

Chapter 2

Theoretical Aspects on Permutation Tests and Shape Analysis

“Normality is a myth; there never has, and never will be, a normal distribution.”.

Testing for Normality. *Biometrika*, **34**, 209–242, 1947.

R.C. Geary

Abstract This chapter deals with inferential aspects in the field of shape analysis. At first we review inferential methods known in the shape analysis literature, highlighting some drawbacks of using Hotelling’s T^2 test statistic. Then we propose an extension of the NonParametric Combination (NPC) methodology to compare shape configurations of landmarks.

Without doubts, the more landmark and semilandmarks are collected for a given object, the more shape information is gathered in the configuration describing the object, and the better is the quantitative evaluation of its morphological features.

However, most of the traditional inferential methods in shape analysis are parametric and they often require large sample size while, in practice, researchers have to deal with few objects/subjects and many landmarks, implying over-dimensional spaces and loss of power.

On the other hand, NPC tests represent an appealing alternative since they are distribution-free and allow for quite efficient solutions when the number of cases is lower than the number of variables (i.e., (semi)landmarks coordinates). This allows to obtain better representations of shapes even in the presence of small sample size. Finally, NPC methodology enables to provide global as well as local evaluation of shapes: it is then possible to establish whether in general two shapes are different and which landmark/subgroup of landmarks mainly contributes to differentiate shapes under study.

Keywords Closed testing procedures • Goodall’s F -test • Hotelling’s T^2 test statistic • NonParametric Combination methodology • Multi-aspect approach • Resampling methods

2.1 Inference and Shape Analysis

The statistical community has shown an increased interest in shape analysis in the last decade and particular efforts have been addressed to the development of powerful statistical methods based on model for shape variation of entire configurations of point corresponding to the locations of morphological landmarks. Rohlf (2000) reviews the main tests used in the field of shape analysis and compares the statistical power of various tests that have been proposed to test for equality of shape in two populations. Even if his work is limited to the simplest case of homogeneous, independent, spherical variation at each landmark and the sampling experiments emphasize the case of triangular shapes, it allows the practitioners to choose the method that has the highest statistical power under a set of assumptions that are appropriate for the data. Through a simulation study, he found that Goodall's F -test had the highest power followed by T^2 -test using Kendall tangent space coordinates. Power for T^2 -tests using Bookstein shape coordinates was good if the baseline was not the shortest side of the triangle. The Rao and Suryawanshi shape variables had much lower power when triangles were not close to being equilateral. Power surfaces for the EDMA-I T statistic revealed very low power for many shape comparisons including those between very different shapes. Power surface for the EDMA-II Z statistic depended strongly on the choice of baseline used for size scaling (Rohlf 2000). We remind the reader that EDMA stands for Euclidean Distance Matrix Analysis (EDMA). Technical details on the above-mentioned tests are provided in Sect. 2.2.

All the above-mentioned tests are based on quite stringent assumptions. In particular, the tests based on the T^2 statistic (e.g., T^2 -tests using Bookstein, Kendall tangent space coordinates, Rao and Suryawanshi shape variables, like Rao-d (1996) and Rao-a (1998)) require independent samples, homogeneous covariance matrices, and shape coordinates distributed according to the multivariate normal distribution. We remark that Hotelling's T^2 test statistic is derived under the assumption of population multivariate normality and it may not be very powerful unless there are a large number of observations available (Dryden and Mardia 1998). It is well known in the literature that Hotelling's T^2 test is formulated to detect any departures from the null hypothesis and therefore often lacks power to detect specific forms of departures that may arise in practice, i.e., the T^2 test fails to provide an easily implemented one-sided (directional) hypothesis test (Blair et al. 1994).

Goodall's F test requires a restrictive isotropic model and assumes that the distributions of the squared Procrustes distances are approximately Chi-squared distributed.

If we consider the methods based on interlandmark distances, EDMA-I T assumes independent samples and the equality of the covariance matrices in the two populations being compared (Lele and Cole 1996), while EDMA-II Z assumes only independent samples and normally distributed variation at each landmark.

In order to complete the review on main tests used in shape analysis, we recall the pivotal bootstrap methods for k -sample problems, in which each sample consists

of a set of real (the directional case) or complex unit vectors (the two-dimensional shape case), proposed in the paper by [Amaral et al. \(2007\)](#). The basic assumption here is that the distribution of the sample mean shape (or direction or axis) is highly concentrated. This is a substantially weaker assumption than is entailed in tangent space inference ([Dryden and Mardia 1998](#)) where observations are presumed highly concentrated. In this paper test statistics like λ_{\min} , Hotelling T^2 , Goodall F , and James F_J have been compared and corresponding p -values have been obtained using both resampling methods (bootstrap or permutation test) and the usual table. In particular, with reference to the pivotal statistic λ_{\min} , 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 $\hat{\mathbf{m}}_i$ be the estimator of \mathbf{m}_0 (i.e., mean shape under H_0) based on sample i , for $i = 1, \dots, k$. Assume that $n^{1/2}\hat{\mathbf{M}}_i\mathbf{m}_0$ has an asymptotic complex normal distribution $\text{CN}_{d-1}(0, \mathbf{G}_i)$, $i = 1, \dots, k$, where \mathbf{G}_i has full rank and $\hat{\mathbf{M}}_i$ represents a projection onto the tangent space at $\hat{\mathbf{m}}_i$.

Define $\hat{\mathbf{A}}_0 = n \sum_{i=1}^k \hat{\mathbf{M}}_i^* \hat{\mathbf{G}}_i^{-1} \hat{\mathbf{M}}_i$ and $T_0(\mathbf{m}) = 2\mathbf{m}^* \hat{\mathbf{A}}_0 \mathbf{m}$, where the $*$ denotes conjugate transpose and \mathbf{m} is a complex unit vector (i.e., $\mathbf{m}^* \mathbf{m} = 1$), thus obtaining

$$\lambda_{\min} \equiv \min_{\mathbf{m}: \|\mathbf{m}\|=1} T_0(\mathbf{m}) = T_0(\hat{\mathbf{m}}_0)$$

where λ_{\min} is the smallest eigenvalue of $\hat{\mathbf{A}}_0$ and $\hat{\mathbf{m}}_0$ is the corresponding unit eigenvector. For further mathematical details we refer the reader to [Amaral et al. \(2007\)](#).

It is proved that this statistic has a limiting Chi-squared distribution $\chi_{2(k-1)(d-1)}^2$ under the null hypothesis of equality of means across populations ([Amaral et al. 2007](#)). Another statistic used in this paper is the James statistic (see [Seber 1984](#)) that represents an effort to solve the multivariate Behrens–Fisher problem and it is given by

$$F_J = (\bar{\mathbf{v}} - \bar{\mathbf{w}})^T \left(\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right)^- (\bar{\mathbf{v}} - \bar{\mathbf{w}}),$$

where $\mathbf{v}_i \sim N(\boldsymbol{\xi}_1, \boldsymbol{\Sigma}_1)$ for $i = 1, \dots, n_1$ and $\mathbf{w}_j \sim N(\boldsymbol{\xi}_2, \boldsymbol{\Sigma}_2)$, for $j = 1, \dots, n_2$ are the partial Procrustes tangent coordinates, \mathbf{v}_i and \mathbf{w}_i are mutually independent, $\bar{\mathbf{v}}$, $\bar{\mathbf{w}}$ and \mathbf{S}_1 , \mathbf{S}_2 are the sample means and sample covariance matrices (with divisors n_1 and n_2) in each group. It is proved that $F_J \sim \chi_M^2$. Although authors focus mainly on the version of the statistic in which neither isotropy within populations nor constant dispersion structure across populations is assumed, they explain how to modify the statistic so that either or both of these assumptions can be incorporated ([Amaral et al. 2007](#)).

As pointed out in [Good \(2000\)](#), the assumption of equal covariance matrices may be unreasonable especially under the alternative, the multinormal model in the tangent space may be doubted and sometimes there are few individuals and many landmarks, implying over-dimensional spaces and loss of power for the Hotelling's

T^2 test. Hence when sample sizes are too small, or the number of landmarks is too large, it is essentially inefficient to assume that observations are normally distributed. An alternative procedure is to consider a permutation version of the test (see Good 2000; Dryden and Mardia 1993; Bookstein 1997; Terriberry et al. 2005). Permutation methods are distribution-free, allow us for quite efficient solutions when the number of cases is less than the number of covariates and may be tailored for sensitivity to specific treatment alternatives providing one-sided as well as two-sided tests of hypotheses (Blair et al. 1994).

In the wake of these considerations, we propose an extension of the NonParametric Combination (NPC) methodology (Pesarin 2001; Pesarin and Salmaso 2010; Brombin 2009; Brombin et al. 2008; Brombin and Salmaso 2009; Brombin et al. 2009a,b; Alfieri et al. 2012). We observe that a key condition for applying permutation tests is the exchangeability of observations under the null hypothesis (Pesarin 2001; Pesarin and Salmaso 2010). Generally permutation tests require homogeneous covariance matrices under H_0 in order to guarantee exchangeability, thus relaxing the stringent assumptions of parametric tests. This is consistent with the notion that if H_0 is true, this implies the equality in multivariate distribution of observed variables, i.e., there is no effect at all.

2.2 Technical Details on Tests Known in Shape Analysis Literature

2.2.1 Hotelling's T^2 and Goodall's F Tests

Let us define two independent random samples X_1, \dots, X_{n_1} and Y_1, \dots, Y_{n_2} from independent populations with mean shapes $[\mu_1]$ and $[\mu_2]$. The hypotheses system is given by

$$H_0 : [\mu_1] = [\mu_2] \text{ versus } H_1 : [\mu_1] \neq [\mu_2]$$

Let v_1, \dots, v_{n_1} and w_1, \dots, w_{n_2} be the partial Procrustes tangent coordinates, where

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

are all mutually independent with common covariance matrices.

An Hotelling's T^2 two sample test in the Procrustes tangent space could be carried out, after performing a GPA superimposition on all $n_1 + n_2$ individuals to compute the average shape. Each specimen is then fit to this overall mean (also called the pole $\hat{\mu}$). Let \bar{v} , \bar{w} and S_v , S_w be respectively the sample means and sample covariance matrices (with divisors n_1 and n_2) in each group. The Mahalanobis distance squared between \bar{v} and \bar{w} is

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

where S_u^- is the Moore–Penrose generalized inverse of $S_u = (n_1 S_v + n_2 S_w) / (n_1 + n_2 - 2)$. Under H_0 we have $\xi_1 = \xi_2$, and we use the test statistic

$$F = \frac{n_1 n_2 (n_1 + n_2 - M - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)M} D^2 \sim F_{M, n_1 + n_2 - M - 1},$$

where $M = km - m - m(m - 1)/2 - 1$ is the dimension of the tangent space. Further Hotelling's T^2 versions can be calculated using Kendall tangent space coordinate, Bookstein coordinates (Edge Superimposition), and Rao and Suryawanshi shape variables, Rao-d (1996) and Rao-a (1998).

The tests based on the T^2 test statistic require independent samples, homogeneous covariance matrices, and shape coordinates normally distributed. We remark that Hotelling's T^2 test statistic is derived under the assumption of population multivariate normality and it may not be very powerful unless there are a large number of observations available (Dryden and Mardia 1998). It is well known in the literature that Hotelling's T^2 test is formulated to detect any departures from the null hypothesis and therefore often lacks power to detect specific forms of departures that may arise in practice, i.e., the T^2 test fails to provide an easily implemented one-sided (directional) hypothesis test (Blair et al. 1994).

Goodall's F -test (1991) compares the Procrustes distance between the means of two samples to the amount of variation found within the samples. It uses a generalized least-squares Procrustes analysis to compute the average shape for each sample. It is given by

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

This result is valid for small σ and $M = 2k - 4$ for 2D data ($M = 3k - 7$ for 3D data). It assumes that configurations are isotropic normal perturbations from mean configurations, and the distributions of the squared Procrustes distances are approximately Chi-squared distributions. When the sphericity assumption is true, this test shows higher power than the usual T^2 -test, especially when sample sizes are small.

2.2.2 Euclidean Distance Matrix Analysis Methods

The form of an object X is all the geometrical information about X that is invariant under translation and rotation (rigid-body transformations) and the form matrix $\mathbf{FM}(X)$ is the $k \times k$ matrix of all pairs of inter-landmark distances in the configuration X .

Let X_1, X_2, \dots, X_n be landmark coordinate matrices for a sample of n individuals from population X . To estimate the average form matrix $\mathbf{FM}(X)$ for the population X , calculate

- $e_{lm,i}$ the squared Euclidean distance between landmarks l and m for the i -th individual,
- $\bar{e}_{lm} = n^{-1} \sum_{i=1}^n e_{lm,i}$ and $s^2 = n^{-1} \sum_{i=1}^n (e_{lm,i} - \bar{e}_{lm})^2$,
- $\hat{\delta}_{lm} = (\bar{e}_{lm}^2 - s^2(e_{lm}))^{0.5}$.
- $\mathbf{FM}(X) = (\hat{\delta}_{lm}^{0.5})_{lm=1,2,\dots,k}$

In the same way the average form matrix $\mathbf{FM}(Y)$ of the sample Y_1, Y_2, \dots, Y_n is calculated from population Y .

EDMA-I test statistic (Lele and Richtsmeier, 1991; Lele, 1993) is given by

$$T = \max(\mathbf{FDM}(X, Y)) / \min(\mathbf{FDM}(X, Y)),$$

where $\mathbf{FDM}(X, Y)$ is the form difference matrix for samples X and Y that is obtained as

$$\mathbf{FDM}(X, Y)_{i,j} = \mathbf{FM}(X)_{i,j} / \mathbf{FM}(Y)_{i,j} \forall i, j = 1, \dots, k,$$

where $\mathbf{FM}(X)$ and $\mathbf{FM}(Y)$ are the average form matrices below the convention $0/0 = 0$.

EDMA-II test statistic (Lele and Cole, 1995; 1996) is calculated as

$$Z = \max |\mathbf{S}_X - \mathbf{S}_Y|,$$

where \mathbf{S}_X and \mathbf{S}_Y are two size-scaled average form matrices (proper scaling factor could be edge length or continuous function of edge lengths). Bootstrap procedures are used to estimate the null distribution of T and Z test statistic.

EDMA-I T assumes independent samples and the equality of the covariance matrices in the two populations being compared, while EDMA-II Z requires independent samples and normally distributed variation at each landmark.

2.3 NPC Approach to Shape Analysis

Let X_1 be the $n_1 \times (k \times m)$ matrix of raw landmark coordinates of specimens belonging to the first group. Similarly X_2 is the $n_2 \times (k \times m)$ matrix of raw landmark coordinates of specimens belonging to the second group. Let $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ be the $n \times (k \times m)$ matrix of raw landmark coordinates of all specimens, i.e., our data set, where $n = n_1 + n_2$. Hence X is a matrix of data with specimens in the rows and landmark coordinates in columns. In the permutation context, in order to denote data sets, it could be useful the unit-by-unit representation given by $X = \{X_{hji}, i = 1, \dots, n, j = 1, 2, h = 1, \dots, km\}$, where it is intended that first $n_1 \times km$ data in the list belong to first sample and the rest to the second.

In practice, denoting by (a_1^*, \dots, a_n^*) a permutation of the labels $(1, \dots, n)$, $X^* =$

$\{X_{hji}^* = X_{hj}(a_i^*), i = 1, \dots, n, j = 1, 2, h = 1, \dots, km\}$ is the related permutation of X , so that $X_{h1}^* = \{X_{h1i}^* = X_{h1}(a_i^*), i = 1, \dots, n_1, h = 1, \dots, km\}$ and $X_{h2}^* = \{X_{h2i}^* = X_{h2}(a_i^*), i = n_1 + 1, \dots, n, h = 1, \dots, km\}$ are the two permuted samples, respectively. For simplicity, we may assume that the landmark coordinates in tangent space behave according to the following model:

$$X_{hji} = \mu_h + \delta_{hj} + \sigma_h Z_{hji},$$

$i = 1, \dots, n, j = 1, 2, h = 1, \dots, km$, where

- k is the number of landmarks in m dimensions;
- μ_h represents a population constant for the h -th variable;
- δ_{hj} represents treatment effect (i.e. the noncentrality parameter) in the j -th group on the h -th variable which, without loss of generality, is assumed to be $\delta_{h1}=0$, $\delta_{h2} \leq (or \geq) 0$;
- σ_h are scale coefficients specific to the h -th variable;
- Z_{hji} are random errors assumed to be exchangeable with respect to treatment levels, independent with respect to units, with null mean vector ($\mathbb{E}(Z) = 0$), and finite second moment.

Hence landmark coordinates in the first group differ from those in the second group by a “quantity” δ , where δ is the km -dimensional vector of effects. Again, X_{hji}^* , $i = 1, \dots, n, j = 1, 2, h = 1, \dots, km$, indicates a permutation of the original data.

Therefore the specific hypotheses may be expressed as

$$H_0 : \bigcap_{h=1}^{km} \{X_{h1} \stackrel{d}{=} X_{h2}\} \quad \text{vs.} \quad H_1 : \bigcup_h^{km} \{(X_{h1} + \delta) \stackrel{d}{>} X_{h2}\},$$

where $\stackrel{d}{>}$ stands for distribution (or stochastic) dominance.

With $T_h^o(0)$ and $T_h^*(0)$ we indicate respectively the observed and permutation values of T_h when $\delta = 0$, i.e., under H_0 .

The assumptions regarding the set of partial tests $\mathbf{T} = \{T_h, h = 1, \dots, km\}$ necessary for NPC are:

1. All permutation partial test T_h are marginally unbiased and significant for large values, so that they are stochastically larger in H_1 than in H_0 .
2. All permutation partial tests T_h are consistent, that is,

$$\Pr\{T_h \geq T_{h\alpha} | U, H_{1h}\} \rightarrow 1, \forall \alpha > 0, h = 1, \dots, km,$$

as n tends to infinity, where $T_{h\alpha} < +\infty$ is the critical value of T_h at level α . In order to obtain global traditional consistency it suffices that at least one partial test is consistent (Pesarin 2001; Pesarin and Salmaso 2010).

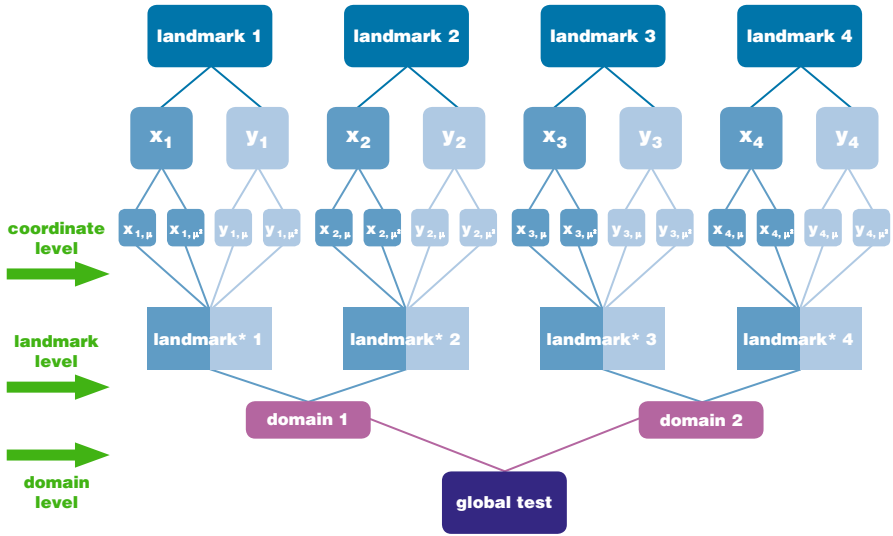


Fig. 2.1 Different levels of combination

Let $\lambda_h, h = 1, \dots, km$ be the set of p -values associated with partial tests in \mathbf{T} that are positively dependent in the alternative and this irrespective of dependence relations among component variables in X .

In shape analysis field, $h = 1, \dots, km$ represents the k landmarks in m dimensions. In order to apply NPC methodology, usually the hypothesis testing problem is broken down into two stages, considering both the coordinate and the landmark level (and, if present, the domain level too). Hence, we formulate partial test statistics for one-sided hypotheses and then we consider the global test T'' obtained after combining at the first stage with respect to m , then with respect to k (of course, this sequence may be reversed).

For example, if we consider 4 landmarks, first of all one can derive a test for each coordinate (x and y coordinates in 2D case) of each landmark. Once decided the aspects of interest, one could focus on the coordinate level or on the landmark level, after combining coordinates, or on the domain level as well and finally on the global test (see Fig. 2.1).

2.3.1 A Suitable Algorithm

We now illustrate the algorithm for calculating the multivariate test, in its simplest version. Then we may add a multi-aspect procedure and adjust partial p -values for multiplicity through closed testing procedure (Finos and Salmaso 2007; Brombin 2009; Brombin and Salmaso 2009).

- The first phase (*coordinate level*) of a procedure estimates the distribution of \mathbf{T} including the following steps:

- 1a. Calculate the vector of observed values of tests $\mathbf{T} : \mathbf{T}_o = \mathbf{T}(\mathbf{X})$.
- 1b. Consider a member g^* , randomly drawn from the set \mathbf{G} of all possible permutations, and the values of vector statistics $\mathbf{T}^* = \mathbf{T}(\mathbf{X}^*)$, where $\mathbf{X}^* = g^*(\mathbf{X})$. In most situations, the data permutation \mathbf{X}^* may be obtained at first by considering a random permutation (a_1^*, \dots, a_n^*) of integers $(1, \dots, n)$ and then by assignment of related individual data vectors to the proper group; thus, according to the unit-by-unit representation, $\mathbf{X}^* = \{\mathbf{X}(a_i^*), i = 1, \dots, n; n_1, n_2\}$.
- 1c. Carry out B independent repetitions of step (b). The set of Conditional Monte Carlo (CMC) sampling results $\{\mathbf{T}_r^*, r = 1, \dots, B\}$ is thus a random sampling from the permutation km -variate distribution of vector test statistics \mathbf{T} .
- 1d. The km -variate EDF $\hat{F}_B(\mathbf{z}|\mathbf{X}) = [\frac{1}{2} + \sum_r \mathbf{I}(\mathbf{T}_r^* \leq \mathbf{z})] / (B + 1)$, $\forall \mathbf{z} \in \mathcal{R}^{km}$, gives an estimate of the corresponding km -dimensional permutation distribution $F(\mathbf{z}|\mathbf{X})$ di \mathbf{T} . Moreover,

$$\hat{L}_h(z|\mathbf{X}) = \left[\frac{1}{2} + \sum_r \mathbf{I}(\mathbf{T}_{hr}^* \geq z) \right] / (B + 1), h = 1, \dots, km,$$

gives an estimate $\forall \mathbf{z} \in \mathcal{R}^1$ of the marginal permutation significance level functions $L_h(z|\mathbf{X}) = \Pr\{T_h^* \geq z|\mathbf{X}\}$; this $\hat{L}_h(T_{ho}|\mathbf{X}) = \lambda_h$. This gives an estimate of the marginal p -value related to test T_h .

At the end of this first phase, we get a p -value for each landmark coordinate, hence in total $2k$ or $3k$, depending from the dimension m , partial p -values.

If, for example, we deal with $k = 4$ landmarks in 2D, hence λ^*_1 is the permutation p -value corresponding to the x coordinate of landmark 1, λ^*_2 the permutation p -value corresponding to the y coordinate of landmark 1, λ^*_3 the permutation p -value corresponding to the x coordinate of landmark 2, λ^*_4 is the permutation p -value corresponding to the y coordinate of landmark 2, and so on (see Fig. 2.2).

- The second phase (*landmark level*) of the algorithm include the following steps:

- 2a. The km observed p -values are estimated from the data \mathbf{X} by $\lambda_h = \hat{L}_h(T_{ho}|\mathbf{X})$, where $T_{ho} = T_h(\mathbf{X})$, $h = 1, \dots, km$, represent the observed values of partial tests and \hat{L}_h is the h th marginal significance level function, the latter being jointly estimated by the CMC sampling method on data set \mathbf{X} , in accordance with step (1.d) above.
- 2b. The combined observed value of the second-order test is evaluated through the same CMC results of the first phase, and is given by the combination of sequential couples (or triplets) of landmark indexes (landmark coordinates) as illustrated in Fig. 2.2. For example the observed statistic related to the first landmark (in 2D case) is given by

$$T''_{1o} = \psi(\lambda_1, \lambda_2).$$

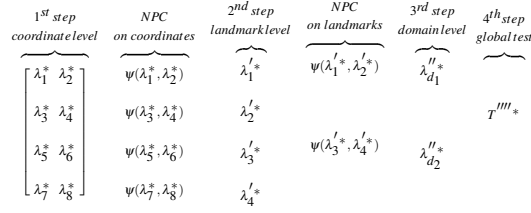


Fig. 2.2 Algorithm for $k = 4$ landmarks in 2D and two domain combinations

- 2c. The r th combined value of vector statistics (step (1.d)) for the first landmark is then calculated by

$$T_{1r}''* = \psi(\lambda_{1r}^*, \lambda_{2r}^*),$$

where $\lambda_{1r}^* = \hat{L}_1(T_{1r}^* | \mathbf{X})$, $r = 1, \dots, B$.

Steps (2.b) and (2.c) will be repeated k times, in order to obtain a partial p -value for each landmark

- The third phase (*domain level*) of the algorithm include the following steps:

- 3a. Let us assume that Z out of k landmarks, $1 \leq Z \leq k$, constitute the first domain (i.e., a subgroup of landmarks sharing anatomical, biological, or locational features); A out of k landmarks, $1 \leq A \leq k$, constitute the second domain and C out of k landmarks, $1 \leq C \leq k$, constitute the third domain. We have just defined three domains but, of course, we may define more than three domains.
- 3b. The combined observed value of the third-order test is evaluated through the same CMC results of the second phase, and is given by

$$T_{Zo}''' = \psi(\lambda_1', \dots, \lambda_Z').$$

corresponding to the first domain,

$$T_{Ao}''' = \psi(\lambda_1', \dots, \lambda_A').$$

corresponding to the second domain, and

$$T_{Co}''' = \psi(\lambda_1', \dots, \lambda_C').$$

corresponding to the third domain.

- 3c. The r th combined value of vector statistics is then calculated by

$$T_{Zr}'''* = \psi(\lambda_{1r}'^*, \dots, \lambda_{Zr}'^*),$$

where $\lambda'_{zr} = \hat{L}_z(T'''_{zr}|\mathbf{X})$, $z = 1, \dots, Z$, $r = 1, \dots, B$, is the permutation p -value corresponding to landmarks belonging to the first domain;

$$T'''_{Ar} = \psi(\lambda'_{1r}, \dots, \lambda'_{Ar}),$$

where $\lambda^*_{ar} = \hat{L}_a(T'''_{ar}|\mathbf{X})$, $a = 1, \dots, A$, $r = 1, \dots, B$, is the permutation p -value corresponding to landmarks belonging to the second domain;

$$T'''_{Cr} = \psi(\lambda'_{1r}, \dots, \lambda'_{Cr}),$$

where $\lambda^*_{cr} = \hat{L}_c(T'''_{cr}|\mathbf{X})$, $c = 1, \dots, C$, $r = 1, \dots, B$, is the permutation p -value corresponding landmarks belonging to the third domain;

Hence at the end of this step we obtain different p -values corresponding to predefined domains. Figure 2.2 illustrates an example where we have defined 2 domains, namely d_1 and d_2 , combining landmarks 1,2 and landmarks 3,4 respectively.

■ The fourth and last phase provides the global p -value.

4a. The combined observed value of the global test is evaluated through the same CMC results in the first phase, and is given by:

$$T_o''' = \psi(\lambda'_1, \lambda'_2, \lambda''_Z, \dots, \lambda''_A, \dots, \lambda''_C).$$

4b. The r th combined value of vector statistics (step (S.d_k)) is then calculated by

$$T_r''' = \psi(\lambda'_{1r}, \lambda'_{2r}, \lambda''_{Zr}, \dots, \lambda''_{Ar}, \dots, \lambda''_{Cr}).$$

4c. Hence, the p -value of the combined test T''' is estimated as

$$\lambda'''_{\psi} = \sum_r \mathbf{I}(T_r''' \geq T_o''')/B.$$

4d. If $\lambda'''_{\psi} \leq \alpha$, the global null hypothesis H_0 is rejected at significance level α .

2.3.2 Including MA Procedure

As said before, this is obviously the simplest version of the combining procedure. Actually we could be interested in emphasizing a particular aspect for each coordinate. Hence, we may apply a multi-aspect (MA) procedure at landmark coordinates level. We briefly present this procedure in the univariate general case. Let us assume, without loss of generality, that observations from a response variable

X on n units are partitioned into two groups, respectively of n_1 and n_2 units, corresponding to two levels of a treatment. Let us also assume that the response variables in the two groups have unknown distributions P_1 and P_2 , both defined on the same probability space $(\mathcal{X}, \mathcal{B})$, where \mathcal{X} is the sample space and \mathcal{B} is an algebra of events. Let $\mathbf{X}_j = \{X_{ji}, i = 1, \dots, n_j\}$ be the data set of n_j elements related to the j -th sample or group, $j = 1, 2$. Let $\mathbf{X}_j^* = \{X_{ji}^*, i = 1, \dots, n_j, j = 1, 2\}$ indicate a permutation of the observed data set \mathbf{X} , where the subscript j emphasizes the group to which permuted elements are assigned. We are interested in testing the global null hypothesis $H_0 : \{X_1 \stackrel{d}{=} X_2\} = \{P_1 = P_2\}$ that the two groups have the same underlying distribution, against the global alternative hypothesis $H_1 : \{X_1 \stackrel{d}{<} X_2\}$ of a stochastic dominance. Thus two CDFs, F_1 and F_2 , are such that in the alternative they do not intersect each other because of the side-assumptions; we also assume, for simplicity, that the two distributions are absolutely continuous (Salmaso and Solari 2005; Brombin 2009; Brombin and Salmaso 2009). H_0 may be broken down into

$$H_0 : \left\{ \bigcap_{i=1}^K H_{0i} \right\} \quad (2.1)$$

where K is the number of considered aspects. Hence H_0 is true if all H_{0i} are jointly true. The alternative may be represented as

$$H_1 : \left\{ \bigcup_{i=1}^K H_{1i} \right\}. \quad (2.2)$$

and it implies that the inequality of two distributions entails the falsity of at least one partial null hypothesis.

In case-control designs, when treatment effects are presumed to influence not only locations but also scale coefficients or other aspects, this may be conveniently examined through several statistics, each one sensitive to differences that affect a particular aspect of the two distributions.

We are interested in the *location-aspect* (l) that summarizes the two distributions in a comparison of two location indices, and in the *distributional-aspect* (d) based on the comparison of the two empirical distribution functions. Of course, other aspects may be included in the MA procedure. In order to evaluate the location-aspect we formulate a system of hypotheses considering test statistics based on both mean and median, while to examine the distributional aspect we construct a hypothesis system based on both Kolmogorov–Smirnov’s and Anderson–Darling’s test statistic.

Thus, in the first case we wish to test

$$\begin{aligned} H_{0l} &: \{[E(X_1) = E(X_2)] \cap [Me(X_1) = Me(X_2)]\} \\ H_{1l} &: \{[E(X_1) < E(X_2)] \cup [Me(X_1) < Me(X_2)]\}. \end{aligned} \quad (2.3)$$

while in the second, the referential system of hypotheses is given by

$$\begin{aligned} H_{0d} &: \{F_1 = F_2\} \\ H_{1d} &: \{F_1 > F_2\} \end{aligned} \quad (2.4)$$

Applying the NPC methodology, we can construct the location-aspect test statistic by combining the permutation p -values λ_μ^* and λ_{Me}^* associated respectively with the two partial tests T_μ^* and T_{Me}^* (the difference between the sample median of the permuted groups) where

$$T_\mu^* = T_\mu(\mathbf{X}^*) = \sum_{i=1}^{n_2} X_{2i}^*$$

and

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

using for example the Tippett combining function

$$T_l^{*''} = \max(1 - \lambda_\mu^*, 1 - \lambda_{Me}^*). \quad (2.5)$$

We can do the same to assess the distributional-aspect by combining the permutation p -values λ_{KS}^* , λ_{AD}^* associated with the two partial tests T_{KS}^* (the permutation version of the two-sample Kolgomorov–Smirnov’s statistic for one-sided alternatives) and T_{AD}^* (the permutation version of the Anderson–Darling’s test statistic) using the Tippett combining function

$$T_d^{*''} = \max(1 - \lambda_{KS}^*, 1 - \lambda_{AD}^*). \quad (2.6)$$

Finally, the global test statistic combines the information from the two-aspect tests into one global test as follows:

$$T_{MA}^{*'''} = \psi(\lambda_l^{*''}, \lambda_d^{*''}), \quad (2.7)$$

where ψ is the selected combining function ([Salmaso and Solari 2005](#)).

Of course it is possible to take any other useful combining function into consideration, e.g., Fisher, Lancaster, Liptak, and Mahalanobis. For the selection of a combining function, see the practical guidelines set out in [Pesarin \(2001\)](#); [Pesarin and Salmaso \(2010\)](#). We have mainly used Fisher omnibus combining function, calculated as

$$T_F'' = \sum_i \log(\lambda_i).$$

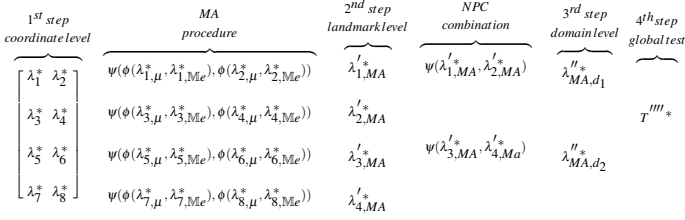


Fig. 2.3 Algorithm for $k = 4$ landmarks in 2D, MA procedure (mean and median aspects) and domain combination. ψ and ϕ are suitable combining functions

It is well known that if the k partial test statistics are independent and continuous, in the null hypothesis T_F'' follows a central χ^2 distribution with $2k$ degrees of freedom. Along with Fisher, we have used Liptak combining function based on the statistics

$$T_L'' = \sum_i \Phi^{-1}(1 - \lambda_i),$$

where Φ is the standard normal c.d.f. Of course if the k test statistics are independent and continuous, then in the null hypothesis T_L'' is normally distributed with mean 0 and variance k .

Instead of presenting the algorithm in this general case, we illustrate the procedure under the shape analysis framework. In particular, when including MA procedure, the first NPC combination presented in Fig. 2.2 is calculated by considering also the aspects (see Fig. 2.3).

One of the main feature and advantage of the proposed approach is that using the MA procedure and the information about domains we are able to obtain not only a global p -value, like in traditional tests, but also a p -value for each of the defined aspects or domains. Hence following our procedure it is possible to construct a hierarchical tree, allowing for testing at different levels of the tree (see Fig. 2.1). On one hand partial tests may provide marginal information for each specific aspect, on the other they jointly provide information on the global hypothesis. In this way, if we find a significant departure from H_0 , we can investigate the nature of this departure in detail. Also, one can move from the top to the bottom of the tree and, for interpreting results in a hierarchical way, from the bottom to the top. It is worth noting that “intermediate” level p -values need to be adjusted for multiplicity.

2.3.3 Closed Testing Procedure in Shape Analysis

Multiple comparisons and multiple testing problems arise frequently in statistical data analysis, and it is important to address them appropriately. Actually, the problem of multiplicity control arises in all cases where the number of hypotheses to be tested is greater than one. Such partial tests, possibly after adjustment for multiplicity (Westfall and Young 1993), may be useful for marginal or separate inferences. If they are jointly considered they provide information on a general overall or global hypothesis, which typically represents the true objective of the majority of multivariate testing problems. In order to produce a valid test for the combination of a large number of p -values, we must guarantee that such test is unbiased and produces, therefore, p -values below the significance level with a probability less than or equal to α itself. This combination could be very troublesome unless we are working in a permutation framework. A Bonferroni correction is valid but the conservativeness of this solution is often unacceptable for both theoretical and practical purposes. Actually, this combination loses power in case of dependence between p -values. On the contrary, using appropriate permutation methods, dependencies may be controlled. With reference to multiple testing procedures mentioned before, these have their starting point in an overall test and look for significant tests on partial contrasts. Conversely combination procedures start with a set of partial tests, each appropriate for a partial aspect, and look for joint analyses leading to global inferences. The global p -value obtained through NPC procedure of p -values associated with sub-hypotheses is an exact test, thus providing a weak control of the multiplicity. The inference in this case must be limited to the global evaluation of the phenomenon. Due to the use of NPC methods, a more detailed analysis may be carried out. Actually what is important is to select potentially active hypotheses (i.e., under the alternative). A correction of each single p -value is hence necessary in this case. A possible solution within a nonparametric permutation framework is represented by Closed testing procedures (Westfall and Wolfinger 2000). A property that is generally required is the strong control of the Familywise Error Rate (FWE), i.e., the probability of making one or more errors on the whole of the considered hypotheses (Marcus et al. 1976). On the other hand, a weak control of the FWE means simply controlling α for the global test (i.e., the test where all hypotheses are null). Although the latter is a more lenient control, it does not allow the selection of active variables because it simply produces a global p -value that does not allow interesting hypotheses to be selected, so the former is usually preferred because it makes inference on each (univariate) hypothesis (Finos and Salmaso 2005). An alternative approach to multiplicity control is given by the False Discovery Rate (FDR). This is the maximum proportion of type I errors in the set of elementary hypotheses. The FWE guarantees a more severe control than the FDR, which in fact only controls the FWE in the case of global null hypotheses, i.e., when all involved hypotheses are under H_0 (Benjamini and Hochberg 1995). In confirmatory studies, for example, it is usually better to strongly control the FWE, thus ensuring an adequate inference when you want to avoid making even

one error. On the contrary, when it is of interest to highlight a pattern of potentially involved variables, especially when dealing with thousands of variables, the FDR would appear to be a more reasonable approach. In this way it is accepted that part (no greater than the α proportion) of the rejected hypotheses are in fact under the null (Finos and Salmaso 2005).

The goal of multiple testing procedures is to control the “maximum overall Type I error rate,” i.e., the maximum probability that one or more null hypotheses is rejected incorrectly. This quantity also goes by the name “Maximum Experiment-wise Error Rate” (MEER).

With reference to the closed testing, here we give just some hints and we refer the reader to Westfall and Wolfinger (2000) and Westfall and Young (1993). Suppose we wish to test hypotheses H_1 , H_2 , H_3 , and H_4 , e.g., concerning four landmarks. Hence, with reference to the Fig. 2.1 we start applying closed testing at landmark level. The closed testing method works as follows:

1. Test each hypothesis H_1 , H_2 , H_3 , and H_4 using an appropriate α -level test.
2. Create the “closure” of the set, which is the set of all possible intersections among H_1 , H_2 , H_3 , and H_4 (in this case the hypotheses H_{12} , H_{13} , H_{14} , H_{23} , H_{24} , H_{34} , H_{123} , H_{134} , H_{234} , and H_{1234}). In Fig. 2.4 we illustrate the procedure. We have enumerated all the possible intersections, but of course we are interested only in some of these intersections. Actually some of these are useful for inferential purpose, and some other are only instrumental and are not investigated. Intersections of interest are represented by the red bounded boxes, corresponding respectively to the landmark level (i.e., H_1 , H_2 , H_3 , and H_4), to the domain level (i.e., H_{12} and H_{34}) and to the global test (H_{1234}).
3. Test each intersection using an appropriate α -level test. In general any test that is valid for the given intersection.
4. You may reject any hypothesis H_i , with control of the MEER, when the following conditions both hold:
 - The test of H_i itself yields a statistically significant result, and
 - The test of every intersection hypothesis that includes H_i is statistically significant.

Hence a statistically significant result has been obtained for the H_3 test, as well as a significant result for all hypotheses that include H_3 , in this case, H_{13} , H_{23} , H_{34} , H_{123} , H_{134} , H_{234} , and H_{1234} (blue boxes in Fig. 2.4). Since the p -value for one of the including tests, the H_{1234} test in this case, is greater than 0.05, you may not reject the H_3 test at the MEER=0.05 level. In this example, we could reject the H_3 hypothesis for MEER levels as low as, but no lower than 0.0618, since this is the largest p -value among all hypotheses containing H_3 . This suggests an informative way of reporting the results of a closed testing procedure. When using a closed testing procedure, the adjusted p -value for a given hypothesis H_i is the maximum of all p -values for tests that include H_i as a special case (including the p -value for the H_i test itself). The adjusted p -value for testing H_3 is, therefore, formally computed as $\max(0.0067, 0.0220, 0.0285, 0.0285, 0.0570, 0.0580, 0.0600, 0.0618) = 0.0618$.

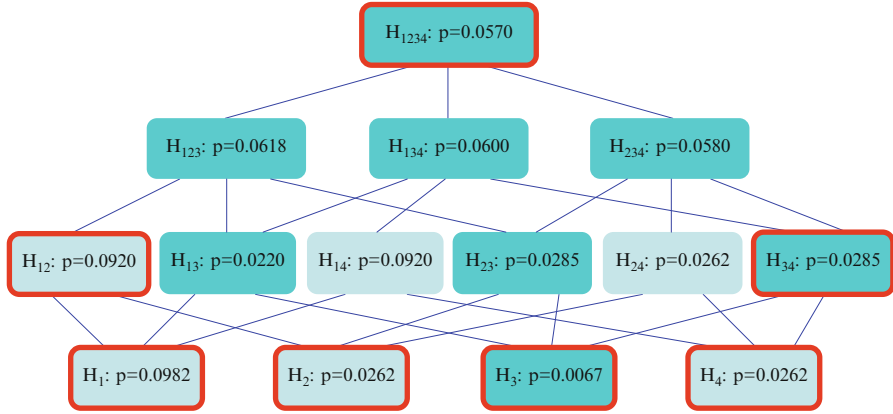


Fig. 2.4 Illustration of the closed testing procedure, focussing on landmark level ($k = 4$)

2.4 Permutation Version of Hotelling's T^2

In order to compare different tests known in the shape analysis literature, we have at first carried out a simulation study on Hotelling's T^2 power which is a sort of milestone for multivariate testing also in shape analysis. Let us consider the two independent sample case and assume that the response variables behave according to the following model:

$$X_{hji} = \mu_h + \delta_{hj} + Z_{hji},$$

$i = 1, \dots, n_j$, $j = 1, 2$, $h = 1, \dots, k$, where n_j is the sample size, μ_h represents a population constant for the h -th variable; δ_{hj} represents the fixed treatment effect (i.e., the noncentrality parameter) in the j -th group on the h -th variable and Z_{hji} are k -dimensional random errors exchangeable with respect to treatment levels with null mean vector ($\mathbb{E}(\mathbf{Z}) = 0$) and finite second moment.

Let \bar{X}_{hj} , $j = 1, 2$, be the sample mean for the h -th variable, S_j the biased sample covariance matrix (with divisors n_1 and n_2) and S the common covariance matrix, given by $S = (n_1 S_1 + n_2 S_2) / (n_1 + n_2 - 2)$.

We define T'' the nonparametric permutation counterpart of Hotelling T^2 given by

$$T'' = \sum_{h=1}^k \left(\frac{\bar{X}_{h1}^* - \bar{X}_{h2}^*}{s_h^*} \right)^2$$

where the symbol $*$ indicates a permutation of the original data, \bar{X}_{hj}^* , $j = 1, 2$ are multivariate permutation sample means, and s_h^* are the diagonal elements of S^* . We remark that the underlying dependence structure is nonparametrically and implicitly “captured” by the permutation procedure (see e.g. [Pesarin 2001](#) and [Pesarin and](#)

Table 2.1 Simulations under H_1 ($n_1 = n_2 = 10$, $\mu = 0$, $\delta = 0.40$, $B = CMC = 1000$)

		$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.20$	$\alpha=0.30$	$\alpha=0.50$
$k = 15$	T^2	0.027	0.118	0.233	0.419	0.566	0.789
	T''	0.231	0.484	0.623	0.771	0.856	0.941
$k = 16$	T^2	0.026	0.098	0.192	0.361	0.504	0.741
	T''	0.228	0.496	0.633	0.792	0.866	0.946
$k = 17$	T^2	0.019	0.081	0.158	0.325	0.455	0.703
	T''	0.258	0.534	0.681	0.811	0.875	0.950
$k = 18$	T^2	0.013	0.067	0.132	0.269	0.414	0.642
	T''	0.253	0.543	0.667	0.816	0.874	0.956
$k = 19$	T''	0.244	0.544	0.700	0.837	0.905	0.977
$k = 20$	T''	0.318	0.552	0.683	0.825	0.904	0.965
$k = 21$	T''	0.307	0.570	0.693	0.832	0.901	0.962
$k = 22$	T''	0.340	0.618	0.744	0.845	0.906	0.964
$k = 23$	T''	0.344	0.629	0.750	0.857	0.918	0.974
$k = 24$	T''	0.338	0.622	0.741	0.862	0.919	0.973
$k = 25$	T''	0.365	0.656	0.774	0.880	0.930	0.970

Table 2.2 Simulations under H_1 ($n_1 = n_2 = 3$, $\mu = 0$, $\delta = 0.40$, $B = MC = 1000$)

		$\alpha=0.10$	$\alpha=0.20$	$\alpha=0.30$	$\alpha=0.50$
$k = 3$	T''	0.059	0.194	0.337	0.554
$k = 18$	T''	0.097	0.278	0.408	0.618
$k = 20$	T''	0.090	0.264	0.390	0.611
$k = 25$	T''	0.117	0.274	0.422	0.643
$k = 30$	T''	0.100	0.270	0.404	0.647
$k = 35$	T''	0.103	0.280	0.436	0.687
$k = 40$	T''	0.089	0.277	0.442	0.667

Salmaso 2010). We also emphasize that in a shape analysis framework X_{hji} will indicate the 2D or 3D landmark coordinates.

When carrying out nonparametric permutation tests we use raw coordinates and not the shape coordinates. Hence we do not use the coordinates obtained after filtering out location, scale, and rotational effects from the original data. However we deal exhaustively with this topic in Chap. 3.

We have compared the traditional parametric Hotelling's T^2 test (T^2) with the nonparametric T^2 -type counterpart (T'') showing that the power for the suggested test increases when increasing the number of the processed variables (see Table 2.1) with the same noncentrality parameter δ , even when the number of covariates (k) is larger than the permutation sample space (see results in Tables 2.2 and 2.3).

We remark that Hotelling's T^2 test considered in Table 2.1 is computed using raw coordinates and not shape variables. Moreover, when $n_1 = n_2 = 10$ and $k = 19$, the

Table 2.3 Simulations under H_1 ($n_1 = n_2 = 3, \mu = 0, \delta = 1, B = MC = 1000$)

		$\alpha=0.10$	$\alpha=0.20$	$\alpha=0.30$	$\alpha=0.50$
$k = 3$	T''	0.187	0.454	0.629	0.800
$k = 20$	T''	0.324	0.779	0.902	0.977
$k = 50$	T''	0.434	0.907	0.985	0.999

test statistic is constantly equal to 0. Hotelling’s T^2 statistic can be related to the F -distribution by the well-known relation

$$T^2 = \frac{n_1 n_2 (n_1 + n_2 - k - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)k} D^2 \sim F_{k, n_1 + n_2 - 1 - k},$$

where D^2 is the Mahalanobis squared distance.
 B is the number of permutations (Monte Carlo sampling) used for estimating the permutation distribution, and CMC is the number of Monte Carlo iterations of the simulation procedure. Note that for $n_1 = n_2 = 3$ we explored the whole permutation sample space.
These interesting findings allow us to assess the usefulness of the nonparametric permutation solution for high-dimensional data in small sample size case. Moreover these preliminary results enable us to evaluate the power of multivariate NPC tests discussed in Chap. 4, thus introducing and extending the notion of “finite-sample consistency (FSC),” widely discussed in Chap. 5.
A comprehensive comparison, based on simulation study, of power behavior of the nonparametric permutation version of Hotelling’s T^2 against that of traditional tests used in shape analysis literature, has been published elsewhere in [Brombin and Salmaso \(2009\)](#). Results highlight the good performances in terms of power behavior of NPC test, while controlling the Type I error.

References

Alfieri R, Bonnini S, Brombin C, Castoro C, Salmaso L (2012) Iterated combination-based paired permutation tests to determine shape effects of chemotherapy in patients with esophageal cancer. Statistical Methods in Medical Research DOI: 101177/0962280212461981 Article published online before print

Amaral G, Dryden I, Wood A (2007) Pivotal bootstrap methods for k -sample problems in directional statistics and shape analysis. Journal of the American Statistical Association 102:695–707

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A new and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57:1289–1300

Blair RC, Higgins JJ, Karniski W, Kromrey JD (1994) A study of multivariate permutation tests which may replace Hotelling’s t^2 test in prescribed circumstances. Multivariate Behavioral Research 29:141–163

- Bookstein FL (1997) Shape and the information in medical images: A decade of the morphometric synthesis. *Computer Vision and Image Understanding* 66:97–118
- Brombin C (2009) A nonparametric permutation approach to statistical shape analysis, ph.D. thesis. Padova, Italy: University of Padova
- Brombin C, Salmaso L (2009) Multi-aspect permutation tests in shape analysis with small sample size. *Computational Statistics & Data Analysis* 53:3921–3931
- Brombin C, Pesarin F, Salmaso L (2008) Dealing with more variables than sample sizes: an application to shape analysis. In: Hunter DR, Richards DSP, Rosenberger JL (eds) *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P. Hettmansperger*, Singapore: World Scientific, pp 28–44
- Brombin C, Mo G, Zotti A, Giuriso M, Salmaso L, Cozzi B (2009a) A landmark analysis-based approach to age and sex classification of the skull of the mediterranean monk seal (*monachus monachus*) (hermann, 1779). *Anatomia, Histologia, Embryologia* 38:382–386
- Brombin C, Salmaso L, Villanova C (2009b) Multivariate permutation shape analysis with application to aortic valve morphology. In: et al. VC (ed) *Stereology and Image Analysis. Ecs10: Proceeding of the 10th European Conference of ISS, The MIRIAM Project Series*, ESCULAPIO Pub. Co., Bologna, Italy, 2009., vol 4, pp 442–449
- Dryden IL, Mardia KV (1993) Multivariate Shape Analysis. *Sankhyā Series A* 55:460–480
- Dryden IL, Mardia KV (1998) *Statistical Shape Analysis*. John Wiley & Sons, London
- Finos L, Salmaso L (2005) A new nonparametric approach for multiplicity control: *Optimal Subset* procedures. *Computational Statistics* 20:643–654
- Finos L, Salmaso L (2007) FDR- and FWE-controlling methods using data-driven weights. *Journal of Statistical Planning and Inference* 137:3859–3870
- Good P (2000) *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer, New York
- Goodall CR (1991) Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, Series B* 53:285–339
- Lele S (1993) Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Mathematical Geology* 25:573–602
- Lele S, Cole III TM (1995) Euclidean distance matrix analysis: a statistical review. In *Current Issues in Statistical Shape Analysis*, University of Leeds, Leeds 3:49–53
- Lele S, Cole TM (1996) A new test for shape differences when variance-covariance matrices are unequal. *The Journal of Human Evolution* 31:193–212
- Lele S, Richtsmeier JT (1991) Euclidean distance matrix analysis: a coordinate free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology* 86:415–427
- Marcus R, Peritz E, Gabriel KR (1976) On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63:655–660
- Pesarin F (2001) *Multivariate Permutation tests: with application in Biostatistics*. John Wiley & Sons, Chichester-New York
- Pesarin F, Salmaso L (2010) *Permutation Tests for Complex Data: Theory, Applications and Software*. Wiley
- Rohlf FJ (2000) Statistical power comparisons among alternative morphometric methods. *American Journal of Physical Anthropology* 111:463–478
- Salmaso L, Solari A (2005) Multiple Aspect Testing for case-control designs. *Metrika* 12:1–10
- Seber G (1984) *Multivariate Observations*. John Wiley, New York
- Terriberry TB, Joshi S, Garig G (2005) Hypothesis testing with nonlinear shape models. In: *Information processing in medical imaging*, vol 3565, Springer Berlin-Heidelberg, pp 15–26
- Westfall PH, Wolfinger RD (2000) *Closed Multiple Testing Procedures and PROC MULTTEST*. Observations, SAS Institute Inc.
- Westfall PH, Young SS (1993) *Resampling-Based Multiple Testing*. Wiley, New York

Permutation Tests in Shape Analysis

Salmaso, L.; Brombin, C.

2013, X, 95 p. 19 illus., 16 illus. in color., Softcover

ISBN: 978-1-4614-8162-1