

# Examining the Variations in the Results of the Hotelling $T^2$ Test in Case of Changing Baseline Landmarks in the Bookstein Coordinates

Ilker Ercan<sup>1</sup> · Deniz Sigirli<sup>1</sup> · Guven Ozkaya<sup>1</sup>

Received: 20 March 2014 / Revised: 15 August 2014 / Accepted: 24 October 2014 / Published online: 22 July 2015  
© International Association of Scientists in the Interdisciplinary Areas and Springer-Verlag Berlin Heidelberg 2015

**Abstract** In many biological and biomedical investigations, the most effective way to analyze the forms of whole biological organs or organisms is by recording geometric locations of landmark points. If we want to compare shapes, then individuals should be translated, rotated and scaled in such a way that all of the individuals lie in a standard position and are centered. Bookstein conducted this process by choosing two landmarks as reference landmarks. Each individual is translated, rotated and scaled according to these reference landmarks. The aim of the present study was to examine the change in the  $p$  values in the case of choosing different baseline landmarks when performing the Hotelling  $T^2$  test, which is commonly used when comparing two sample shape configurations based on Bookstein coordinates. For this purpose, the changes in the  $p$  values were investigated in shape configurations that are composed of a different number of landmarks by taking all of the possible paired landmark combinations at different variance levels and sample sizes. As a result of the present study, it was observed that with the increase in the landmark number, the number of possible baseline landmark combinations also increases and, for this reason, a substantial number of variations occur in the  $p$  values. Therefore, it is an important to decide which landmarks should be taken as reference landmarks when using the Bookstein coordinates.

**Keywords** Bookstein coordinates · Hotelling  $T^2$  test · Statistical shape analysis · Morphometrics

---

✉ Ilker Ercan  
ercan@uludag.edu.tr

<sup>1</sup> Department of Biostatistics, Faculty of Medicine, Uludag University, Gorukle Campus, 16059 Bursa, Turkey

## 1 Introduction

Most of the studies in medicine are related to the examination of geometrical properties of an organ or organism. While in these studies, the datasets used for statistical analysis are composed of quantitative or qualitative measuring values, currently, an organ's or organism's appearance or shape is treated as input data to develop imaging techniques [1].

In many biological and biomedical investigations, the most effective way to analyze forms of whole biological organs or organisms is to record the geometric locations of landmark points [2]. Several different morphometric algorithms are available for analyzing landmark data. All of these methods utilize the same raw data: the coordinate values of landmarks in two or three dimensions. Some current morphometric techniques that use landmark-based methods are the finite-element scaling analysis (FESA), thin-plate splines (TPS), procrustes analysis and Euclidean distance matrix analysis (EDMA) [3].

When digitizing landmarks from a number of specimens, it is in general impossible to ensure that each specimen is measured in the same position and orientation. If we want to compare shapes, it is, therefore, necessary to translate and rotate the specimens into a standardized position and orientation. In addition, we would like to scale the specimens to a standard size. Bookstein suggested performing this task by selecting two landmarks to form a baseline [4, 5].

With the shape coordinates that are suggested by Bookstein, different analysis can be performed on these coordinates. While some of these analyses can be performed by comparing two samples to examine the differences between the shape configurations, some of them can be performed on the same subject to examine a single individual's shape configuration or can compare two individuals with each other [4, 6].

One of the tests that is implemented on Bookstein shape coordinates to compare shape configurations of the two samples is the Hotelling  $T^2$  test. This test assumes that the samples are independent, that Bookstein shape coordinates have a multivariate normal distribution, and that the samples are drawn from populations that have the same covariance matrix [7].

The aim of the present study is to examine the change in  $p$  values for choosing different baseline landmarks when performing the Hotelling  $T^2$  test, which is commonly used to compare two sample shape configurations that are based on Bookstein coordinates. For this purpose, the changes in  $p$  values were investigated in shape configurations that comprise a different number of landmarks, by taking all of the possible paired landmark combinations at different variance levels and sample sizes.

## 2 Methods

### 2.1 Bookstein Coordinates

Let  $(x_j, y_j), j = 1, \dots, k$ , be  $k \geq 3$  landmarks in a plane ( $m = 2$  dimensions). Bookstein suggests removing the similarity transformations by translating, rotating and rescaling such that landmarks 1 and 2 are sent to a fixed position. If landmark 1 is sent to  $(0, 0)$  and landmark 2 is sent to  $(1, 0)$ , then suitable shape variables are the coordinates of the remaining  $k - 2$  coordinates after these operations. To preserve symmetry, we consider the coordinate system in which the baseline landmarks are sent to  $(-0.5, 0)$  and  $(0.5, 0)$ .

Bookstein coordinates  $(u_j^B, v_j^B)^T, j = 3, \dots, k$ , are the remaining coordinates of an object after translating, rotating and rescaling the baseline to  $(-0.5, 0)$  and  $(0.5, 0)$ , which results in the following:

$$u_j^B = \{(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)\} / D_{12}^2 - \frac{1}{2}, \tag{1}$$

$$v_j^B = \{(x_2 - x_1)(y_j - y_1) - (y_2 - y_1)(x_j - x_1)\} / D_{12}^2 \tag{2}$$

where  $j = 3, \dots, k, D_{12}^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 > 0$  and  $-\infty < u_j^B, v_j^B < \infty ((v_j^B, v_j^B)^T$  is the Bookstein coordinates and  $T$  shows the transpose).

If the baseline is taken as  $(0, 0)$  and  $(1, 0)$ , then there is no  $-0.5$  in the equation for  $u_j^B$  [8].

### 2.2 Two Independent Sample Hotelling’s $T^2$ Test

Consider two random samples  $X_1, X_2, \dots, X_{n_1}$  and  $Y_1, Y_2, \dots, Y_{n_2}$  from independent populations with mean

shapes  $[\mu_1]$  and  $[\mu_2]$ . To test between  $H_0 : [\mu_1] = [\mu_2]$  versus  $H_1 : [\mu_1] \neq [\mu_2]$ , Hotelling’s  $T^2$  test can be performed.

Hotelling’s  $T^2$  two-sample test based on Bookstein coordinates has been explicitly described. Consider two samples of shapes with sizes  $n_1$  and  $n_2$ . The multivariate normal model is proposed for Bookstein coordinates ( $M$ -vectors)  $v_1, \dots, v_{n_1}$  and  $w_1, \dots, w_{n_2}$ :

$$v_i \sim N_M(\xi_1, \Sigma), \quad w_j \sim N_M(\xi_2, \Sigma), \\ i = 1, \dots, n_1; \quad j = 1, \dots, n_2,$$

and the  $v_i$  and  $w_j$  are all mutually independent; common variance matrices are assumed, and  $M$  is the dimension of the shape space (which is  $2k - 4$  for planar data). We write  $\bar{v}, \bar{w}$  and  $S_v, S_w$  for the sample means and sample covariance matrices (with divisors  $n_1$  and  $n_2$ ) in each group. The squared Mahalanobis distance between  $\bar{v}$  and  $\bar{w}$  is

$$D^2 = (\bar{v} - \bar{w})^T S_u^{-1} (\bar{v} - \bar{w}), \tag{3}$$

where  $S_u = (n_1 S_v + n_2 S_w) / (n_1 + n_2 - 2)$  and  $S_u^{-1}$  is the inverse of  $S_u$ , which will usually exist. Under  $H_0, \xi_1 = \xi_2$ , and the test statistic given below is used.

$$F = \frac{n_1 n_2 (n_1 + n_2 - M - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)M} D^2 \tag{4}$$

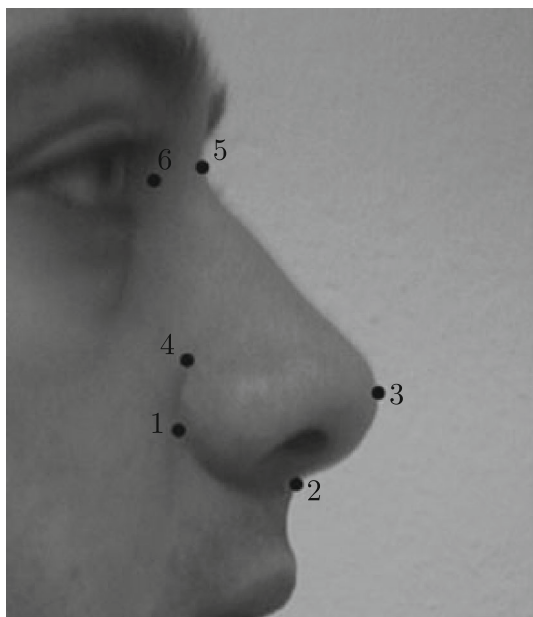
The test statistic has an  $F_{M, n_1 + n_2 - M - 1}$  distribution under  $H_0$ . Hence,  $H_0$  is rejected for large values of  $F$  [8].

### 2.3 Sample

To investigate the effect of the baseline choice on the results of the Hotelling  $T^2$  test, a total of six landmarks are included in the present study, which are marked on the nose images of 100 subjects (Fig. 1). Landmarks have been identified through some of the anatomical landmarks of nose which have been described in at [9]. These landmarks were selected so that they can show the effects of the landmarks which were found near and far. In the present study, to examine the effect of baseline choice in the different number of landmarks, separate evaluations were performed for 3 (landmark 1, landmark 2 and landmark 3), 4 (landmarks 1, 2, 3 and 4), 5 (landmarks 1, 2, 3, 4 and 5) and 6 (landmarks 1, 2, 3, 4, 5 and 6) landmarks.

### 2.4 Design of Simulation

In the present study, simulations were performed according to an isotropic model for four different shape configurations, which comprise 3, 4, 5 and 6 landmarks, respectively. In the simulation study, the mean vectors of landmarks were obtained from a real dataset as below.



**Fig. 1** Landmarks used in the present study

$$X = [586.19, 667.34, 624.90, 440.70, 451.41, 416.55] \quad (5)$$

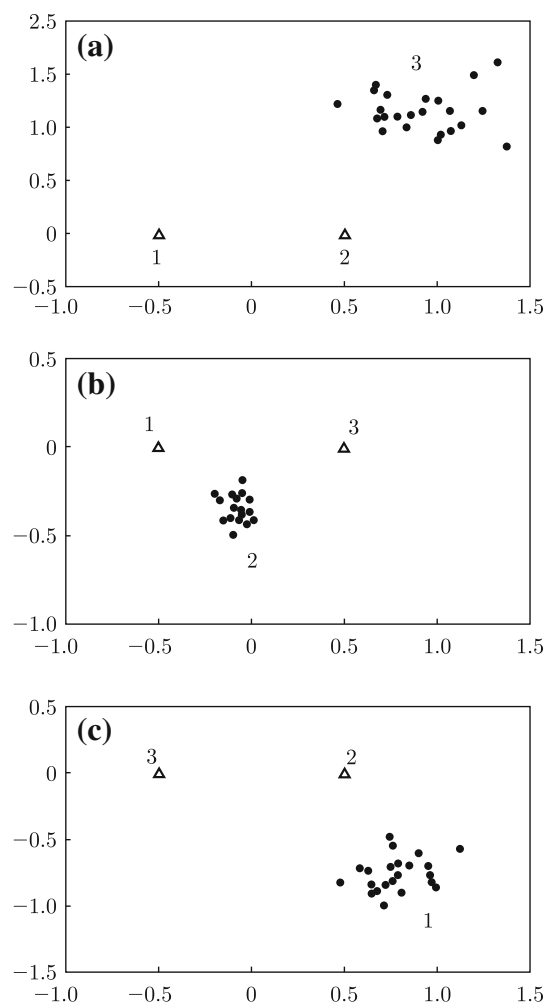
Data were generated from a multivariate normal distribution for the variances 0.01 and 0.05. Additionally, to investigate high variance levels, data were generated for the minimum and maximum variances (737 and 2949) obtained from the real dataset.

After generating the data, the Hotelling  $T^2$  test was performed on the Bookstein coordinates, which are obtained by taking landmark 1 and landmark 2, landmark 1 and landmark 3, and landmark 2 and landmark 3 as baseline landmarks, with the same dataset for the shape configuration, which consists of three landmarks (Fig. 2).

The same application was performed for the other shape configurations, which consist of 4, 5 and 6 landmarks, by taking all of the possible landmark pairs as baseline landmarks.

Maximum change rates were calculated from the percentage changes of maximum and the minimum  $p$  values, which were obtained from the test results that were performed on the same individual (shape) by taking different landmark combinations as baseline landmarks. Change rates obtained for all of the individuals (shapes) were compared according to the landmark number in different sample sizes and different variance levels. The Kruskal–Wallis test and the Mann–Whitney  $U$  test were used in these comparisons. Bonferroni correction was applied to the post hoc comparison results ( $\alpha^* = \alpha/k = 0.05/6 = 0.008$ ).

In the simulation study, 1,000,000 repetitions were performed for the  $n = 25, 50, 100, 250, 500$  and 1000



**Fig. 2** Bookstein coordinates of **a** baseline landmarks 1 and 2, **b** baseline landmarks 1 and 3, and **c** baseline landmarks 2 and 3

sample sizes. Simulations were performed on an HP Z800 Workstation with a 2.40 GHz dual processor and 16 Gb of RAM, using R software.

### 3 Results

Simulation results are given in Table 1 for  $\sigma^2 = 0.01$ , in Table 2 for  $\sigma^2 = 0.05$ , in Table 3 for  $\sigma^2 = 737$  and in Table 4 for  $\sigma^2 = 2949$ .

In both large and small sample sizes, it is seen that changes in the  $p$  values increase as the landmark number increases. Also the increase in the variances shows an effect in the way of increasing the changes in the  $p$  values. In the present study, while the lowest change in the  $p$  values was obtained for the lowest variance level ( $\sigma^2 = 0.01$ ) and for the three-landmark situation, the biggest change in the  $p$  values obtained for the highest variance level which

**Table 1** Statistical comparison of the maximum percentage changes in the *p* values according to the number of landmarks, for a variance of 0.01

Sample size		Number of landmarks				Overall <i>p</i>	Post hoc <i>p</i>
		3	4	5	6		
<i>n</i> = 25	Median	0.0075	0.0119	0.0119	0.0148	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0039–0.0131	0.0066–0.0195	0.0065–0.0194	0.0081–0.0241		4–5: <0.001 4–6: <0.001 5–6: <0.001
<i>n</i> = 50	Median	0.0074	0.0120	0.0120	0.0149	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0038–0.013	0.0066–0.0197	0.0066–0.0196	0.0082–0.0243		4–5: 0.005 4–6: <0.001 5–6: <0.001
<i>n</i> = 100	Median	0.0073	0.0120	0.0120	0.0150	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0038–0.013	0.0065–0.0198	0.0065–0.0197	0.0082–0.0245		4–5: 0.227 4–6: <0.001 5–6: <0.001
<i>n</i> = 250	Median	0.0073	0.0119	0.0120	0.0151	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0037–0.0129	0.0065–0.0198	0.0065–0.0197	0.0082–0.0247		4–5: 0.963 4–6: <0.001 5–6: <0.001
<i>n</i> = 500	Median	0.0073	0.0119	0.0120	0.0151	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0037–0.0129	0.0065–0.0197	0.0065–0.0198	0.0082–0.0247		4–5: 0.062 4–6: <0.001 5–6: <0.001
<i>n</i> = 1000	Median	0.0073	0.0119	0.0120	0.0150	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0037–0.0129	0.0064–0.0197	0.0065–0.0197	0.0081–0.0247		4–5: 0.010 4–6: <0.001 5–6: <0.001

Q1 first quartile, Q3 third quartile

we obtained from the real dataset ( $\sigma^2 = 2949$ ) and for the six-landmark situation.

### 4 Discussion

Morphometrics is a field that is concerned with the variation and change in the form (size and shape) of objects and organisms. There are various methods that are used for the

analysis of shapes. These methods can be considered to be traditional methods that use linear distances, ratios or angles in the analysis or can be considered to be modern geometric morphometric methods, which include statistical shape analysis applications based on landmarks [1, 10, 11]. One of the approaches in modern geometric morphometrics is the use of Bookstein coordinates. Bookstein coordinates are the coordinates of the remaining  $k - 2$  coordinates after removing the similarity transformations by translating,

**Table 2** Statistical comparison of the maximum percentage change in the  $p$  values according to the number of landmarks, for a variance of 0.05

Sample size		Number of landmarks				Overall $p$	Post hoc $p$
		3	4	5	6		
$n = 25$	Median	0.0166	0.0265	0.0263	0.0326	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0087–0.029	0.0147–0.0431	0.0146–0.0427	0.018–0.0531		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 50$	Median	0.0164	0.0266	0.0266	0.0331	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0085–0.0289	0.0146–0.0436	0.0146–0.0432	0.0181–0.0536		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 100$	Median	0.0163	0.0266	0.0266	0.0332	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0084–0.0288	0.0145–0.0437	0.0145–0.0435	0.0182–0.054		4–5: 0.009 4–6: <0.001 5–6: <0.001
$n = 250$	Median	0.0162	0.0264	0.0266	0.0334	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0084–0.0286	0.0144–0.0436	0.0145–0.0436	0.0181–0.0543		4–5: 0.059 4–6: <0.001 5–6: <0.001
$n = 500$	Median	0.0162	0.0264	0.0266	0.0334	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0083–0.0286	0.0144–0.0436	0.0145–0.0437	0.0182–0.0544		4–5: 0.001 4–6: <0.001 5–6: <0.001
$n = 1000$	Median	0.0161	0.0264	0.0267	0.0333	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.0083–0.0286	0.0144–0.0436	0.0145–0.0436	0.0181–0.0544		4–5: 0.007 4–6: <0.001 5–6: <0.001

Q1 first quartile, Q3 third quartile

rotating and rescaling such that landmarks 1 and 2 are sent to a fixed position. There are debates on the choice of baseline landmarks in the Hotelling  $T^2$  test for when two groups are compared based on Bookstein coordinates.

There are no found similar studies to the present study in the literature, but Rohlf [12] stated that different choices for the base (reference) landmarks of the shape lead to somewhat different results. The present study was conducted to

investigate this discussion. For this reason, we have performed a simulation study with different landmark numbers, in cases of changing baseline landmarks, for low and high variance levels and in small, moderate and large sample sizes.

In the present study, significant differences were observed in terms of the variations in the  $p$  values derived from the changes in the baseline landmarks, according to

**Table 3** Statistical comparison of the maximum percentage change in the  $p$  values according to the number of landmarks, for a variance of 737

Sample size		Number of landmarks				Overall $p$	Post hoc $p$
		3	4	5	6		
$n = 25$	Median	0.6652	0.8105	0.8481	0.8759	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4755–0.8223	0.6868–0.9036	0.7401–0.9259	0.7811–0.9417		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 50$	Median	0.6749	0.8249	0.8625	0.8873	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4821–0.8319	0.7047–0.9127	0.7612–0.9342	0.7996–0.9477		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 100$	Median	0.6820	0.8337	0.8731	0.8976	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4878–0.8381	0.7161–0.9185	0.7766–0.9403	0.8155–0.9530		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 250$	Median	0.6865	0.8422	0.8834	0.9078	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4909–0.8434	0.7263–0.9243	0.7914–0.9461	0.8308–0.9585		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 500$	Median	0.6899	0.8467	0.8891	0.9135	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4930–0.8464	0.7314–0.9271	0.7994–0.9494	0.8400–0.9616		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 1000$	Median	0.6914	0.8495	0.8936	0.9178	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4937–0.8484	0.7350–0.9292	0.8057–0.9519	0.8466–0.9638		4–5: <0.001 4–6: <0.001 5–6: <0.001

Q1 first quartile, Q3 third quartile

the landmark numbers. These differences were observed in both high and low variance levels and in small, moderate and large sample sizes. It was observed that the variations in the  $p$  values increased with an increase in the landmark number. When the effect of the increment of the landmark number on the  $p$  values was examined in the low and high variance levels, major variations were observed in the  $p$  values, especially at high variance levels.

## 5 Conclusion

While working with a small number of landmarks leads to fewer effects on the  $p$  values because of the selection of baseline landmarks, this approach can be insufficient for describing the shape. As we know, the more landmarks we study with, the more power we obtain when describing the shape. However, the results of the present study which is

**Table 4** Statistical comparison of the maximum percentage change in the  $p$  values according to the number of landmarks, for a variance of 2949

Sample size		Number of landmarks				Overall $p$	Post hoc $p$
		3	4	5	6		
$n = 25$	Median	0.6668	0.8199	0.8936	0.9063	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4770–0.8232	0.7004–0.9085	0.8057–0.9519	0.8350–0.9550		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 50$	Median	0.6772	0.8313	0.8871	0.9157	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4854–0.8332	0.7153–0.9158	0.8046–0.9448	0.8513–0.9594		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 100$	Median	0.6829	0.8393	0.8949	0.9230	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4902–0.8387	0.7248–0.9211	0.8160–0.9492	0.8630–0.9692		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 250$	Median	0.6881	0.8461	0.9019	0.9295	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4932–0.8445	0.7328–0.9258	0.8253–0.9532	0.8736–0.9667		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 500$	Median	0.6902	0.8498	0.9058	0.9330	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4940–0.8464	0.7370–0.9285	0.8310–0.9556	0.8788–0.9688		4–5: <0.001 4–6: <0.001 5–6: <0.001
$n = 1000$	Median	0.6923	0.8521	0.9082	0.9359	<0.001	3–4: <0.001 3–5: <0.001 3–6: <0.001
	Q1–Q3	0.4949–0.8490	0.7395–0.9302	0.8347–0.9573	0.8830–0.9703		4–5: <0.001 4–6: <0.001 5–6: <0.001

Q1 first quartile, Q3 third quartile

based on the Bookstein coordinates show that with the increase in the landmark number, the number of possible baseline landmark combinations also increases and, for that reason, a substantial number of variations occur in the  $p$  values. Therefore, it is important to determine which landmarks should be taken as reference landmarks in the case of using Bookstein coordinates.

**Acknowledgments** This study was supported by Uludag University Scientific Research Project Committee UAP(T)-2010/34.

## References

1. Ercan I, Ocakoglu G, Sigirli D, Ozkaya G (2012) Statistical shape analysis and usage in medical sciences. *Turk Klin J Biostat* 4:27–35
2. Bookstein FL (1991) *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, Cambridge
3. Richtsmeier JT, Cheverud JM, Lele S (1992) Advances in anthropological morphometrics. *Annu Rev* 21:283–305
4. Bookstein FL (1984) A statistical method for biological shape comparisons. *J Theor Biol* 107:475–520

5. Hammer Ø, Harper DAT (2005) *Paleontological data analysis*. Wiley, Massachusetts
6. Baskan EB, Ercan I, Ozdemir ST, Ozkaya G, Yilmaz M, Ozgenel Y, Saricaoglu H (2013) A case of Parry–Romberg syndrome: examining efficacy of lipofilling by statistical shape analysis. *Turk Klin J Med Sci* 33:284–289
7. Rohlf FC (2000) Statistical power comparisons among alternative morphometric methods. *Am J Phys Anthropol* 111:463–478
8. Dryden I, Mardia KV (1998) *Statistical shape analysis*. Wiley, Chichester
9. Ercan I, Etoz A, Guney I, Ocakoglu G, Ozdemir ST, Kan I, Kahveci R (2007) Statistical shape analysis of nose in Turkish young adults. *J Craniofac Surg* 18(1):219–224
10. Ocakoglu G, Ercan I (2013) Traditional and modern morphometrics. *Turk Klin J Biostat* 5:37–41
11. Sigirli D, Ercan I (2013) Growth and allometry in modern morphometrics. *Turk Klin J Biostat* 5:42–48
12. Rohlf FC (1999) Shape statistics: procrustes superimpositions and tangent space. *J Classif* 16:197–223