Hacettepe Journal of Mathematics and Statistics
Volume 43 (1) (2014), 141 – 152

TYPE I ERROR RATE FOR TWO-SAMPLE TESTS IN STATISTICAL SHAPE ANALYSIS

Gokhan Ocakoglu * † and Ilker Ercan * ‡

Received 19 : 02 : 2013 : Accepted 25 : 06 : 2013

Abstract

Nowadays, with the help of advanced imaging techniques the image or shape of an organ or organism can be used as input data. Therefore, the statistical analysis of shape has recently become more important in the medical and biological sciences. Methods related to two-sample tests have been developed for statistical shape analysis, giving rise to considerable interest in research that evaluates the performance of these tests. In this study, two sample procedures are used to compare the mean shapes from the statistical shape analysis literature according to type I error rate.

Keywords: Statistical shape analysis, two-sample tests, type I error rate.

2000 AMS Classification:

1. Introduction

In the biological and medical sciences, morphometric methods are frequently preferred for examining the morphologic structures of organs or organisms with regard to diseases or environmental factors. Therefore, the statistical analysis of shapes has recently become more important in the medical and biological sciences. Data sets include qualitative and quantitative measurements for use in the statistical analyses associated with medical research. Nowadays, with the help of advanced imaging techniques the image or shape of an organ or organism can be used as input data [1].

Shape is defined as all the geometrical information that remains when location, scale and rotational effects are filtered from an object [2], [3], [4], [5]. Statistical shape analysis is a geometrical analysis of the statistics measured from sets of shapes that determines the features of similar shapes or of different groups comprising similar shapes. Distance between shapes, mean shape and shape variation can be predicted and obtained using statistical shape analysis [3]. A comparison of shapes between groups can also be done at a particular significance level.

*Department of Biostatistics, Faculty of Medicine, Uludag University, Bursa, Turkey.
†Corresponding author. Email: (G. Ocakoglu) ocakoglu@uludag.edu.tr
‡Email: (I. Ercan) ercan@uludag.edu.tr

Inferential methods described in the shape analysis literature make use of landmark configurations that are optimally superimposed via either a least-squares procedure or an analysis of interlandmark distance matrices [6].

Methods concerning two-sample tests have been developed for statistical shape analysis, giving rise to considerable interest in research that evaluates the performance of these tests. In this study, the Hotelling T^2, Goodall's F and James F_j tests as well as the $\lambda_{\min}$ test statistic are used to compare the mean shapes of two samples from the statistical shape analysis literature according to type I error rates derived from various variance values in different sample sizes. This simulation study considers both isotropic and anisotropic cases for which tangent space is used as shape space and considers methods that use complex arithmetic and exploit the geometry of the shape space.

2. Materials and Methods

2.1. Shape Space. The shape space is the set of all possible shapes [3]. For any set of landmarks $\{X_i\}$ in the original Euclidean plane, we can imagine the set of shapes derived by holding all but one of the X's at fixed position and varying that one in a circle about its original position. We would like the metric assigned to shape space (the set of "shapes" of all such sets of X's, correcting for centroid, orientation, and scale, all of which usually change whenever one of the X's moves) to be such that the shapes generated by circles in the original landmark plane are all at the same distance from the original shape $\{X_i\}$ in the shape space. That is, to a circle around one landmark in data space should correspond something very nearly a circle in shape space [7]. Although shape spaces defined by superimposition methods have less dimensions than raw data or non-redundant measurements, they are non-Euclidean and correspond to a curved surface. Nobody will recommend applying traditional statistics directly in this space because traditional statistics relies on the Euclidean metric, which is not the same as the Procrustes one [8]. Special statistical methods (rather than the usual linear multivariate methods) are required to take into account the non-Euclidean geometry of Kendall's shape space for both two and three-dimensional landmarks [4]. To perform usual statistical methods, one must first project the surface of the hyperhemisphere onto a "flat" tangent space where the Euclidean metrics allows us to use Euclidean statistics. The data are projected on a tangent shape space (also called Kendall tangent space or Kent tangent space). The contact between spaces is chosen as the mean shape. Working on variation in the tangent space is a rather perilous estimation since the projection can introduce distortion for the largest distances. However, provided that variation is small, one can assume that the portion of the shape hyperhemisphere and tangent space are nearly flat and nearly confused [8].

The projection onto a Euclidean space can be orthogonal or stereographic. Note that both projections will introduce biases for shapes being very different from the mean shape: the orthogonal projection minimizes large differences while stereographic projection accentuates them. The stereographic projection is produced by adjusting the size scale factor for the configuration to be projected onto the tangent space. To perform this projection, we use simple trigonometric relationships and divide the coordinates of the aligned configurations by the cosine of the Procrustes distance ρ between shapes and the mean shape [8].

In this study the performances of two-sample test procedures that examine differences in mean shape between two independent populations were evaluated in case of using tangent shape space as a shape space. For these test procedures the case in terms of using complex arithmetic and exploiting the geometry of the shape space which is an alternative computational method was also considered for examining tests performances.

2.2. Two-Sample Hotelling T^2 Test. The two-sample Hotelling T^2 test is used to test an alternative hypothesis related to the differences of the mean shapes of two groups and is accordingly applied to shape coordinates [9]. The Hotelling T^2 test assumes that the samples have multivariate normal distributions and equal variance-covariance matrices [10].

Consider two independent random samples $X_1, \ldots, X_{n1}$ and $Y_1, \ldots, Y_{n2}$ from two independent populations with mean shapes $[\mu_1]$ and $[\mu_2]$. To test the hypothesis $H_0 : [\mu_1] = [\mu_2]$, a two-sample Hotelling T^2 test can be performed in the Procrustes tangent space where the pole corresponds to overall pooled full Procrustes mean shape $\widehat{\mu}$. Let $v_1, \ldots, v_{n1}$ and $w_1, \ldots, w_{n2}$ be the partial Procrustes tangent coordinates (with pole $\widehat{\mu}$) [3].

A multivariate normal model is proposed in the tangent space, where $v_i \sim N(\xi_1, \sum_1)$ for $i = 1, \ldots, n_1$, $w_j \sim N(\xi_2, \sum_2)$ for $j = 1, \ldots, n_2$, and the v_i and w_j values are all mutually independent. $\bar{v}$ and $\bar{w}$ and S_v, S_w represent the sample means and sample covariance matrices respectively (with divisors n_1 and n_2) in each group. If the covariance matrices are assumed to be equal ($\sum_1 = \sum_2$), then the squared Mahalanobis distance between $\bar{v}$ and $\bar{w}$ is given by Equation-2.1.

$$D^2 = (\bar{v} - \bar{w})^T S_U^+ (\bar{v} - \bar{w}) \tag{2.1}$$

where $S_U = (n_1 S_1 + n_2 S_2)/(n_1 + n_2 - 2)$ and S_U^+ is the Moore-Penrose generalized inverse of S_U. Under the null hypothesis, we have $\xi_1 = \xi_2$ and the two-sample Hotelling statistic, which is given by Equation 2.2

$$F_H = \frac{n_1 n_2 (n_1 + n_2 - M - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)M} D^2 \tag{2.2}$$

where $M = 2d - 2$ is the dimension of the planar shape space. The test statistic has an $F_{M, n_1 + n_2 - M - 1}$ distribution under the null hypothesis [2], [3].

2.3. James F_j Test. When covariances are not assumed to be equal, an alternative method is to use the statistic proposed by James, which represents an effort to solve the multivariate Behrens-Fisher problem [2], [6].

$$F_j = (\bar{v} - \bar{w})^T \left(\frac{1}{n_1} S_v + \frac{1}{n_2} S_w \right)^+ (\bar{v} - \bar{w}) \tag{2.3}$$

The J-statistic has an asymptotic χ^2_M distribution under the null hypothesis regardless of whether $\sum_1$ and $\sum_2$ are equal, and we reject the null hypothesis for large values of this statistic [2].

2.4. Two-Sample Goodall's F Test. Goodall presented a statistical framework for analyzing Procrustes shape data and developed a possible F test. This test is based on the Procrustes chord distance and should work under the assumption that variation is isotropic and is equal for each landmark [8]. This assumption implies that the variances of all landmarks (that is, the amount of dispersion) are expected to be the same. The assumption also implies that the patterns of dispersion across landmarks are expected to be uncorrelated [11].

If $\sum_1 = \sum_2 = \sum$ and we have isotropic covariance structure ($\sum = \sigma^2 I$) [2].

In an isotropic variance structure, the diagonal elements and the variance values of the covariance matrix are equal for each landmark, and all elements except the diagonal elements are equal to zero. Perhaps the simplest type of covariance structure for the perturbation distribution is one in which all landmarks are perturbed with the same variance irrespective of direction. This isotropic variance structure is easy to visualize, but may not be biologically realistic in the study of certain biological structures or certain populations [12]. An isotropic normal model with mean μ and transformed by an additional location, rotation and scale effects are given by Equation-2.4

$$x_i = \beta_i(\mu + E_i)\Gamma_i + 1_k\gamma_i^T \quad vec(E_i) \sim N(0, \sigma^2 I_{km}) \tag{2.4}$$

where $\beta_i > 0$ (scale), $\Gamma_i \in SO(m)$ (rotation) and $\gamma_i \in \mathbb{R}^m$ (translation), and σ is small.

Consider independent random samples $x_1, x_2, \ldots, x_n$ from a population modeled by Equation-2.4 with μ_1 and $y_1, y_2, \ldots, y_n$ from Equation-2.4 with mean μ_2. Both populations are assumed to have a common σ^2 variance for each coordinate [3].

We wish to test $H_0 : [\mu_1] = [\mu_2](= [\mu_0])$ against $H_1 : [\mu_1] \neq [\mu_2]$. $[\widehat{\mu}_1]$ and $[\widehat{\mu}_2]$ are the full Procrustes means of each sample. Under the H_0 hypothesis, with a small σ the Procrustes distances are approximately distributed as

$$\sum_{i=1}^{n_1} d_F^2(X_i, \widehat{\mu}_1) \sim \tau_0^2 \chi^2_{(n_1-1)M} \tag{2.5}$$

$$\sum_{i=1}^{n_2} d_F^2(Y_i, \widehat{\mu}_2) \sim \tau_0^2 \chi^2_{(n_2-1)M} \tag{2.6}$$

$$d_F^2(\widehat{\mu}_1, \widehat{\mu}_2) \sim \tau_0^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \chi^2_M \tag{2.7}$$

where $\tau = \sigma/\delta$, $\delta_0 = S(\mu_0)$ and d_F^2 represents the squared full Procrustes distance between two configurations. In addition, these statistics are approximately mutually independent [3]. Hence, under the null hypothesis, we have the approximate distribution as given in equation-2.8.

$$F_G = \frac{n_1 + n_2 - 2}{n_1^{-1} + n_2^{-1}} \frac{d_F^2(\widehat{\mu}_1, \widehat{\mu}_2)}{\sum_{i=1}^{n_1} d_F^2(X_i, \widehat{\mu}_1) + \sum_{i=1}^{n_2} d_F^2(Y_i, \widehat{\mu}_2)} \sim F_{M,(n_1+n_2-2)M} \tag{2.8}$$

We reject the null hypothesis for large values of this test statistic. The Hotelling T^2 procedure is less powerful than Goodall's F test, for which the isotropic normal model holds [3], [13].

2.5. $\lambda_{\min}$ Test Statistic. Amaral et al. [2] proposed a novel bootstrap approach to k-sample testing problems in which each sample consists of a set of real or complex unit vectors. The basic assumption is that the distribution of the sample mean shape (or direction or axis) is highly concentrated [6]. Consider k samples of unit vectors in $\mathbb{C}^d$ (in most traditional applications, $d = 2; 3$, but sometimes the case $d \geq 4$ is also relevant), and let $\widehat{\mu}_i$ be the estimator of μ_0 (i.e., the mean shape under the hypothesis) based on sample i, for $i = 1, \ldots, k$. Assume that $n^{\frac{1}{2}} \widehat{M}_i \mu_0 \xrightarrow{D} \mathbb{C}N_{d-1}(0, G_i)$ for $i = 1, \ldots, k$ where G_i denotes asymptotic covariance matrix has full rank and $\widehat{M}_i$ represents a projection onto the tangent space at $\widehat{\mu}_i$ [6].

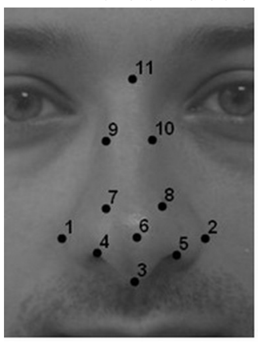
Define $\widehat{A}_0 = n \sum_{i=1}^{k} \widehat{M}_i^T \widehat{G}_i^{-1} \widehat{M}_i$ and $T_0(\mu) = 2\mu^T \widehat{A}_0 \mu$, where T denotes the conjugate transpose, μ is a complex unit vector and $\widehat{G}_i$ is a consistent estimator of G_i. We thus obtain

$$\lambda_{\min} \equiv \min_{\mu:\|\mu\|=1} T_0(\mu) = T_0(\widehat{\mu}) \tag{2.9}$$

where $\lambda_{\min}$ is the smallest eigenvalue of $\widehat{A}_0$ and $\widehat{\mu}_0$ is the corresponding unit eigenvector [2], [6]. It is proven that $\lambda_{\min} \xrightarrow{D} \chi^2_{2(k-1)(d-1)}$ as $n \to \infty$ under the null hypothesis of equality of means across populations [6].

2.6. A Simulation Study. In this study we aim to compare type I error rates of the tabular, bootstrap and permutation adaptations of Hotelling T^2, Goodall's F and James F_j tests as well as the $\lambda_{\min}$ test statistic. A mean vector and a variance-covariance matrix are computed from a data set obtained from the landmark markings of the nose in the anterior views of the faces of 50 subjects. Eleven landmarks (Figure 1) are applied to the images in the manner described by Ercan et al. [14]. In the present study, the data are simulated from a multivariate normal distribution under isotropic and anisotropic models.

FIGURE 1. Landmark markings for the source data set used in the simulation study.



The samples for which type I error rates are examined in the simulation study are $n_1 = n_2 = 20, 50, 100$ and 500.

A mean vector that computed from a data set obtained from the landmark markings as mentioned above is $(\bar{x}_1, \ldots, \bar{x}_{11}, \bar{y}_1, \ldots, \bar{y}_{11})$ = (501, 590, 546, 522, 568, 546, 521, 570, 532, 563, 547, 399, 398, 384, 398, 397, 409, 425, 426, 469, 469, 500).

Variance values are determined to be $0.001, 0.01, 0.05, 0.1, 0.5, 1, 5, 737, 1703$ and 2949 in the isotropic case. The values $737, 1703$ and 2949 values are the minimum, maximum and mean variance values of the variance-covariance matrix, which contains real values from the sample data set.

Isotropic structures are used in studies and when comparing the methods; however it is not the case as in real-world applications; therefore, in our study we also compare methods by simulating with anisotropic structures. The real variance-covariance matrix computed from the sample data set is used as input for the simulation of the anisotropic case.

In the examination of type I error rate in the simulation study, it is assumed that related tests use tangent space as shape space, that they use complex arithmetic and that they exploit the geometry of shape space.

The simulation study has been conducted with 1000 replications, and the number of bootstrap and permutation resamples is set to 100.

We used TPSDIG 2.04 software to mark the landmarks on the images. The simulation study and analyses were performed using R 2.12.0 software [15].

3. Results and Discussion

In Table 1, we give type I error rates as determined for both cases according to the exploitation of shape space, according to various variance values for the isotropic model and according to the variance-covariance matrix computed from the real data set for the anisotropic model in different sample sizes.

It has been observed that applications of statistical shape analysis have recently been used more than ever before in medical and biological sciences to compare the structures of shapes [14], [16], [17], [18]. For example, forensics analyses [19], computer-assisted neurosurgery methods [20] anthropological studies [14], [17], [18], [21] and MRI-based morphological analyses of the brain [22], [23], [24] make use of statistical shape analysis. Therefore, it is of great importance that shape objects be recognized, measured and compared.

Newly developed methods utilize two-sample tests in statistical shape analysis, which is a geometric morphometric concept. However, more emphasis has been placed on studies of the comparative performance of related tests. In this study, we aim to compare the type I error rates of the Hotelling T^2, Goodall's F and James F_j tests as well as the $\lambda_{\min}$ test statistic, which are all used in the shape analysis literature to compare mean shapes. In this simulation study, the performance of tabular, bootstrap and permutation adaptations of the related procedures are examined in terms of type I error rate. We also consider isotropic and anisotropic cases for different variance values and sample sizes using the tangent space as the shape space. Finally, we consider related procedures that use complex arithmetic and exploit the geometry of the shape space.

We examined the procedures of bootstrap adaptations through simulation results, considered isotropic covariance structure, exploited tangent space and used complex arithmetic with the geometry of the shape space, thus evaluating small samples. In light of these findings, the application of the Hotelling T^2, James F_j and Goodall's F tests in tangent space put the type I error rate under the determined nominal level. Additionally, we observe that the type I error rates remained under the nominal level following the application of $\lambda_{\min}$ test statistic with the Hotelling T^2, Goodall's F and James F_j tests when complex arithmetic was applied and the geometry of the shape space was exploited. In a similar study of small samples, Brombin and Salmaso [6] conducted the Hotelling T^2, Goodall's F and James F_j tests and generally found that the type I error rate was under the nominal level in the isotropic covariance structure when using complex arithmetic with the geometry of the shape space. Brombin and Salmaso [6] also observed a value close to the determined nominal level when using the $\lambda_{\min}$ test statistic. Amaral et al. [2] carried out a similar study with small samples and observed a value close to

Table-1: Type I error rates for n_1=n_2=20, 50, 100, 500 and σ^2= 0.001, 0.01, 0.05, 0.1 in the case of using shape space as tangent space and exploiting complex arithmetic with geometry of shape space.

		$\sigma^2 = 0.001$				$\sigma^2 = 0.01$				$\sigma^2 = 0.05$				$\sigma^2 = 0.1$			
		n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500
Using Tangent space as shape space	H_bootstrap	0.000	0.024	0.051	0.049	0.000	0.033	0.037	0.055	0.000	0.021	0.036	0.066	0.000	0.027	0.040	0.048
	H_permutation	0.047	0.042	0.057	0.048	0.037	0.045	0.054	0.056	0.048	0.040	0.046	0.057	0.062	0.046	0.042	0.050
	H_tabular	0.018	0.024	0.044	0.044	0.032	0.041	0.037	0.056	0.047	0.042	0.044	0.061	0.052	0.052	0.043	0.043
	G_ bootstrap	0.017	0.020	0.047	0.041	0.019	0.036	0.033	0.058	0.019	0.038	0.036	0.053	0.028	0.041	0.038	0.041
	G_ permutation	0.046	0.032	0.050	0.041	0.051	0.050	0.043	0.057	0.045	0.047	0.045	0.063	0.060	0.053	0.057	0.051
	G _ tabular	0.044	0.030	0.044	0.035	0.053	0.048	0.037	0.059	0.054	0.049	0.041	0.056	0.059	0.052	0.052	0.042
	J_ bootstrap	0.000	0.024	0.051	0.049	0.000	0.033	0.037	0.055	0.000	0.021	0.036	0.066	0.000	0.027	0.040	0.048
	J_ permutation	0.047	0.032	0.057	0.048	0.037	0.045	0.054	0.056	0.048	0.040	0.046	0.057	0.062	0.046	0.042	0.050
	J_ tabular	0.120	0.035	0.053	0.044	0.167	0.051	0.042	0.056	0.191	0.062	0.046	0.062	0.228	0.066	0.052	0.045
Using complex arithmetic and exploit geometry of shape space	H_ bootstrap	0.000	0.012	0.047	0.049	0.000	0.032	0.038	0.059	0.000	0.022	0.039	0.054	0.000	0.031	0.041	0.046
	H_ permutation	0.050	0.034	0.043	0.042	0.043	0.051	0.044	0.056	0.047	0.048	0.038	0.061	0.051	0.054	0.046	0.046
	H_ tabular	0.044	0.033	0.059	0.045	0.041	0.048	0.042	0.060	0.047	0.042	0.044	0.061	0.052	0.052	0.043	0.043
	G_ bootstrap	0.020	0.017	0.040	0.039	0.030	0.042	0.036	0.056	0.019	0.036	0.042	0.050	0.026	0.042	0.042	0.043
	G_ permutation	0.050	0.037	0.053	0.039	0.050	0.050	0.043	0.060	0.047	0.044	0.044	0.058	0.048	0.050	0.055	0.047
	G _ tabular	0.054	0.032	0.051	0.039	0.055	0.049	0.042	0.060	0.053	0.049	0.041	0.056	0.059	0.052	0.052	0.042
	J_ bootstrap	0.000	0.012	0.047	0.049	0.000	0.032	0.038	0.059	0.000	0.022	0.039	0.054	0.000	0.031	0.041	0.046
	J_ permutation	0.050	0.034	0.043	0.042	0.043	0.051	0.044	0.056	0.047	0.048	0.038	0.061	0.051	0.054	0.046	0.046
	J_ tabular	0.197	0.066	0.063	0.046	0.198	0.067	0.049	0.060	0.191	0.062	0.046	0.062	0.228	0.066	0.052	0.045
	λ_{min}_ bootstrap	0.045	0.029	0.044	0.041	0.056	0.058	0.043	0.059	0.049	0.046	0.038	0.053	0.072	0.038	0.045	0.043
	λ_{min}_ permutation	0.052	0.033	0.051	0.037	0.050	0.057	0.038	0.058	0.044	0.051	0.052	0.060	0.056	0.050	0.051	0.049
	λ_{min}_ tabular	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.108	0.112	0.076	0.066	0.286	0.106	0.076	0.047

Table-1 (continued): Type I error rates for n_1=n_2=20, 50, 100, 500 and σ^2= 0.5, 1, 5, 737 in the case of using shape space as tangent space and exploiting complex arithmetic with geometry of shape space.

		$\sigma^2 = 0.5$				$\sigma^2 = 1$				$\sigma^2 = 5$				$\sigma^2 = 737$			
		n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500
Using Tangent space as shape space	H_bootstrap	0.000	0.023	0.041	0.043	0.000	0.028	0.045	0.052	0.000	0.024	0.039	0.054	0.000	0.023	0.040	0.047
	H_permutation	0.045	0.054	0.047	0.048	0.049	0.045	0.055	0.045	0.054	0.044	0.045	0.048	0.047	0.049	0.043	0.046
	H_tabular	0.049	0.046	0.055	0.051	0.047	0.049	0.058	0.048	0.052	0.041	0.045	0.054	0.046	0.049	0.047	0.048
	G_ bootstrap	0.022	0.028	0.042	0.046	0.019	0.040	0.044	0.047	0.020	0.034	0.044	0.054	0.021	0.035	0.038	0.048
	G_ permutation	0.054	0.047	0.046	0.046	0.060	0.050	0.055	0.042	0.049	0.047	0.049	0.050	0.047	0.045	0.037	0.052
	G _ tabular	0.046	0.045	0.048	0.049	0.057	0.050	0.058	0.044	0.046	0.046	0.049	0.053	0.046	0.051	0.041	0.049
	J_ bootstrap	0.000	0.023	0.041	0.043	0.000	0.028	0.045	0.052	0.000	0.024	0.039	0.054	0.000	0.023	0.040	0.047
	J_ permutation	0.045	0.054	0.047	0.048	0.049	0.045	0.055	0.045	0.054	0.044	0.045	0.048	0.047	0.049	0.043	0.046
	J_ tabular	0.222	0.066	0.056	0.053	0.207	0.066	0.065	0.050	0.232	0.048	0.051	0.055	0.188	0.070	0.053	0.048
Using complex arithmetic and exploit geometry of shape space	H_ bootstrap	0.000	0.023	0.051	0.047	0.000	0.022	0.045	0.047	0.000	0.028	0.042	0.059	0.000	0.031	0.045	0.044
	H_ permutation	0.052	0.045	0.055	0.053	0.051	0.046	0.059	0.050	0.056	0.041	0.047	0.054	0.050	0.054	0.048	0.053
	H_ tabular	0.049	0.046	0.055	0.051	0.047	0.049	0.058	0.048	0.052	0.041	0.045	0.054	0.047	0.052	0.048	0.047
	G_ bootstrap	0.022	0.029	0.046	0.044	0.024	0.029	0.048	0.045	0.023	0.038	0.038	0.048	0.008	0.039	0.035	0.044
	G_ permutation	0.049	0.042	0.058	0.050	0.053	0.044	0.054	0.045	0.042	0.051	0.048	0.053	0.053	0.052	0.046	0.051
	G _ tabular	0.046	0.045	0.048	0.049	0.057	0.050	0.059	0.044	0.048	0.048	0.049	0.053	0.486	0.524	0.548	0.568
	J_ bootstrap	0.000	0.023	0.051	0.047	0.000	0.022	0.045	0.047	0.000	0.028	0.042	0.059	0.000	0.031	0.045	0.044
	J_ permutation	0.052	0.045	0.055	0.053	0.051	0.046	0.059	0.050	0.056	0.041	0.047	0.054	0.050	0.054	0.048	0.053
	J_ tabular	0.222	0.066	0.056	0.053	0.207	0.066	0.065	0.050	0.232	0.058	0.051	0.055	0.185	0.071	0.053	0.047
	λ_{min}_ bootstrap	0.009	0.026	0.047	0.048	0.013	0.036	0.046	0.046	0.010	0.029	0.038	0.054	0.004	0.039	0.042	0.041
	λ_{min}_ permutation	0.042	0.038	0.052	0.059	0.057	0.045	0.044	0.048	0.049	0.048	0.043	0.053	0.047	0.050	0.043	0.049
	λ_{min}_ tabular	0.281	0.117	0.078	0.055	0.264	0.110	0.087	0.054	0.294	0.109	0.075	0.057	0.240	0.109	0.073	0.048

Table-1 (continued): Type I error rates for $n_1=n_2=20, 50, 100, 500$ and $\sigma^2= 1703, 2949$ and anisotropic covariance structure in the case of using shape space as tangent space and exploiting complex arithmetic with geometry of shape space.

		$\sigma^2 = 1703$				$\sigma^2 = 2949$				Anisotropic covariance structure			
		n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500	n=20	n=50	n=100	n=500
Using Tangent space as shape space	H_bootstrap	0.000	0.034	0.037	0.050	0.000	0.022	0.034	0.046	0.000	0.016	0.038	0.045
	H_permutation	0.053	0.052	0.043	0.050	0.049	0.042	0.044	0.052	0.060	0.048	0.050	0.047
	H_tabular	0.054	0.054	0.047	0.054	0.053	0.042	0.041	0.053	0.046	0.049	0.050	0.045
	G_ bootstrap	0.030	0.040	0.033	0.047	0.024	0.041	0.027	0.043	0.052	0.041	0.051	0.051
	G_ permutation	0.057	0.055	0.041	0.053	0.051	0.052	0.038	0.055	0.061	0.046	0.051	0.050
	G _ tabular	0.057	0.054	0.044	0.054	0.053	0.047	0.033	0.048	0.169	0.140	0.122	0.154
	J_ bootstrap	0.000	0.034	0.037	0.050	0.000	0.022	0.034	0.046	0.000	0.016	0.038	0.045
	J_ permutation	0.053	0.052	0.043	0.050	0.049	0.042	0.044	0.052	0.060	0.048	0.050	0.047
	J_ tabular	0.214	0.073	0.055	0.055	0.212	0.066	0.046	0.054	0.207	0.064	0.055	0.050
Using complex arithmetic and exploit geometry of shape space	H_ bootstrap	0.000	0.033	0.031	0.036	0.000	0.016	0.018	0.028	0.000	0.022	0.039	0.005
	H_ permutation	0.054	0.059	0.045	0.053	0.051	0.046	0.037	0.046	0.051	0.056	0.050	0.042
	H_ tabular	0.053	0.055	0.050	0.051	0.056	0.043	0.041	0.053	0.046	0.049	0.050	0.045
	G_ bootstrap	0.011	0.038	0.033	0.054	0.004	0.021	0.027	0.047	0.047	0.042	0.050	0.049
	G_ permutation	0.093	0.065	0.059	0.050	0.146	0.077	0.046	0.051	0.069	0.047	0.053	0.054
	G _ tabular	0.914	0.931	0.927	0.936	0.988	0.992	0.995	0.997	0.168	0.145	0.122	0.151
	J_ bootstrap	0.000	0.033	0.031	0.036	0.000	0.016	0.018	0.028	0.000	0.022	0.039	0.045
	J_ permutation	0.054	0.059	0.045	0.053	0.051	0.046	0.037	0.046	0.051	0.056	0.050	0.042
	J_ tabular	0.218	0.075	0.053	0.054	0.216	0.067	0.044	0.055	0.207	0.064	0.055	0.047
	λ_{min}_ bootstrap	0.006	0.040	0.036	0.051	0.003	0.023	0.028	0.048	0.016	0.040	0.048	0.040
	λ_{min}_ permutation	0.059	0.055	0.037	0.044	0.050	0.051	0.035	0.045	0.061	0.055	0.051	0.051
	λ_{min}_ tabular	0.243	0.110	0.062	0.058	0.239	0.116	0.064	0.056	0.315	0.146	0.117	0.093

In Table-1, H indicates Hotelling T^2 test, G indicates Goodall's F test and J indicates James F_J test respectively.

the determined level in terms of type I error rates in related procedures. As for large samples, while the type I error rates converged to the nominal level in both usages of shape space, we found results under the nominal level in the simulation study of high variance values.

In the simulation study in which we exploited the variance-covariance matrix of real landmark values, the anisotropic covariance structure and the procedures of bootstrap adaptations, we found that type I error rates stayed under the nominal level according to the Hotelling T^2, Goodall's F and James F_j tests as well as the $\lambda_{\min}$ test statistic for both usages of shape space in small samples. When large samples were evaluated, we found that the type I error rates remained under the determined nominal level only when the Hotelling T^2 test was applied in the case of exploiting complex arithmetic with the geometry of the shape space.

Following the examination of the permutation adaptation of procedures through the simulation results and considering the isotropic covariance structure, the tests showed an overall performance in all sample sizes in both usages of shape space. However, the Goodall's F test tends to overestimate the nominal level in small samples in the case of exploiting complex arithmetic with the geometry of the shape space. In a similar study of small samples, Amaral et al. [2] found an overall results that were close the nominal level for the type I error rates; however, Amaral et al. [2] reported that as the variance values in the Goodall's F test increased, the related procedure tended to overestimate the nominal level of the type I error rate. Compared to the variance values in Amaral et al. [2], the variance values of the Goodall's F test are close to the values of the nominal level of the type I error rate. Brombin and Salmaso [6] stated that the Hotelling T^2 and James F_j tests showed similar values but that the Goodall's F test and the $\lambda_{\min}$ test statistic tended to underestimate the nominal level. In the anisotropic covariance structure, the examined procedures showed similar results to the nominal type I error rate in small and large sample sizes.

When tabular versions of procedures were analyzed through simulation results, the James F_j test tended to overestimate the nominal level in small samples in both usages of shape space in the case of isotropic covariance structure. The Hotelling T^2 test underestimated the nominal level in small samples in tangent space with reference to type I error rate in low variance values, but the Goodall's F test overestimated the nominal level in the case of exploiting complex arithmetic with the geometry of the shape space in high variance values. We found that comparison with the $\lambda_{\min}$ test statistic generally underestimated and overestimated the nominal level. We found that the James F_j and Goodall's F tests as well as the $\lambda_{\min}$ test statistic underestimated and overestimated the nominal level; on the other hand, the Hotelling T^2 test revealed values close to the nominal level, which Brombin and Salmaso [6] also observed in a similar study of small samples in the case of exploiting isotropic covariance structure and in the cases of related procedures that use complex arithmetic and exploit the geometry of the shape space. Amaral et al. [2] also found that the Goodall's F test and the $\lambda_{\min}$ test statistic overestimated the nominal level; however, the Hotelling T^2 and James F_j tests resulted in values close to the nominal level in a similar study of small samples. The Goodall's F test overestimated the nominal level in large samples when exploiting complex arithmetic with the geometry of the shape space and in the case of high variance values. It was observed that the Goodall's F test and $\lambda_{\min}$ test statistic overestimated the nominal level in both usages of shape space in anisotropic covariance structure.

When the present study is compared with the similar studies [2], [6] in the literature, performances of two-sample test procedures used in this study were examined in terms of both using tangent space as a shape space and using complex arithmetic with exploiting the geometry of shape space. This study also differs from other literatures in terms of using variance-covariance matrix of real-life data set to examine the performances of related procedures in anisotropic case. In addition, it has been observed that the variance values given in simulation scenarios in similar studies are smaller than the variance values of real-life data sets. For this reason, in this study two-sample test procedures' performances were also examined for large variance values computed from a real-life data set. Present study also differs in terms of including large sample size values.

4. Conclusions

As predicted, the results of the present study indicate that tests perform better with large samples than with small samples. For small samples, permutation test adaptations gave the most favorable results in all isotropic and anisotropic covariance structures. For large samples, permutation test adaptations gave the most favorable results with regard to type I error rate in all low and high variance values and in all isotropic and anisotropic covariance structures. It was concluded that bootstrap adaptations of tests gave the most unfavorable results in all isotropic and anisotropic covariance structures in small samples.

Conflict of Interests. The authors declare that they have no conflict of interest.

Acknowledgements. This study was supported by Uludag University Scientific Research Project. Project number: (UAP(T)-2010/34). The numerical calculations reported in this paper were performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TR-Grid e-Infrastructure).

References

[1] Ercan, I., Ocakoglu, G., Sigirli, D. and Ozkaya, G. *Statistical shape analysis and usage in medical sciences*, Turkiye Klinikleri J Biostat. 4(1), 27-35 (2012).

[2] Amaral, G.J.A, Dryden, I. and Wood, A.T.A. *Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis*, Journal of the American Statistical Association. 102(478), 695-707 (2007).

[3] Dryden, I.L. and Mardia, K.V. Statistical shape analysis (Vol. 4). Wiley, New York, 1998.

[4] Rohlf, F.J. *Shape statistics. Procrustes superimpositions and tangent spaces*, Journal of Classification. 16(2), 197-223, 1999.

[5] Rohlf, F.J. *On the use of shape spaces to compare morphometric methods*, Hystrix-the Italian Journal of Mammalogy, 11(1), 12-13, 2000.

[6] Brombin, C. and Salmaso, L. *Multi-aspect permutation tests in shape analysis with small sample size*, Computational Statistics & Data Analysis, 53(12), 3921-3931, 2009.

[7] Bookstein, F. L. Morphometric tools for landmark data: Geometry and Biology. Cambridge Univ. Press, New York, 1991.

[8] Claude, J. Morphometrics with R. Springer Verlag, 2008.

[9] Zelditch, M., Swiderski, D.L., Sheets, H.D. and Fink, W.L. Geometric morphometrics for biologists. A primer. Academic Press, 2004.

[10] Hammer, Ø. and Harper, D.A.T. Paleontological data analysis. Malden (Massachusetts), Blackwell Publishing Ltd Oxford, 2005.

[11] Richtsmeier, J.T., Lele, S. and Cole, T. *Landmark morphometrics and the analysis of variation*, In. Hallgrimsson, B, Hall, BK(eds.)Variation. A Central Concept in Biology. Amsterdam. Elsevier, 2005.

[12] Lele, S. and Richtsmeier, J.T. *An invariant approach to statistical analysis of shapes*, Florida. Chapman & Hall, 2001.

[13] Goodall, C. *Procrustes methods in the statistical analysis of shape*, Journal of the Royal Statistical Society. Series B (Methodological), 53(2), 285-339, 1991.

[14] Ercan, I., Etoz, A., Guney, I., Ocakoglu, G., Ozdemir, S.T., Kan, I. and Kahveci, R. *Statistical shape analysis of nose in Turkish young adults*, Journal of Craniofacial Surgery, 18(1), 219, 2007.

[15] Dryden, I. L. Shapes package. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[16] Ercan, I., Ocakoglu, G., Guney, I. and Yazici, B. *Adaptation of generalizability theory for inter-rater reliability for landmark localization*, International Journal of Tomography & Statistics, 9(S08), 51-58, 2008.

[17] Ercan, I., Ozdemir, S.T, Etoz, A., Sigirli, D., Tubbs, R.S., Loukas, M. and Guney, I. *Facial asymmetry in young healthy subjects evaluated by statistical shape analysis*, Journal of Anatomy, 213(6), 663-669, 2008.

[18] Ozdemir, S.T., Ercan, I., Ozkaya, G., Cankur, N.S. and Erdal Y.S. *Geometric Morphometric Study and Cluster Analysis of Late Byzantine and Modern Human Crania*, Collegium Antropologicum, 34(2), 493-499, 2010.

[19] De Greef, S., Claes, P., Vandermeulen, D., Mollemans, W., Suetens, P. and Willems, G. *Large-scale in-vivo Caucasian facial soft tissue thickness database for craniofacial reconstruction*, Forensic science international, 159, S126-S146, 2006.

[20] Aldridge, K., Marsh, J.L., Govier, D. and Richtsmeier, J.T. *Central nervous system phenotypes in craniosynostosis*, Journal of Anatomy, 201(1), 31-39, 2002.

[21] Jonke, E., Schaefer, K., Freudenthaler, J.W., Prossinger, H. and Bookstein, F.L. *A cephalometric comparison of skulls from different time periods-the bronze age, the 19th century and the present*, Collegium antropologicum, 27(2), 789-802, 2003.

[22] Bookstein, F.L., Sampson, P.D., Streissguth, A.P. and Connor, P.D. *Geometric morphometrics of corpus callosum and subcortical structures in the fetal-alcohol-affected brain*, Teratology. 64(1).4-32, 2001.

[23] Colak, C., Ercan, I., Dogan, M., Ozdemir, S.T., Sener, S. and Alkan, A. *Detecting the shape differences of the corpus callosum in behcet's disease by statistical shape analysis*, Anatomical Record-Advances in Integrative Anatomy and Evolutionary Biology, 294(5), 870-874, 2011.

[24] Ozdemir, S.T., Ercan, I., Sevinc, O., Guney, I., Ocakoglu, G., Aslan, E. and Barut, C. *Statistical shape analysis of differences in the shape of the corpus callosum between genders The Anatomical Record*, Advances in Integrative Anatomy and Evolutionary Biology, 290(7), 825-830, 2007.