lambda (3), basion (4), and staphylion (5), on hominid skulls chosen for our study are indicated in Fig. 4.

All of the triangles in Fig. 4 a and b have a common vertex, 1 and 5, respectively, and in Fig. 4c, they have a common side: 2–3. In general, there may be some advantage in using Delaunay triangulation, which provides triangles close to the equilateral. In our case, the triangulation indicated in Fig. 4b corresponds to Delaunay triangulation. Because each triangle can be specified by two angles, there are altogether six independent angles describing the shape of an object. It may be noted that the triangulation in Fig. 4c is similar to Bookstein's scheme of choosing a line joining any two landmarks and another line perpendicular to it as coordinate axes to specify the coordinates of the rest of the landmarks.

What triangulation should one choose for statistical analysis? There are two stages in statistical analysis in comparing populations for differences in shape. One is to establish by an appropriate test whether there are any shape differences. The other is to specify the nature of differences in shape. For the first object, any particular triangulation will do, provided that we can find an appropriate stochastic model for the corresponding angles. In practice one may choose two or more different triangulations to check the consistency of results. Once differences in shape are established, it may be necessary to consider all possible triangles formed by choosing all possible sets of three landmarks to specify the nature of differences in shape. First, we examine the differences in the shapes of hominid crania by types of apes (Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus) and sex (male and female), by using the data collected by Paul O'Higgins and studied by O'Higgins and Dryden (8). We compare the results of two triangulations (Fig. 4 a and b) for consistency. At the next step, we examine the nature of the shape difference between the males of Pan and Pongo, the two apes found to be most dissimilar among the three types of apes.

The mean values for the angles as indicated in Fig. 4 *a* and *b* for two triangulations are given in Tables 4 and 5.

The square of the Mahalanobis distance between two populations with mean values μ_1 , μ_2 , and common covariance matrix Σ is defined by

$$(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2),$$

which is estimated by



FIG. 2. Mean shapes of English (thick line) and Naqada (dotted line) triangles.



FIG. 3. Shapes of triangles corresponding to different positions of points.

where $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ are the sample mean vectors and *S* is the pooled sum of squares and products matrix based on *n* degrees of freedom [see Rao (9)]. Hotelling's T^2 , which provides a test for the hypothesis $\mu_1 = \mu_2$, is

$$T^{2} = \frac{n-p+1}{p} \frac{n_{1}n_{2}}{n_{1}+n_{2}} \frac{1}{n} D^{2}$$
$$= \frac{n-p+1}{p} \frac{n_{1}n_{2}}{n_{1}+n_{2}} (\bar{\mathbf{x}}_{1}-\bar{\mathbf{x}}_{2}) S^{-1} (\bar{\mathbf{x}}_{1}-\bar{\mathbf{x}}_{2}),$$

where n_1 and n_2 are sample sizes on which $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ are based. The above statistic has an *F* distribution on *p* and n - p + 1 degrees of freedom. In our problem, n = 167 - 6 = 161. The D^2 and T^2 values for testing differences between groups by sex and differences between sexes within each group are reported in Tables 6 and 7.



FIG. 4. Different possible triangulations of landmarks.

Table 4. Mean values of angles-triangulation I

	Pan		Go	Gorilla		Pongo	
	Male (28)	Female (26)	Male (29)	Female (30)	Male (30)	Female (24)	
$\overline{\theta_1}$	0.0301	0.0365	0.0947	0.0406	0.1446	1.3070	
θ_2	0.2727	0.3018	0.2917	0.2910	0.3472	0.3644	
θ_3	0.5202	0.4929	0.4229	0.4850	0.2585	0.3332	
ψ_1	3.0752	3.0607	2.9249	3.0457	2.8417	2.8627	
ψ_2	2.3262	2.2674	2.1789	2.2715	2.0500	2.0980	
ψ3	0.3429	0.3164	0.3094	0.3416	0.2079	0.2202	

All T^2 values have low *p*-values except for the difference between *Pan* males and females, showing significant differences in shape. Some interesting observations arising out of the study of T^2 and D^2 values are as follows. There are no inconsistencies in the conclusions based on the two triangulations of the landmarks. The D^2 values for comparing the males of different species are somewhat larger than the corresponding D^2 values for females, indicating that shapes of female crania of different apes are more similar than the shapes of the male crania of different apes. Among the hominids, *Pan* and *Gorilla* are closer in the shape of the crania, and *Pongo* is more distant.

3.2. Test for Sexual Dimorphism. The difference in shape between male and female crania seems to be of different orders of magnitude, judged by the D^2 values, in the three species, indicating sexual dimorphism, which can be tested as follows. Let \mathbf{d}_c , \mathbf{d}_g , and \mathbf{d}_0 be the vectors of differences in mean values of six angles between males and females in the *Pan* (chimpanzee), *Gorilla*, and *Pongo* (orangutan) samples. To \mathbf{d}_c we attach a weight $w_c = (28 \times 26)/(28 + 26)$, where the numbers 28 and 26 are the sample sizes for *Pan* males and females. Similarly, we compute the weights w_g and w_0 , for *Gorilla* and *Pongo* (orangutan) samples. Then, we compute what is called the sum of squares and products matrix between the species using the formula

$$B = w_c \mathbf{d}_c \mathbf{d}'_c + w_a \mathbf{d}_a \mathbf{d}'_a + w_0 \mathbf{d}_0 \mathbf{d}'_0 - w \mathbf{d} \mathbf{d}',$$

where $w = w_c + w_g + w_0$ and

$$\mathbf{d} = (w_c \mathbf{d}_c + w_q \mathbf{d}_q + w_0 \mathbf{d}_0) / w.$$

To this (6×6) matrix we attach q = 2 degrees of freedom. The pooled sum of squares and products matrix used in the computations of D^2 and T^2 values is the 6×6 matrix S, based on n = 161 degrees of freedom. We compute the Wilks Λ statistic to test for sexual dimorphism

$$\Lambda = \frac{|S|}{|S+B|} = 0.608919.$$

The significance of Λ is assessed by using Rao's transformation of Λ into *F* by the following computations:

Table 5. Mean values of angles-triangulation II

	Pan		Go	rilla	Pongo		
	Male (28)	Female (26)	Male (29)	Female (30)	Male (30)	Female (24)	
θ_1	1.3287	1.3072	1.2336	1.2815	1.3684	1.2324	
θ_2	1.3762	1.3763	1.2597	1.3273	0.9987	1.1428	
θ_3	0.4381	0.4921	0.4326	0.4601	0.4746	0.4914	
ψ_1	0.7924	0.7791	0.8088	0.8068	0.7503	0.8266	
ψ_2	1.2580	1.2771	1.3101	1.2617	1.6524	1.5056	
ψ3	0.3783	0.4007	0.4086	0.3681	0.4617	0.4061	

Table 6. D^2 and T^2 values for differences between species by sex

		Males			Females		
Species	D^2	T^2	<i>p</i> -value	D^2	T^2	p-value	
Pan ~ Gorilla							
Triangulation I	10.37	23.85	< 0.001	2.17	4.88	< 0.001	
Triangulation II	12.50	28.76	< 0.001	5.84	13.13	< 0.001	
Gorilla ~ Pongo							
Triangulation I	28.57	68.03	< 0.001	18.06	38.89	< 0.001	
Triangulation II	30.51	72.65	< 0.001	19.84	42.72	< 0.001	
$Pongo \sim Pan$							
Triangulation I	43.11	100.84	< 0.001	17.77	35.81	< 0.001	
Triangulation II	45.31	105.96	< 0.001	20.56	41.45	< 0.001	

$$m = n + \frac{q - p - 1}{2} = 161 + \frac{2 - 6 - 1}{2} = 158.5$$

$$s = \sqrt{\frac{p^2 q^2 - 4}{p^2 + q^2 - 5}} = 2$$

$$\lambda = \frac{pq - 2}{4} = 2.5$$

$$F = \frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \frac{ms - 2\lambda}{pq} = 7.3191.$$

F is approximately distributed as *F* on pq = 12 and ms = 312 degrees of freedom. The *p*-values for F = 7.3991 based on 12 and 312 degrees of freedom is small, indicating sexual dimorphism.

3.3 Canonical Coordinates for Graphical Representation. Rao (10) developed the concept of canonical coordinates for representing the relative positions of the populations under study, which are characterized by a number of measurements (6 angles in the present problem), in a low-dimensional space. For this we consider the 6×6 matrix X of mean values with rows representing the variables (6 angles) and columns the populations (6 groups of hominids) and compute the "between sums of squares and products" matrix

$$B = X \left(I_6 - \frac{1}{6} J_6 \right) X',$$

where I_6 is a diagonal matrix with unities on the diagonal and J_6 is a 6×6 matrix with unity as all entries. Let $W = n^{-1}S$, where *n* is the degrees of freedom and *S* is the pooled sums of squares and products. Then we compute the eigenvalues and eigenvectors using the determinantal equation

$$|W^{-1/2}BW^{-1/2} - \lambda I| - 0.$$

where $W^{1/2}$ is the symmetric square root of W. If λ_i and l_i , $i = 1, \ldots, 6$ are the eigenvalues and the corresponding eigenvectors, then the canonical coordinates in different dimensions (after translation to a suitable origin) are $X'W^{-1/2}l_i$, $i = 1, \ldots, 6$, as given Table 8.

The eigenvalue λ_i represents the variance between populations in the *i*th dimension or variance as explained by the *i*th canonical coordinates. The values of λ_i and the percentage of

Table 7. D^2 and T^2 values for differences between sexes within species

	7	Friangulat	ion I	Triangulation II			
Species	D^2	T^2	<i>p</i> -value	D^2	T^2	<i>p</i> -value	
Pan	0.75	1.63	0.141	1.29	2.80	0.013	
Gorilla	5.51	13.12	< 0.001	5.90	14.04	< 0.001	
Pongo	9.04	19.46	< 0.001	9.29	21.36	< 0.001	

Table 8. Canonical coordinates in the first three dimensions

Dimension				
1	2	3		
0.256	1.461	1.526		
1.625	1.906	1.041		
1.863	-1.140	0.539		
0.800	0.749	0.779		
6.763	0.658	1.646		
3.695	2.093	0.0123		
	1 0.256 1.625 1.863 0.800 6.763 3.695	Dimension 1 2 0.256 1.461 1.625 1.906 1.863 -1.140 0.800 0.749 6.763 0.658 3.695 2.093		

Table 9. Percentage of variance explained

i	1	2	3	4	5	6
λ_i	0.9585	0.0384	0.0028	0.0002	0.0000	0.0000
%	95.90	99.70	99.98	100.00	—	_

variance explained by canonical coordinates are given in Table 9. It is seen that the first two canonical coordinates account for 99.7% of the variance and the first three canonical coordinates explain most of the variance due to six angles.

The relative positions of the six populations under study are shown in Fig. 5, where the x and y axes represent the first two canonical coordinates and the third canonical coordinates are plotted on the vertical line to indicate any additional differences between groups in the third dimension. The relative positions of the groups are as inferred in Section 3.1 based on D^2 values and tests of significance.

3.4. Interpretation of Differences Between Populations. When overall differences in shape between populations are indicated by appropriate tests, it may be of interest to examine the nature of differences and to determine whether the differences are localized to some subconfigurations of the landmarks. We illustrate the method for such a study using the male *Pan* and *Pongo* apes.

We consider all possible sets of three out of five landmarks chosen for study. There are 10n such sets giving rise to 10 triangles, and we examine the difference between the two groups in the shape of each triangle. Table 10 gives the mean values of angles for each triangle for each of the two groups and the associated D^2 and T^2 values. Here the Hotelling's T^2 test follows the *F* distribution with 2 and 55 degrees of freedom.

It is seen that the D^2 values for the triangles 125 and 345 are small, indicating that the relative positions of the landmarks 1, 2, and 5 and 3, 4, and 5 are nearly the same for *Pan* and *Pongo*. The major difference is in the relative positions of landmarks 2, 3, and 5, with 2 moving toward 3 and with the angle 235 remaining the same. The mean shapes of the crania of *Pan* and *Pongo* apes are shown in Fig. 6.

This raises the question of whether the difference in the shape of triangle 235 has caused the difference in the shapes of other triangles or whether there are other factors also



FIG. 5. The configuration of the male and female apes in the dimensions of the first three canonical coordinates.

Table 10. Mean angles, Mahalanobis's D^2 , and Hotelling's T^2 for all possible triangles in the *Pan* and *Pongo* samples

	Pan			Pan Pongo					
Δ	ψ	θ	ϕ	ψ	θ	ϕ	D^2	T^2	<i>p</i> -value
123	0.520	0.343	2.280	0.258	0.208	2.676	26.252	186.705	< 0.001
124	0.793	0.668	1.682	0.606	0.592	1.945	18.940	134.706	< 0.001
125	0.792	1.329	1.022	0.750	1.368	1.024	0.624	4.438	0.016
134	0.272	2.326	0.544	0.347	2.050	0.746	7.699	54.757	< 0.001
135	0.272	2.705	0.166	0.492	2.367	0.284	10.621	75.534	< 0.001
145	0.030	3.075	0.037	0.146	2.842	0.156	8.890	63.227	< 0.001
234	0.597	1.658	0.887	0.732	1.457	0.953	6.497	46.210	< 0.001
235	1.258	1.376	0.508	1.652	0.999	0.492	25.954	184.588	< 0.001
345	0.378	0.438	2.326	0.462	0.474	2.206	2.564	18.235	< 0.001
245	0.660	1.814	0.668	0.921	1.473	0.749	19.332	137.492	< 0.001

affecting the differences in the other triangles. To test this phenomenon, let us consider triangles 235, 123, and 234, which specify the configuration of the five landmarks. The D_2^2 value for triangle 235 (2 angles) is 25.954, as given in Table 10. The D_6^2 value for all of the triangles, 235, 123, and 234 (6 angles), is 37.688. The additional D^2 due to triangles 123 and 234 independently of the triangle 235 is $D_6^2 - D_2^2 = 37.688 -$ 25.954 = 11.734, whereas the individual D^2 values due to these triangles are 26.252 and 6.497, respectively (as given in Table 10). Thus the differences is shapes of triangles 123 and 234 are largely explained by the difference in the shape of triangle 235.

The significance of $D_6^2 - D_2^2$ can, however, be tested by Rao's U statistic [see Rao (9), p. 568]:

$$U = \frac{n_1 + n_2 - 6 - 1}{6 - 2} \frac{n_1 n_2 (D_6^2 - D_2^2)}{(n_1 + n_2)(n_1 + n_2 - 2) + n_1 n_2 D_2^2}$$

= 6.017,

which as F on 4 and 51 degrees of freedom (using the values $n_1 = 28, n_2 = 30$) has a p-value of 0.002. The test indicates some



Pango

FIG. 6. Differences in shapes of Pan and Pongo.



FIG. 7. Angular measurements of the human profile.

additional differences due to triangles 123 and 234 to be explained, though smaller in magnitude.

What is the mean configuration of landmarks on an object? There are several definitions in the literature depending on the choice of shape measurements characterizing the configuration of landmarks on an object. We refer to a recent review paper by Molchanov (11) on this subject. We believe that the mean configuration has to be viewed in terms of the mean configurations of all possible triangles formed from different sets of three landmarks, as in Table 10. Further work is in progress.

4. Angular Measurements of Profile Between Landmarks

Although the angular data based on any particular triangulation of landmarks specify the configuration of landmarks, some further data may be generated to characterize the profile between landmarks if available, which may provide some additional information in problems of discrimination and identification. Let us consider the human facial profile (Fig. 7) and the triangle formed by the landmarks h, a, and b. We divide the angle $\angle ahb$, say h^0 , into k equal parts and draw lines at angles h^0/k , $h^0/2k$, ... to the line ah. With k = 4, we have three lines, as shown in Fig. 7, which meet the profile between the landmarks a and b at three points. We now measure the four angles (1, 2, 3, 4), as shown in Fig. 7, which provide measurements on the shape of the profile. The process can be repeated with all the other basic triangles by choosing suitable values of k for each of the triangles. The angles of the basic triangles and the new angles generated by the process described above can be used in statistical analysis.

- Rao, C. R. & Suryawanshi, S. (1996) Proc. Natl. Acad. Sci. USA 93, 12132–12136.
- 2. Lele, S. (1993) Math. Geol. 25, 573-602.
- Bookstein, F. L. (1991) Morphometric Tools for Landmark Data Geometry and Biology (Cambridge Univ. Press, Cambridge, U.K.).
- 4. Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* (Chapman and Hall).
- 5. Pukkila, T. M. & Rao, C. R. (1988) Information Sci. 45, 379-389.
- Mardia, K. V. (1972) Statistics of Directional Data (Academic, New York).
- 7. Rao, C. R. & Ali, H. (1998) Student, in press.
- 8. O'Higgins, P. & Dryden, I. L. (1993) J. Hum. Evol. 24, 183-205.
- 9. Rao, C. R. (1973) *Linear Statistical Inference and Its Applications* (Wiley, New York).
- 10. Rao, C. R. (1948) J. R. Stat. Soc. B 10, 159-203.
- 11. Molchanov, I. S. (1997) Proc. Int. Stat. Inst. 1, 119-122.