

SPRINGER BRIEFS IN STATISTICS

Chiara Brombin
Luigi Salmaso

Permutation Tests in Shape Analysis

 Springer

SpringerBriefs in Statistics

For further volumes:

<http://www.springer.com/series/8921>

Chiara Brombin • Luigi Salmaso

Permutation Tests in Shape Analysis

 Springer

Chiara Brombin
CUSSB, Vita-Salute
University of San Raffaele
Milan, Italy

Luigi Salmaso
Department of Management
and Engineering
University of Padova
Vicenza, Italy

ISSN 2191-544X

ISBN 978-1-4614-8162-1

DOI 10.1007/978-1-4614-8163-8

Springer New York Heidelberg Dordrecht London

ISSN 2191-5458 (electronic)

ISBN 978-1-4614-8163-8 (eBook)

Library of Congress Control Number: 2013943529

© Springer Science+Business Media New York 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Fortunato Pesarin

WHO SAID CAN'T?

*Someone is always doing something someone
else said was impossible.*

TRY TRYING.

Unknown

Preface

The statistical community has shown an increased interest in shape analysis in the last decade, in particular with reference to the development of robust inferential statistical methods. In this book we present an extension of NonParametric Combination (NPC) methodology (Pesarin, 2001; Pesarin and Salmaso, 2010) to shape analysis. At first we introduce basic concepts and terms that will be used throughout the book. In particular we provide a brief overview of statistical shape analysis and geometric morphometric techniques, focussing on landmark and semilandmark-based representations of shapes (Chap. 1). Then we face with inferential aspects in the field of shape analysis. In particular, we review inferential methods known in the shape analysis literature, highlighting some drawbacks of using Hotelling's T^2 test statistic, and we introduce NPC methodology for the analysis of shape configurations. Multiple Aspect (MA) procedures and domain combinations are also illustrated (Chap. 2). The case of heterogeneous variation and nonzero correlation among landmarks is also investigated, along with the effects of superimposition on the power of NPC tests (Chap. 3). Permutation tests have been evaluated also in the particular case in which the number of variables is larger than the cardinality of permutation sample space. We have performed a simulation study to evaluate the power of multivariate NPC tests, showing that the power for the proposed tests increases when increasing the number of the processed variables provided that the noncentrality parameter increases, even when the number of covariates is larger than the permutation sample space (Chap. 4).

These preliminary results allowed us to extend the notion of *finite-sample consistency* for permutation tests combination-based to the shape analysis field. Sufficient conditions are given in order that the rejection rate converges to one, for fixed sample sizes at any attainable α -value, when the number of variables diverges, provided that the noncentrality induced by test statistics also diverges (Chap. 5).

The last chapter is mainly devoted to practical applications. In particular we present an application concerning the facial expression of emotion along with a case study aimed at analyzing aortic valve morphology. Moreover we also introduce two innovative topics: biometric morphing and nonparametric iterated combination for paired data (Chap. 6).

We wish to thank our colleagues Bruno Cozzi, PierFrancesco Galzignato, and Carla Villanova for kindly suggesting us interesting case studies and providing us with the datasets used in this book.

A great thanks to our inspiring Master, Advisor, and colleague Fortunato Pesarin for his passion and encouragement for our research. A final thanks to James Rohlf for introducing us in the fascinating field of shape analysis.

Milano, Italy
Vicenza, Italy

Chiara Bombin
Luigi Salmaso

References

- 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

Contents

1	A Brief Overview on Statistical Shape Analysis	1
1.1	Some Historical Notes	2
1.2	How to Describe Shapes	3
1.2.1	Monk Seal Skulls Study	6
1.2.2	Example of Principal Facial Landmarks	8
1.3	Multivariate Morphometrics	10
	References	14
2	Theoretical Aspects on Permutation Tests and Shape Analysis	17
2.1	Inference and Shape Analysis	18
2.2	Technical Details on Tests Known in Shape Analysis Literature	20
2.2.1	Hotelling's T^2 and Goodall's F Tests	20
2.2.2	Euclidean Distance Matrix Analysis Methods	21
2.3	NPC Approach to Shape Analysis	22
2.3.1	A Suitable Algorithm	24
2.3.2	Including MA Procedure	27
2.3.3	Closed Testing Procedure in Shape Analysis	31
2.4	Permutation Version of Hotelling's T^2	33
	References	35
3	Evaluating Power Behavior of Nonparametric Combination Testing Methodology After Generalized Procrustes Analysis and Under Different Correlation Structures	37
3.1	Generalized Procrustes Analysis and Power of NonParametric Combination methodology	37
3.2	Introducing Correlation Between Landmarks	39
3.3	Paired Data Problems: Study of Symmetric Structures	47
	References	51

4 Power Behavior of Permutation Tests with High Dimensional Data	53
4.1 High Dimensional Data with Small Sample Sizes.....	53
4.2 Simulation Study and Results	54
4.3 Final Remarks	57
References	59
5 Finite-Sample Consistency of Combination-Based Tests in Shape Analysis	61
5.1 How to Obtain a Tangent Space: A Brief Overview.....	61
5.2 Main Theorems and General Characterization of Finite Sample Consistency.....	63
5.3 A Toy Example.....	71
References	74
6 Applications to Real Case Studies	77
6.1 Biometric Morphing and Facial Expression of Emotions	77
6.2 Some Remarks on Iterated Combination.....	82
6.3 On Morphology of Aortic Valve	88
6.3.1 Introduction	88
6.3.2 Inferential Results	92
6.4 Final Remarks	93
References	94

Chapter 1

A Brief Overview on Statistical Shape Analysis

“We have the duty of formulating, of summarizing, and of communicating our conclusions, in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions”.

Statistical methods and scientific induction. *Journal of the Royal Statistical Society*, B, **17**, 69–78, 1955.

R.A. Fisher

Abstract In this chapter we introduce the basic concepts and terms that will be used throughout the book. In particular we provide a brief overview of statistical shape analysis and geometric morphometric techniques, focussing on landmark and semilandmark-based representations of shapes.

Shape is described as the geometric property of an object invariant under rotation, scale, or translation. Morphometric is a new promising branch of statistics that integrates knowledge from mathematics, geometry, biometrics, computer science, and modern engineering (essential especially for complicated three-dimensional object) to study shape and size of objects, along with their covariations with other variables. In this context the shape of an object is considered as a whole, align with the interdependence of its parts and conclusions are drawn under conditions of uncertainty. Geometric morphometrics is a more recent area of morphometrics that, by means of statistical tools, analyzes geometric information of objects, focusing on exactly where points or parts of the organism are located with respect to each other. To illustrate how landmarks and semilandmarks are chosen and then classified in real applications, we propose two case studies.

Keywords Landmark-based analysis • Geometric morphometrics • Multivariate morphometrics • Procrustes analysis • Semilandmarks • Statistical shape analysis

1.1 Some Historical Notes

Statistical shape analysis may be considered a cross-disciplinary field characterized by flexible theory and techniques: specific applications may be found in archeology, architecture, biology, geography, geology, agriculture, genetics, medical imaging, neuroanatomical research, security applications such as face recognition, entertainment industry (movies, games), computer-aided design, and manufacturing, and so on. The range of applications of shape analysis is enormous.

Statistical shape analysis offers a mathematical quantification of shape. It relates to the study of random objects, where the concept of shape corresponds to some geometrical information that is invariant under translation, rotation, and scale effects (Small 1996). In everyday language, *shape* refers to the external form or appearance characteristic of an object or the outline of an area/figure.

An intuitive definition of shape is given by Kendall (1977).

Definition *Shape is all the geometrical information that remains when location, scale, and rotational effects are filtered out from an object.*

Hence two objects have the same shape if they are invariant under the Euclidean similarity transformations of translation, scaling, and rotation (Dryden and Mardia 1998). The pioneers in the field of shape analysis are Kendall (1977, 1984, 1989) and Bookstein (1978, 1986, 1989, 1991). In 1977, Kendall published a brief note in which he introduced a new representation of shapes as elements of complex projective spaces. But at that time, applied statisticians found difficult to appreciate the practical utility of his result. Full details of Kendall's theory of shape were finally given in 1984. Showing also potential areas of research and applications, this work fascinated both probabilists and statisticians.

Mardia et al. (1977) investigated the distribution of the shapes of triangles generated by certain processes, and in particular considered whether towns in a plain are spread regularly with equal distances between neighboring towns. Dryden and Mardia's interests in statistical shape analysis began in 1986, when they start collaborating with Paul O'Higgins and David Johnson (Department of Anatomy at the University of Leeds, UK) for the analysis of the shape of some mouse vertebrae (Dryden and Mardia 1998).

In 1986, Kendall was invited to be a discussant for an inspiring article by Bookstein in the journal of *Statistical Science*, then published in Volume 1 (Bookstein 1986). At that time, it was natural to perceive the close connection between the mathematical work in shape theory from Kendall landmark paper (Kendall 1984) and the practical application proposed in Bookstein's paper.

Kendall and Bookstein developed their theories independently, but they came to the same conclusion about the possibility of representing spaces on manifolds independently. Even if they had the same intuition, their proposals were innovative and unique in both theoretical aspects and practical applications. On one hand, Kendall represented the shapes of triangles in the plain as points on a sphere, i.e., a shape of positive curvature, he focussed on the differential geometry of shape

analysis and his main applications were in archeology and astronomy. On the other hand, Bookstein suggested to represent the shapes of triangles as points on a Poincarè half plane, i.e., a space of negative curvature; his main interests were in biological and medical sciences and drew on the tradition of researchers such as D'Arcy Thompson (1961), Julian Huxley (1932), and later researchers in allometry and multivariate morphometrics (Small 1996).

Moreover, while Kendall studied the shapes of random sets of points (as generated by a Poisson process), Bookstein focused on points called landmark, i.e., biologically active sites on organisms.

Some references and reviews include Goodall (1991), Le and Kendall (1993), Kent (1994), Kent (1995), Dryden and Mardia (1993), Small (1988), Stoyan et al. (1995), Stoyan and Stoyan (1994), and Mardia (1995). Recent books on the topic are Small (1996), Mardia and Dryden (1989), Bookstein (1991), Lele and Richtsmeier (2001), Slice (2005), and Weber and Bookstein (2011).

1.2 How to Describe Shapes

A substantial role in shape analysis research has been played by shape analysis based on landmark data, where shapes are represented by a discrete sampling of the object contours (Dryden and Mardia 1998; Small 1996). Hence landmarks provide one way of sampling form, they are distinguishable equivalent points on the objects under study, i.e., they are positioned in some way equivalent on all specimens.

Bookstein and his colleagues recommended the use of landmarks for the analysis of biological features and constrains the choice of landmarks to prominent features of the organism or biological structure (Dryden and Mardia 1998). Hence these points were biologically active sites on organisms and defined in Dryden and Mardia (1998) as follows:

Definition *A landmark is a point of correspondence on each object that matches between and within population.*

These loci have the same name, i.e., they are homologues, as well as Cartesian coordinates, and correspond in some sensible way over the forms of a data set. We recall that in geometric morphometrics the term homologous has no meaning other than the same name is used for corresponding parts in different species or developmental stages (Slice et al. 1996). Moreover these points represent a foundation for the explanations of the biological processes, and still nowadays many of the explanations of form accepted as epigenetically valid adduce deformations of the locations of landmarks (Bookstein 1986).

Srivastava et al. (2005) emphasized some limitations of the landmark-based representations. Despite the effectiveness of this approach in the applications where landmarks are readily available (e.g., physician-assisted medical image analysis), automatic detection of landmarks is not straightforward and the resulting shape analysis is extremely determined by the choice of landmarks. In addition, shape interpolation with geodesics in this framework lacks a physical interpretation.

Landmarks could be basically classified into three groups: anatomical, mathematical, and pseudo-landmarks.

- An *anatomical landmark* is a point assigned by an expert that corresponds between organisms in some biologically meaningful way, e.g., the corner of an eye or the meeting of two sutures on a skull.
- *Mathematical landmarks* are points located on an object according to some mathematical or geometrical property of the figure, e.g., at a point of high curvature or at an extreme point. Mathematical landmarks are particularly useful in automatic recognition and analysis.
- *Pseudo-landmarks* are constructed points on an organism, located either around the outline or in between anatomical or mathematical landmarks. Continuous curves can be approximated by a large number of pseudo-landmarks along the curve. Also, pseudo-landmarks are useful in matching surfaces, when points can be located on a regular grid over each surface.

Furthermore they could be grouped into three further types, as described in [Dryden and Mardia \(1998\)](#).

- *Type I landmarks* (usually the easiest and the most reliable to locate) are mathematical points whose homology from case to case is supported by the strongest (local) evidence, such as meeting of structures or tissues or a small patch of some unusual histology. Type III landmarks have at least one deficient coordinate (which means that they can be reliably located to an outline or surface but not at a specific location, e.g., tip of a rounded bump).
- *Type II landmarks* are defined by local properties such as maximal curvatures, i.e., they are mathematical point whose homology is strengthened only by geometric, not histological, evidence: for instance, the sharpest curvature of a tooth. Actually this type of landmarks include landmarks which are not homologous in a developmental or evolutionary sense but which are equivalent functionally such as wing tips.
- *Type III landmarks* are the most difficult and the least reliable to locate. They have at least one deficient coordinate (i.e., they can be reliably located to an outline or surface but not at a specific location, e.g., tip of a rounded bump) and they occur at extremal points or constructed landmarks (e.g., maximal diameters and centroids). They characterize more than one region of the form and they could be treated by geometric morphometrics as landmark points, even if they could be tricky because of the deficiency they embody.

Anatomical landmarks are usually of type I or II and mathematical landmarks are usually of type II or III. Pseudo-landmarks are commonly taken as equi-spaced along outlines between pairs of landmarks of type I or II, and in this case the pseudo-landmarks are type III landmarks.

Landmark coordinates, or measurements derived from them, support and are fundamental in many studies since they record equivalences or homologies. Deciding which and how many landmarks to analyze is crucial: several aspects should be taken into account. First of all, by definition, landmarks must be homologous

loci identifiable on each specimen in the study. Then, configurations of landmark should provide an adequate summary of morphology since, within the geometric framework, landmarks represent the unique data in any analysis. When research interest is the study of curves or perimeters of structures on an organism, then semilandmarks points should be collected. Actually these points are located on a curve and allowed to slip a small distance with respect to another corresponding curve. The term “semi” is used because the landmarks lie in a lower number of dimensions than other types of landmarks, e.g., along a one-dimensional curve in a two-dimensional image (Dryden and Mardia 1998). Semilandmarks are defined in relation to other landmarks, for example “midway between landmarks 1 and 2.” Indeed they have no anatomical identifiers but remain corresponding points in a sense satisfactory for subsequent morphometric interpretation (Bookstein 1997). Hence these loci fail to be true landmarks in the fact that they do not enjoy homology property, as previously defined, since they lie on homologous curves while their exact position along these usually smooth regions or curves is unclear.

Defining semilandmarks could be useful to study substantial regions in a object that cannot be defined simply using anatomical or mathematical landmarks, or a region comprises between two or more real landmark points (Adams et al. 2004).

On the basis of these considerations, Katina et al. (2007) proposed another landmark classification, including the information carried by semilandmarks on curves and surfaces. In particular it is possible to define the following landmark types:

Type 1: discrete juxtaposition of tissues;

Type 2: extreme of curvature characterizing a single structure;

Type 3: landmark points characterized locally by information from multiple curves and surfaces and by symmetry:

- *Type 3a:* intersection of a ridge curve and the midcurve on the same surface;
- *Type 3b:* intersection of an observed curve and the midcurve;
- *Type 3c:* intersection of a ridge curve and an observed curve on the same surface;

Type 4: semilandmarks on ridge curves and symmetric curve (midsagittal curve);

Type 5: semilandmarks on surfaces;

Type 6: constructed semilandmarks.

To provide a reliable analysis, landmarks points should be consistently located and reproduced with a high degree of accuracy. Moreover, they should have conserved topological positions relative to other landmarks.

Two-dimensional landmark coordinates can be easily obtained from a digital image of a specimen using free image processing softwares such as tpsDig (Rohlf 2007), ImageJ, or Scion Image (Rasband 1997–2012; Schneider et al. 2012). Of course, image quality and the quality of the final data are strongly related. Reliable landmark data are unlikely to be obtained from a poor-contrast, out-of-focus photograph of a specimen. The quality of the specimen and/or photograph should be enhanced (e.g., cleaning, whitening, careful mounting, and illumination) prior to digitizing the landmarks.

However, progresses in landmark acquisition have been made. Modern imaging techniques such as laser scanning, MRI, and the introduction of three-dimensional (3D) reconstruction in CT technology have revolutionized the way landmark coordinates are obtained from computer-generated images using software probes (Spoor et al. 2000). Advances in software and hardware are also making possible lightweight stereophotogrammetric devices using digital cameras directly linked to portable computers (O’Higgins 2000). In order to illustrate how landmarks and semilandmarks are chosen and then classified in real applications, we briefly introduce our case studies on monk seal skulls and on emotion recognition through facial expression (see Chap. 6 for more information and results).

1.2.1 Monk Seal Skulls Study

Data at hand consist of 17 Mediterranean monk seal (*Monachus monachus*) skulls and information about sex and age class category have been collected by fellows of the Department of Experimental Veterinary Sciences of the University of Padova (Mo 2005). In particular 4 seals are male and 5 are female, while for 8 of them we do not have information about sex. Left-lateral, frontal, posterior, dorsal, and ventral views of the skull are also available for each subject.

Further analysis of the full data set may be found in Brombin et al. (2009), Brombin (2009), Brombin et al. (2008), Brombin and Salmaso (2009), Pesarin (2001), and Pesarin and Salmaso (2010).

Figures 1.1 and 1.2 typify how landmarks data lie upon images. Here we wish to show the design of the experiment performed to describe the shape of the monk seal skulls. A description of the landmarks used in the design is given below.

Type 1:

- nasospinale (denoted by *ns*), a point where the midsagittal plane meets the inferior inner rim of the nasal aperture;
- rhinion (denoted by *rhi*), midline point at the inferior free end of the internasal suture;
- nasion (denoted by *n*), midline point where the two nasal bones and the frontal intersect;
- maxillonasofrontale (nasomaxilla, denoted by *mnf*), a point on the crossing of frontonasal, frontomaxillare, nasomaxillare sutures;

Type 2:

- jugale (denoted by *ju*), point in the depth of the notch between the temporal and frontal process of the zygomatic;
- mastoideale (denoted by *ms*), most inferior point on the mastoid process;

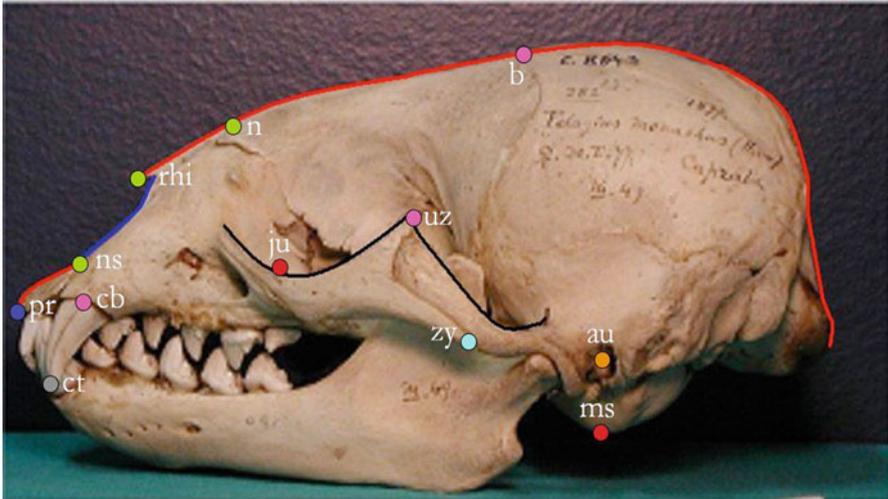


Fig. 1.1 Left-lateral view (Legend: Type 1 (green dot), Type 2 (red dot), Type 3 (blue dot), Type 4 (magenta dot), Type 5 (cyan dot), Type 6 (orange dot), unknown Type (light-gray dot), nasal curve (blue curve), midsagittal curve (red curve), zygomatic curve (black curve))

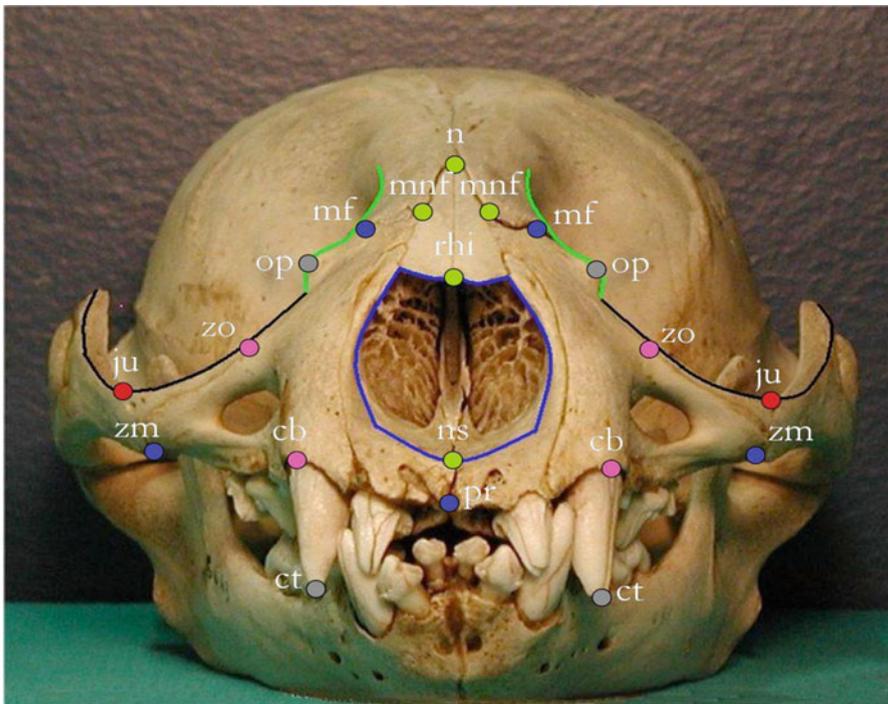


Fig. 1.2 Frontal view (Legend: Type 1 (green dot), Type 2 (red dot), Type 3 (blue dot) Type 4 (magenta dot), unknown Type (light-gray dot), nasal curve (blue curve), zygomatic curve (black curve), part of orbital curve (green curve))

Type 3:

- prosthion (denoted by pr), point on the maxillary bone where the midsagittal plane meets a tangent that goes through the alveolar margins of the central incisors;
- maxillofrontale (denoted by mf), point where the anterior lacrimal crest of the maxilla meets the frontomaxillary suture;

Type 4:

- bregma (denoted by b), the juncture of the coronal and sagittal sutures in the median sagittal plane; should an ossicle be present, the landmark can be located by drawing in pencil a continuation of the sutures until these lines intersect;
- canine base (denoted by cb), most mesial point on the outer alveolar margin of the canine;
- superior zygomaticum (upper zygomatic, denoted by uz), most superior point on the suture that separates zygomatic and parietal bone;
- zygomaxillare (denoted by zm), most inferior point on the zygomaticomaxillary suture;
- zygoorbitale (denoted by zo); point where the orbital rim intersects the zygomaticomaxillary suture;

Type 5:

- zygion (denoted by zy), most inferior point on the suture that separates zygomatic and parietal bone;

Type 6:

- auriculare (denoted by au), point vertically above the center of the external auditory meatus at the root of the zygomatic process.

In the left-lateral view we have chosen 5 midplane landmarks and 7 bilateral landmarks (see Table 1.1 and Fig. 1.1).

In the frontal view we have chosen 4 midplane landmarks and 8 bilateral landmarks. Landmarks marked by gray bullets have not been classified, in particular canine tip (ct) landmark point could be defined as Type 5, but it is difficult to classify univocally. In Fig. 1.1 we show the landmarks and the curves we have chosen. The gray bullet refers to an undefined landmark type.

Similar arguments could be applied to orbital process (op) landmark point, where op can be defined as orbital spine on the orbital process of the maxillary bone (Fig. 1.2).

1.2.2 Example of Principal Facial Landmarks

Standard anthropometric anatomical landmarks, i.e., landmarks on the skull from which craniometric measurements can be taken, are shown in Fig. 1.3a, b. Most of

these landmarks are frequently used in plastic and reconstructive surgery. Pictures reproduced below derived from a rhinoseptoplasty case study, in which 14 patients were evaluated before and after surgery. Results of this case study are published elsewhere in [Brombin et al. \(2011\)](#).

Table 1.1 Anatomical landmarks

<i>Midsagittal</i>
Prosthion, pr
Nasospinale, ns
Rhinion, rhi
Nasion, n
Bregma, b
<i>Bilateral</i>
Canine base, cb
Canine tip, ct
Maxillonasofrontale (Nasomaxilla), mnf
Maxillofrontale, mf
Orbital process, op
Zygoorbitale, zo
Zygomaxillare, zm
Jugale, ju
Superior zygomatic (Upper zygomatic), uz
Zygion, zy
Auriculare, au
Mastoideale, ms

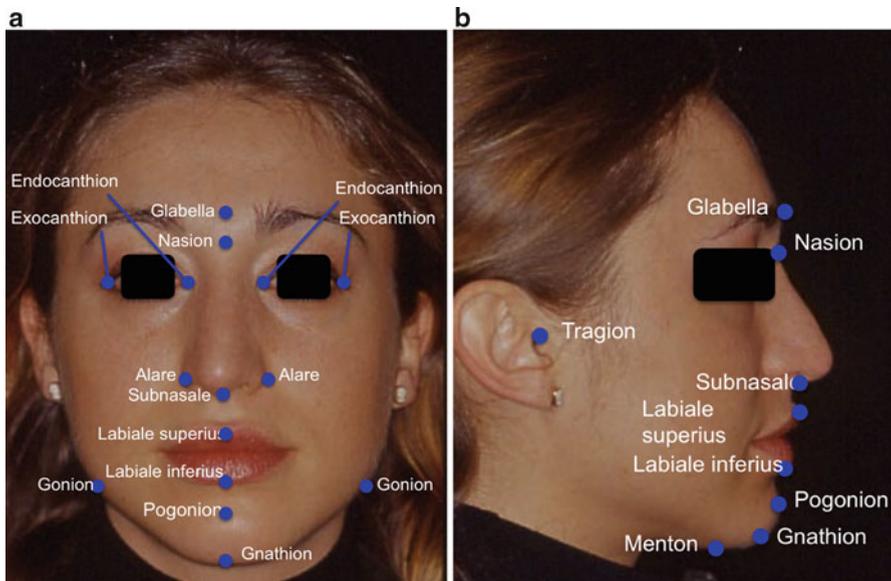


Fig. 1.3 Principal facial landmarks: frontal (a) and right-lateral views (b)

All the analyzed patients gave their informed consent to participate in the experiment (included the photographic documentation only for scientific purposes). Nowadays, anthropometric analysis and surgical practice intersected at the point to treat congenital or posttraumatic facial disfigurements in various racial or ethnic groups successfully (Farkas 1994; Farkas et al. 2005).

While anthropometric analysis of nose allows to provide data which should contribute to satisfactory results of the cosmetic nasal surgery (Etöz 2011), landmark-based geometric morphometric techniques represent a powerful tool to analyze nasal shapes in any population.

Midsagittal landmarks	
Name of landmark	Description of landmark
Glabella	The most prominent part in the midline between the brows
Nasion	The midpoint of the nasofrontal suture
Subnasale	The junction between the lower border of the nasal septum, the partition that divides the nostrils, and the cutaneous portion of the upper lip in the midline
Labiale superius	The midpoint of the vermilion border of the upper lip
Labiale inferius	The midpoint of the vermilion border of the lower lip
Gnathion	The lowest point in the midline on the lower border of the chin
Pogonion	The craniometric point that is the most forward-projecting point on the anterior surface of the chin
Menton	The most inferior point of the chin
Bilateral landmarks	
Name of landmark	Description of landmark
Endocanthion	The inner corner of the eye fissure where the eyelids meet, not the caruncles (the red eminences at the medial angles of the eyes)
Exocanthion	The outer corner of the eye fissure where the eyelids meet
Tragion	The point located at the notch just above the tragus of the ear This point corresponds approximately to the upper edge of the ear hole
Alare	The most lateral point on the nasal ala
Gonion	The most lateral point at the angle of the mandible

1.3 Multivariate Morphometrics

Morphometrics is the study of shape variation and its covariation with other variable and it represents an integral part of organismal biology (Adams 1999). Or again, following the definition given in Bookstein (1991), *morphometrics is the exploration of the relationship between extrinsic (e.g. geography, species, sex, etc.) and intrinsic (e.g. growth) factors, and patterns of form variation. It also concerns itself with the localization and characterization of such form differences.* Its goal is the objective description of the changes in the form of an organism—its shape and size—during

ontogeny or during the course of evolution (Bookstein 1986). This description is an abstract representation (e.g., configuration of sets of landmarks) of the specimens under study. As a consequence, different representations of specimens are therefore to be expected to generate results that differ to some degree (O'Higgins 2000).

Database of landmark locations are usually processed using techniques such as multivariate morphometrics and deformation analysis. Actually one can evaluate configurations of landmark points by means of variables expressing aspects of size or shape of single specimens, like distances or ratios of distances, or can directly measure the relation between one form and another as a deformation (Dryden and Mardia 1998). Both strategies represent useful tools to examine group differences in size and shape or between size change and shape change (Bookstein 1986). With reference to multivariate morphometrics, this approach is often applied without regard for homology, i.e., it does not require that size or shape measures derive from the locations of homologous landmarks. As a consequence, the homology of linear distances is difficult to assess, because many distances (e.g., maximum width) are not defined by homologous points. The large amount of measurements obtained through this method is analyzed in the conventional multivariate statistical analysis—canonical variates analysis, principal components analysis, factor analysis, linear modeling, discriminatory analysis, component extraction—and any findings are interpreted coefficient by coefficient. In classic morphometric analysis, measurements are taken in such a way that the geometry of the full landmark configuration is lost. Moreover visualizing results through graphical representations of shape is very demanding because the geometric relationships among the variables (linear distances) are not preserved, thus losing some aspects of shape. Results seem to be more abstract/mathematical rather than pictorial and anatomical.

To conclude these approaches provide precise mathematical descriptions of patterns of covariance between collected variables, but they do not lead to readily interpretable analyses of the size and shape differences under study (O'Higgins 2000). Another approach consists in describing differences between sets of landmark coordinates not in terms of absolute movements but in terms of deformations using “transformation grids” (Thompson 1961): stretchings and contractions of space are then illustrated.

Deformation analysis has been introduced into descriptive biology by Thompson (1961) under the label of “Cartesian Transformation.” We recall the notion of deformation as given in Dryden and Mardia (1998).

Definition *A deformation is a mapping which takes neighbouring points to neighbouring points and which alters lengths of little segments by factors which never get too large or too small. It is an informal version of what the mathematician calls a diffeomorphism, a one-to-one transformation which, along with its inverse, has a derivative at every point of a region and its image.*

Thompson (1961) suggested to observe directly a comparison of biological forms, as a geometric object of measurement in its own right, rather than as the mere numerical difference of measures made upon forms separately. In particular he

proposed to represent the form change as a deformation of the picture plane corresponding closely to what biologists already knew as homology: the smooth mapping of one form onto the other sending landmarks onto their homologues and interpolated suitably in between (Bookstein 1986).

But the field of morphometrics has lately experienced a revolution. Actually, in the 1980s, various authors, among whom we mention Fred Bookstein and James Rohlf, proposed to combine traditional multivariate morphometrics and deformation analysis, calling this synthesis *geometric morphometrics*. This field deals with methods for the analysis of configurations of landmarks in which their full geometry is preserved throughout the analyses and which operate in a specific shape space, Kendall's shape space (Rohlf 1999; O'Higgins 2000).

The term "geometric" referred the geometry of Kendall's shape space: the estimation of mean shapes and the description of sample variation of shape using the geometry of Procrustes distance. Multivariate morphometrics is usually carried out in a linear tangent space to the non-Euclidean shape space in the vicinity of the mean shape.

It could be defined as a collection of approaches for the multivariate statistical analysis of Cartesian coordinate data, often limited to landmark point locations. More directly it is described as the class of morphometric methods that capture the geometry of the morphological structures of interest and preserve complete information about the relative spatial arrangements of the data throughout the analyses. As a consequence, results of high-dimensional multivariate analyses can be mapped back into physical space to achieve appealing and informative visualizations, contrary to alternative traditional methods (Slice 2005).

Geometric morphometric approaches represent a powerful and sophisticated tool allowing to simultaneously visualize and statistically quantify differences in form. Given the rigor of geometric morphometric approaches, the easiness of digitally acquiring landmark data and the availability of free and user-friendly image processing softwares, geometric morphometric studies are becoming more and more frequent in the biological and biomedical literature.

The direct analysis of databases of landmark locations is not convenient because of the presence of nuisance parameters, such as position, orientation, and size. Once obtained the raw landmark data for a number of specimens, the next step of any analysis is to translate and rotate the landmark configurations into a common position and remove size differences between them. This operation is called superimposition and facilitates comparison of configurations by removing variation associated with differences in their location, orientation, and size. Such differences are irrelevant in a comparison of configuration shape, as follows from the definition of shape given by Kendall (1977).

Hence, to carry out a valuable statistical shape analysis, a generalized least-squares superimposition (GLS or Generalized Procrustes Analysis, GPA) is performed to eliminate non-shape variation in configurations of landmarks and to align the specimens to a common coordinate system (Rohlf and Slice 1990). Along with GPA, we mention another registration method, i.e., the two-point registration, that provides Bookstein's shape coordinates. The aligned specimens identify points

in a non-Euclidean space, which is approximated by a Euclidean tangent space for standard multivariate statistical analyses (Slice et al. 1996). With reference to GPA superimposition method, at first, the centroid of each configuration is translated to the origin, and configurations are scaled to a common unit size (by dividing by centroid size, see Bookstein 1986). Finally, the configurations are optimally rotated to minimize the squared differences between corresponding landmarks (Gower 1975; Rohlf and Slice 1990). This is an iterative process and it is useful to compute the mean shape, which is inestimable prior to superimposition. Generalized resistant-fit (GRF) procedures, providing median and repeated median-based estimates of fitting parameters rather than least-squares estimates, are also available (Slice et al. 1996). In particular they are more efficient for revealing differences between two objects when the major differences are mostly in the relative positions of a few landmarks (Rohlf and Slice 1990). Even if they lack the well-developed distributional theory associated with the least-squares fitting techniques, being robust, these methods seem to be protected against departure from the assumptions of the analysis (e.g., independent, identically, and normally distributed errors) and seem to be unresponsive to the potentially strong influences of atypical or incorrect data values (Siegel and Benson 1982).

In the presence of semilandmarks, a newsworthy method is that of “sliding semilandmarks,” allowing outlines to be combined with landmark data in one analysis, providing a richer description of the shapes. The iterative procedure involves at first sliding the semilandmarks to the left or right along a curve during the GPA superimposition in an attempt to minimize the distance between the adjusted position and the corresponding point in the consensus or to reduce the overall bending energy required to fit the specimens to the sample average configuration. Computations are iterative and the algorithm provides smooth and interpretable deformation grids among the forms. For details, see Bookstein (1997), Adams et al. (2004), Slice et al. (1996), and TpsRelw soft4ware guide by Rohlf (2008).

After superimposition, differences in shape can be described either in terms of differences in coordinates of corresponding landmarks between objects (Bookstein 1996) or in terms of differences in the deformation grids representing the objects, e.g., using the thin-plate spline method (Bookstein 1991).

The thin-plate spline is a global interpolating function that maps the landmark coordinates of one specimen to the coordinates of the landmarks in another specimen and represent a mathematically rigorous realization of Thompson (1961) idea of transformation grids, where one object is deformed or “warped” into another. The parameters describing these deformations (partial warp scores) can be used as shape variables for statistical comparisons of variation in shape within and between populations (Adams 1999). As a result, the thin-plate splines can be interpreted as one method of generating a coordinate system for tangent space mentioned above. Along with the superimposition methods, several alternative procedures for obtaining shape information from landmark data have been proposed (Adams et al. 2004). Here we mention Euclidean Distance Matrix Analysis (EDMA) methods proposed by (Lele and Richtsmeier 1991), a related approach using

standard multivariate methods on logs of size-scaled interlandmark distances (Rao and Suryawanshi 1996) and methods based on interior angles (Rao and Suryawanshi 1998).

References

- Adams DC (1999) Methods for shape analysis of landmark data from articulated structures. *Evolutionary Ecology Research* 1:959–970
- Adams DC, Rohlf FJ, Slice DE (2004) Geometric morphometrics: ten years of progress following the ‘revolution’. *Italian Journal of Zoology* 71:5–16
- Bookstein FL (1978) The measurement of biological shape and shape change. Lecture notes on biomathematics. Springer-Verlag, New York, vol 24
- Bookstein FL (1986) Size and shape spaces for landmark data in two dimensions. *Statistical Science* 1:181–242
- Bookstein FL (1989) Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11:567–585
- Bookstein FL (1991) *Morphometric Tools For Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge
- Bookstein FL (1996) Combining the tools of geometric morphometrics. In: *Advances in morphometrics*, vol 284, Plenum Press, New York, pp 131–151
- 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, Giurisato M, Salmaso L, Cozzi B (2009) 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, Ferronato G, Galzignato P (2011) Multi-aspect procedures for paired data with application to biometric morphing. *Communications in Statistics - Simulation and Computation* 40:1–12
- 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
- Etöz A (2011) Anthropometric analysis of the nose. In: Brenner DM (ed) *Rhinoplasty*, InTech
- Farkas LG (1994) *Anthropometry of the Head and Face*. Raven Press, New York
- Farkas LG, Katic MJ, Forrest CR (2005) International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery* 16:615–646
- Goodall CR (1991) Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, Series B* 53:285–339
- Gower JC (1975) Generalized procrustes analysis. *Psychometrika* 40:33–50
- Huxley JS (1932) *Problems of Relative Growth*. The Dial Press, New York
- Katina S, Bookstein FL, Gunz P, Schaefer K (2007) What is worth digitizing all those curve? a worked example from craniofacial primatology. *American Journal of Physical Anthropology, Abstracts of American Academy of Physician Assistants (AAPA) poster and podium presentations* 132, S44:140

- Kendall DG (1977) The diffusion of shape. *Advances in Applied Probability* 9:428–430
- Kendall DG (1984) Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* 16:81–121
- Kendall DG (1989) A survey of the statistical theory of shape. *Statistical Science* 4:87–120
- Kent JT (1994) The complex bingham distribution and shape analysis. *Journal of the Royal Statistical Society: Series B* 56:285–299
- Kent JT (1995) Current issues for statistical inference in shape analysis. In: Mardia KV, Gill CA (eds) *Proceedings in Current Issues in Statistical Shape Analysis*, University of Leeds, University of Leeds Press, pp 167–175
- Le H, Kendall DG (1993) The riemannian structure of euclidean shape spaces: a novel environment for statistics. *Annals of Statistics* 21:1225–1271
- 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
- Lele SR, Richtsmeier JT (2001) *An invariant approach to statistical analysis of shapes*. Chapman & Hall/CRC, Boca Raton, Florida
- Mardia KV (1995) Shape advances and future perspectives. In: Mardia KV, Gill CA (eds) *Proceedings in Current Issues in Statistical Shape Analysis*, University of Leeds, University of Leeds Press, pp 57–75
- Mardia KV, Dryden IL (1989) The statistical analysis of shape data. *Biometrika* 76:271–281
- Mardia KV, Edwards R, Puri ML (1977) Analysis of Central Place Theory. *Bulletin of the International Statistical Institute* 47:93–110
- Mo G (2005) *Anatomy of the skull of mediterranean monk seal, Monachus monachus, functional morphology, densitometry, morphometrics and notes on natural history*. PhD thesis, University of Padova
- O’Higgins P (2000) The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. *Journal of Anatomy* 197:103–120
- 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
- Rao C, Suryawanshi S (1996) Statistical analysis of shape of objects based on landmark data. *Proceedings of the National Academy of Sciences of the United States of America* 93:12,132–12,136
- Rao C, Suryawanshi S (1998) Statistical analysis of shape through triangulation of landmarks: a study of sexual dimorphism in hominids. *Proceedings of the National Academy of Sciences of the United States of America* 95:4121–4125
- Rasband W (1997–2012) *ImageJ*. U. S. National Institutes of Health, Bethesda, Maryland, USA, State University of New York at Stony Brook, <http://imagej.nih.gov/ij/>
- Rohlf FJ (1999) Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification* 16:197–225
- Rohlf FJ (2007) *tpsDig2*, digitize landmarks and outlines, version 2.12. Department of Ecology and Evolution, State University of New York at Stony Brook
- Rohlf FJ (2008) *tpsRelw*, Relative warps analysis, version 1.46. Department of Ecology and Evolution, State University of New York at Stony Brook
- Rohlf FJ, Slice DE (1990) Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* 39:40–59
- Schneider C, Rasband W, Eliceiri K (2012) Nih image to imagej: 25 years of image analysis. *Nature Methods* 9:671–675
- Siegel AF, Benson RH (1982) A robust comparison of biological shapes. *Biometrics* 38:341–350
- Slice DE (2005) *Modern Morphometrics In Physical Anthropology*. Springer-Verlag New York, LLC
- Slice DE, Bookstein FL, Marcus LF, Rohlf FJ (1996) A glossary for geometric morphometrics. *Advances in Morphometrics* 284:531–551

- Small CG (1988) Techniques of shape analysis on sets of points. *International Statistical Review* 56:243–257
- Small CG (1996) *The Statistical Theory of Shape*. Springer, New York
- Spoor F, Jefferies N, Zonneveld F (2000) Imaging skeletal growth and evolution. In: O’Higgins P, Cohn MJ (eds) *Development, Growth and Evolution : Implications for the Study of the Hominid Skeleton*, London: Academic Press, pp 123–161
- Srivastava A, Joshi S, Mio W, Liu X (2005) Statistical shape analysis: Clustering, learning and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27:590–602
- Stoyan D, Stoyan H (1994) *Fractals, Random Shapes and Point Fields: Methods of geometric statistics*. Wiley, Chichester
- Stoyan D, Kendall WS, Mecke J (1995) *Stochastic geometry and its applications*, 2nd Edition. Wiley, Chichester, England
- Thompson DW (1961) *On Growth and Form*. University Press, Cambridge
- Weber GW, Bookstein FL (2011) *Virtual Anthropology: a guide to a new interdisciplinary field*. Springer: New York

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

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

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}_j 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

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

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

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

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

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

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).

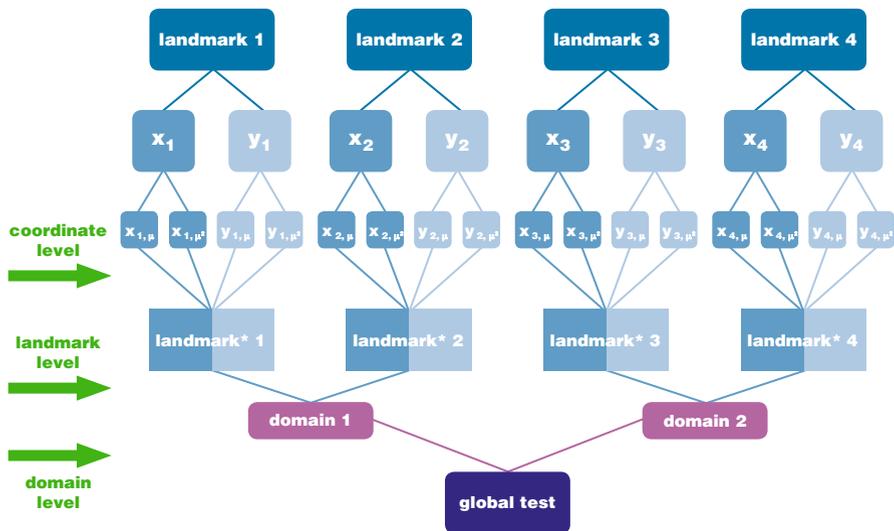


Fig. 2.1 Different levels of combination

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

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

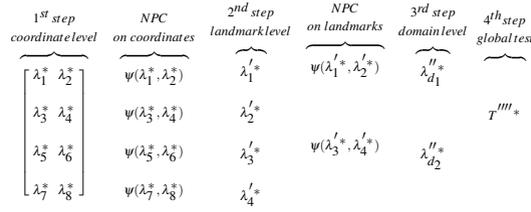


Fig. 2.2 Algorithm for $k = 4$ landmarks in 2D and two domain combinations

- 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 Kolmogorov–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

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

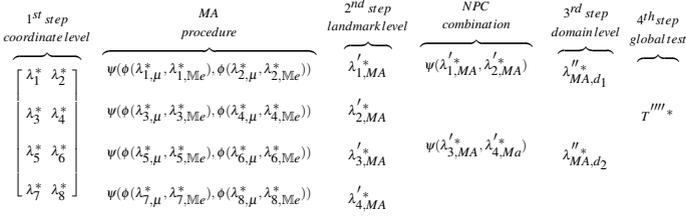


Fig. 2.3 Algorithm for $k = 4$ landmarks in 2D, MA procedure (mean and median aspects) and domain combination. ψ and ϕ are suitable combining functions

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 $\max(0.0067, 0.0220, 0.0285, 0.0285, 0.0570, 0.0580, 0.0600, 0.0618) = 0.0618$.

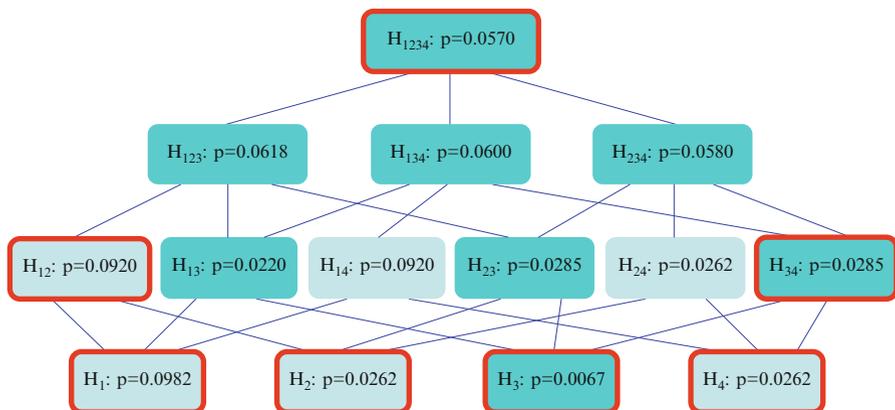


Fig. 2.4 Illustration of the closed testing procedure, focussing on landmark level ($k = 4$)

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, Giurionato 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
- Terriberly 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

Chapter 3

Evaluating Power Behavior of Nonparametric Combination Testing Methodology After Generalized Procrustes Analysis and Under Different Correlation Structures

“We call a thing big or little with reference to what it is wont to be, as we speak of a small elephant or a big rat.”.

On Growth and Form, 1917

D’Arcy Thompson

Abstract In this chapter we investigate the effect of Generalized Procrustes Analysis (GPA) superimposition on the power of NonParametric Combination (NPC) testing methodology. Through a simulation study, we show how GPA alters power function. Since GPA superimposition provides permutationally non-equivalent transformations, NPC tests are approximate.

Moreover, we examine the case of correlated landmarks. Through a toy example, considering hypothetical configurations representing some 3D monkey skulls, we compare the behavior of traditional tests with that of nonparametric permutation tests in this particular case.

Finally we introduce paired data problem that, in the context of shape analysis, relates to the study of symmetric structures. Inferential techniques for symmetric shapes are presented.

Keywords Correlated landmarks • Generalized Procrustes Analysis GPA superimposition • Paired landmarks • Symmetric structures

3.1 Generalized Procrustes Analysis and Power of NonParametric Combination methodology

In Chap. 2, in order to carry out NPC tests, we have used raw coordinates and not shape coordinates. This choice could be questioned. For example, one could ask how the NPC tests properly ensure the invariances to translation, rotation, and

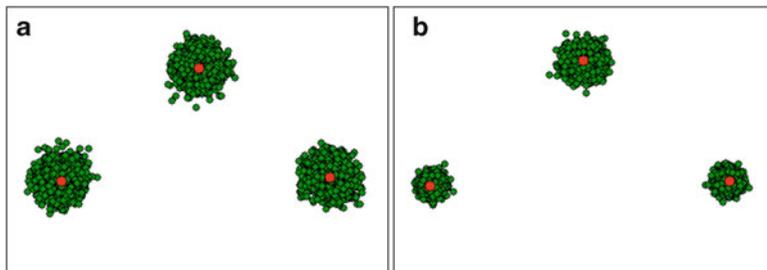


Fig. 3.1 Raw data (a) and superimposed data (b)

scale needed for proper shape tests. We could replay that we were considering configurations of landmarks different by construction. In addition, possible difference in power behavior is associated with transformations induced by GPA. Actually, including GPA, NPC tests are approximate, since GPA superimposition provides permutationally non-equivalent transformations (Brombin 2009). Moreover, the probability distribution of transformed data after GPA may be altered with respect to the initial distribution. Hence GPA privileges the shape, but it may alter the dependency structures and, as a result, the distribution producing permutationally non-equivalent tests within the permutation testing framework. In the extreme case, if we consider two shapes that differ only for a scale factor (e.g., a big and a small circle), without GPA, inferential results obtained using NPC tests lead us to accept the alternative hypothesis, i.e., the two shapes are significantly different. On the other hand, after superimposition, we just accept the null hypothesis, stating the equality of the two shapes. Hence, inferential conclusions may be highly different.

We could regard GPA superimposition as a method for standardizing shapes. It is well known that different results may be obtained using standardized or original data. For example, in multivariate statistics, this situation occurs in principal component analysis (PCA). Actually, the components obtained using variance or correlation matrix are in general not the same, nor is possible to pass from one solution to another by a simple scaling of the coefficient.

We recall this issue in Sect. 3.2, examining simulation results in the presence of correlation between landmarks. In Fig. 3.1 we show the effects of GPA superimposition. The scatterplot has been realized using the tutorial program TPStri (Rohlf 2008), allowing to show some of the relationships between shape coordinates and to perform sampling experiments for triangles. In particular we have generated and plotted 2,000 random triangles (shown as small green dots) from normal distributions, centered on the target shapes (to simulate the effects of random shape variation such as digitizing error). The target (mean) shape is an equilateral triangle represented by the red dots (with $(-\frac{1}{2}, -\frac{1}{2\sqrt{3}})$, $(\frac{1}{2}, \frac{1}{2\sqrt{3}})$ as the endpoints of the base) and it has been chosen close to the reference (an equilateral triangle too with $(0,0)$, $(0,1)$ as endpoints of the base).

A “medium” standard deviation $\sigma = 0.05$ for the scatter around each landmark has been chosen (i.e., we simulate digitizing error). In Fig. 3.1a we have displayed the raw scatter, while in Fig. 3.1b we have displayed the scatter after Procrustes aligning each sample to the reference configuration.

As you can see, after GPA superimposition, the variance around each landmark is greatly reduced and this of course can influence power behavior.

Here we propose the same simulation study presented in Chap. 2, considering only NPC tests. We compare their power behavior, in the case in which superimposition step is included or not. Let n_j , $j = 1, 2$, be the sample size in the two samples. Three domains have been defined. We denote with G the global test obtained after combining all partial tests and with G_d the global test that takes into account the information about domains. Fisher (F) and Liptak (L) are the possible combining functions used and MA denotes the Multi Aspect procedure previously discussed. For the sake of space we report only the results for the following simulations:

- 1st simulation: $n_1 = n_2 = 10$, $\sigma^2 = 0.25$,
- 2nd simulation: $n_1 = n_2 = 10$, $\sigma^2 = 0.50$,
- 3rd simulation: $n_1 = 50$, $n_2 = 20$, $\sigma^2 = 0.25$.
- 4th simulation: $n_1 = n_2 = 50$, $\sigma^2 = 0.25$.

We also show simulation results under the null hypothesis, in order to evaluate if the nominal α -level, after GPA superimposition, is still under control.

As expected, power function after GPA is slightly different, thus confirming that transformations are induced by GPA (see the rows highlighted in gray in Tables 3.1–3.4).

The type I error is under control (see the column of nominal level $\alpha = 0.05$ in Tables 3.5–3.8). By means of GPA superimposition we are able to compare shapes on the basis of points of correspondence (i.e., landmarks). It is worth noting that other methods are available, allowing for comparisons of entire curves. In particular, it is possible to compare the parameters or coefficients describing a curve of interest within functional data analysis field.

3.2 Introducing Correlation Between Landmarks

In this nonparametric framework we also analyze the case of heterogeneous variation at each landmark. We have evaluated power and the achieved α -level. Superimposition step has been included in the routine. Tippett (T) and Fisher (F) combining functions have been used, considering both location and distributional aspects.

Full simulation study has been published elsewhere in Brombin (2009), Brombin et al. (2009), Pesarin (2001), and Pesarin and Salmaso (2010).

In order to obtain nonsingular covariance matrix, we have performed an eigenvalue decomposition (ED) of the original variance covariance matrix and transformed the original eigenvalues λ . We have considered transformations like $\lambda^{1/3}$ and $\lambda^{1/10}$,

Table 3.1 1st simulation ($n_1 = n_2 = 10$, $B = CMC = 1,000$, $\sigma^2 = 0.25$)

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L) & GPA	0.054	0.172	0.279	0.434	0.549	0.741
G (L)	0.049	0.171	0.281	0.461	0.581	0.766
G_d (L) & GPA	0.048	0.165	0.265	0.419	0.526	0.724
G_d (L)	0.040	0.168	0.275	0.436	0.567	0.743
G (F) & GPA	0.079	0.214	0.326	0.480	0.615	0.769
G (F)	0.072	0.233	0.365	0.544	0.650	0.810
G_d (F) & GPA	0.067	0.203	0.302	0.452	0.581	0.753
G_d (F)	0.067	0.216	0.343	0.512	0.619	0.787
G (L, MA) & GPA	0.056	0.175	0.275	0.440	0.549	0.748
G (L, MA)	0.057	0.176	0.303	0.458	0.586	0.749
G_d (L, MA) & GPA	0.052	0.170	0.265	0.419	0.527	0.723
G_d (L, MA)	0.052	0.177	0.289	0.442	0.566	0.739
G (F, MA) & GPA	0.075	0.218	0.312	0.477	0.604	0.767
G (F, MA)	0.075	0.235	0.365	0.533	0.659	0.801
G_d (F, MA) & GPA	0.071	0.203	0.290	0.447	0.568	0.755
G_d (F, MA)	0.065	0.211	0.344	0.508	0.630	0.780

Table 3.2 2nd simulation ($n_1 = n_2 = 10$, $B=CMC=1,000$, $\sigma^2 = 0.50$)

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L) & GPA	0.031	0.098	0.180	0.330	0.436	0.647
G (L)	0.027	0.108	0.186	0.339	0.443	0.648
G_d (L) & GPA	0.029	0.098	0.167	0.316	0.424	0.632
G_d (L)	0.020	0.097	0.183	0.333	0.438	0.630
G (F) & GPA	0.033	0.110	0.196	0.365	0.475	0.647
G (F)	0.034	0.134	0.227	0.369	0.498	0.689
G_d (F) & GPA	0.030	0.115	0.188	0.345	0.467	0.631
G_d (F)	0.036	0.120	0.210	0.356	0.485	0.675
G (L, MA) & GPA	0.030	0.108	0.176	0.330	0.439	0.643
G (L, MA)	0.022	0.109	0.206	0.345	0.446	0.651
G_d (L, MA) & GPA	0.028	0.101	0.170	0.311	0.427	0.627
G_d (L, MA)	0.021	0.098	0.193	0.334	0.438	0.647
G (F, MA) & GPA	0.029	0.110	0.201	0.362	0.475	0.645
G (F, MA)	0.032	0.142	0.242	0.372	0.493	0.695
G_d (F, MA) & GPA	0.027	0.113	0.190	0.347	0.467	0.634
G_d (F, MA)	0.028	0.126	0.221	0.356	0.491	0.678

Table 3.3 3rd simulation: $n_1 = 50, n_2 = 20, \sigma^2 = 0.25$

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L) & GPA	0.200	0.422	0.564	0.738	0.829	0.929
G (L)	0.000	0.787	0.870	0.932	0.959	0.987
G.d (L) & GPA	0.162	0.390	0.524	0.691	0.790	0.914
G.d (L)	0.088	0.761	0.854	0.923	0.955	0.985
G (F) & GPA	0.305	0.588	0.737	0.858	0.918	0.965
G (F)	0.385	0.677	0.797	0.907	0.953	0.990
G.d (F) & GPA	0.264	0.549	0.700	0.833	0.904	0.958
G.d (F)	0.337	0.634	0.771	0.887	0.940	0.984
G (L, MA) & GPA	0.195	0.415	0.560	0.742	0.828	0.926
G (L, MA)	0.488	0.690	0.811	0.889	0.931	0.971
G.d (L, MA) & GPA	0.157	0.382	0.516	0.696	0.802	0.918
G.d (L, MA)	0.476	0.686	0.791	0.878	0.920	0.965
G (F, MA) & GPA	0.287	0.579	0.736	0.857	0.920	0.961
G (F, MA)	0.564	0.775	0.863	0.920	0.954	0.982
G.d (F, MA) & GPA	0.254	0.542	0.699	0.835	0.908	0.959
G.d (F, MA)	0.549	0.767	0.853	0.911	0.948	0.982

Table 3.4 4th simulation ($n_1 = n_2 = 50, B = MC = 1,000, \sigma^2 = 0.25$)

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L) & GPA	0.429	0.702	0.812	0.907	0.952	0.989
G (L)	0.354	0.644	0.759	0.867	0.916	0.972
G.d (L) & GPA	0.386	0.674	0.795	0.895	0.937	0.981
G.d (L)	0.332	0.615	0.752	0.856	0.912	0.966
G (F) & GPA	0.653	0.884	0.942	0.980	0.992	1.000
G (F)	0.677	0.873	0.944	0.976	0.989	1.000
G.d (F) & GPA	0.610	0.869	0.931	0.972	0.990	0.997
G.d (F)	0.641	0.855	0.922	0.970	0.985	0.998
G (L, MA) & GPA	0.428	0.707	0.814	0.904	0.956	0.985
G (L, MA)	0.359	0.630	0.762	0.877	0.920	0.979
G.d (L, MA) & GPA	0.382	0.671	0.795	0.883	0.938	0.979
G.d (L, MA)	0.340	0.619	0.755	0.864	0.909	0.973
G (F, MA) & GPA	0.634	0.882	0.936	0.980	0.993	1.000
G (F, MA)	0.678	0.883	0.939	0.977	0.991	0.999
G.d (F, MA) & GPA	0.595	0.866	0.923	0.971	0.987	0.997
G.d (F, MA)	0.651	0.863	0.925	0.971	0.985	0.999

Table 3.5 Simulations under H_0 ($n_1 = n_2 = 10$, $B = \text{CMC} = 1,000$, $\sigma^2 = 0.25$) using GPA

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L)	0.017	0.054	0.107	0.215	0.307	0.512
G_d (L)	0.014	0.055	0.115	0.213	0.304	0.512
G (F)	0.014	0.058	0.120	0.208	0.307	0.506
G_d (F)	0.013	0.057	0.121	0.205	0.313	0.490
G (L, MA)	0.016	0.051	0.109	0.217	0.311	0.516
G_d (L, MA)	0.013	0.046	0.116	0.218	0.310	0.514
G (F, MA)	0.010	0.058	0.118	0.214	0.311	0.504
G_d (F, MA)	0.012	0.061	0.115	0.210	0.317	0.484

Table 3.6 Simulations under H_0 ($n_1 = n_2 = 10$, $B = \text{CMC} = 1,000$, $\sigma^2 = 0.50$) using GPA

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L)	0.007	0.053	0.099	0.198	0.300	0.518
G_d (L)	0.012	0.060	0.101	0.203	0.313	0.520
G (F)	0.007	0.047	0.107	0.196	0.299	0.509
G_d (F)	0.012	0.055	0.107	0.205	0.299	0.500
G (L, MA)	0.008	0.055	0.099	0.201	0.297	0.514
G_d (L, MA)	0.007	0.059	0.104	0.207	0.312	0.519
G (F, MA)	0.005	0.051	0.105	0.195	0.298	0.504
G_d (F, MA)	0.008	0.049	0.096	0.200	0.295	0.503

Table 3.7 Simulations under H_0 ($n_1 = 50$, $n_2 = 20$, $B = \text{CMC} = 1,000$, $\sigma^2 = 0.25$) using GPA

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L)	0.014	0.054	0.098	0.196	0.289	0.491
G_d (L)	0.014	0.056	0.096	0.200	0.296	0.502
G (F)	0.013	0.053	0.103	0.188	0.289	0.488
G_d (F)	0.013	0.052	0.097	0.196	0.293	0.489
G (L, MA)	0.014	0.056	0.095	0.194	0.292	0.494
G_d (L, MA)	0.013	0.053	0.095	0.198	0.299	0.492
G (F, MA)	0.017	0.053	0.098	0.189	0.284	0.485
G_d (F, MA)	0.014	0.048	0.101	0.194	0.285	0.497

rescaled by their trace (see the effect of transforming eigenvalues on the scatterplot in Fig. 3.2). Then we have recalculated the covariance matrix Σ^* , using the relation $\Sigma^* = V\Lambda^*V'$, where Λ^* is a diagonal matrix with the transformed eigenvalues, V is an orthogonal matrix, containing the corresponding eigenvectors and V' means V transposed.

Under the alternative, data have been generated using different means and the same covariance matrix Σ^* . In Table 3.9 we display hypothetical mean configurations, representing 3D male and female *Macaca fascicularis* monkey skulls (for details, see Frost et al. 2003).

Table 3.8 Simulations under H_0 ($n_1 = n_2 = 50$, $B=CMC=1,000$, $\sigma^2 = 0.25$) using GPA

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$	$\alpha = 0.30$	$\alpha = 0.50$
G (L)	0.011	0.048	0.097	0.200	0.316	0.510
G.d (L)	0.013	0.048	0.098	0.204	0.311	0.516
G (F)	0.013	0.045	0.108	0.204	0.297	0.515
G.d (F)	0.010	0.051	0.113	0.207	0.303	0.521
G (L, MA)	0.009	0.050	0.096	0.212	0.309	0.503
G.d (L, MA)	0.011	0.049	0.096	0.207	0.313	0.513
G (F, MA)	0.011	0.051	0.105	0.198	0.302	0.514
G.d (F, MA)	0.010	0.050	0.107	0.203	0.294	0.515

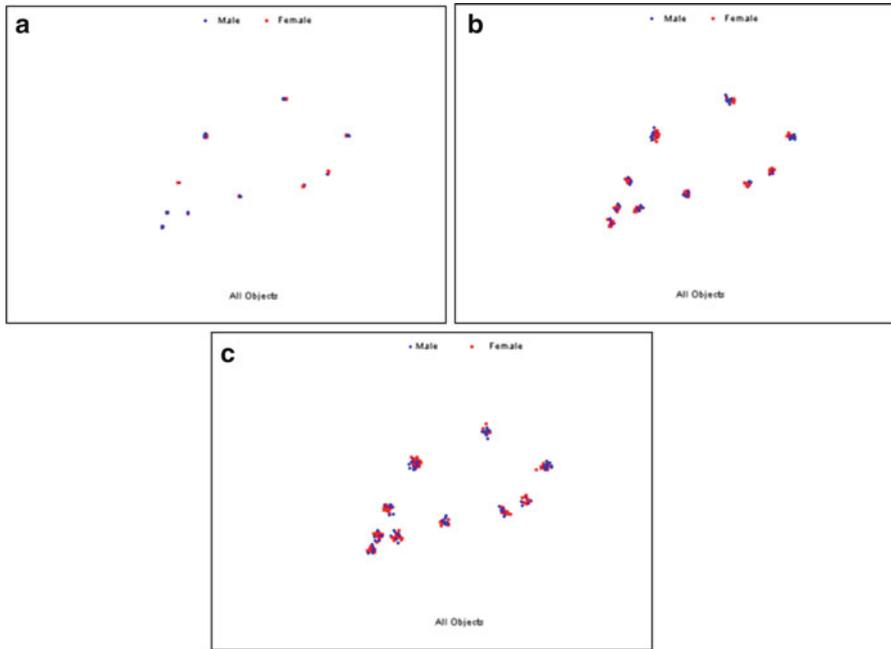


Fig. 3.2 Original eigenvalues λ (a), $\lambda^{1/3}$ (b), and $\lambda^{1/10}$ (c)

Let n_j , $j = 1, 2$, denote the sample size in the two groups. In particular we have considered these settings: $n_1 = n_2 = 5$, $n_1 = n_2 = 10$, $n_1 = 5$, $n_2 = 10$. In the simulation study we have evaluated power and α -level when the number of 3D landmarks k was, in turn, equal to 3, 6, 9, 11. Three domains have been considered, i.e., the first includes landmarks 1, 2, and 11; the second one includes landmarks from 3 to 7; and the third one includes landmarks from 8 to 10. We denote with T'' the Hotelling's T^2 permutation counterpart, with G the combination of all partial tests, with G.d the combination using domains.

Table 3.9 Configurations

Landmark		Male			Female		
#	Lnd. name	x_M	y_M	z_M	x_F	y_F	z_F
1	Inion	17.7752	18.9981	6.9585	17.5252	18.9981	6.9585
2	Bregma	15.9101	16.3499	9.2159	15.9101	16.4499	9.2159
3	Glabella	13.6833	12.7086	7.6433	13.6833	12.7086	7.6433
4	Nasion	13.6799	12.6892	7.5628	13.8299	12.6892	7.5628
5	Rhinion	12.9273	11.2649	5.1792	12.9273	11.2149	5.1792
6	Nasospinale	12.6114	10.5523	3.6257	12.6114	10.5523	3.6257
7	Prosthion	12.4725	10.233	2.8531	12.4725	10.2330	2.8531
8	Opisthion	17.1882	17.8852	5.0014	17.1882	17.8852	5.1514
9	Basion	16.5070	16.7665	4.4799	16.5070	16.7165	4.4799
10	Staphylion	14.6975	13.8755	4.1783	14.6075	13.8755	4.1783
11	Incisivion	13.2442	11.4665	3.5466	13.2442	11.4665	3.5166

We remark that transforming eigenvalues as above mentioned, it is difficult to quantify the amount of the correlation introduced in the dataset. We did not calculate any multivariate correlation index, since we are just interested in evaluating what happens to power behavior after introducing some correlations between landmarks. By the way we would like to mention that the original model by [Goodall \(1991\)](#) described a perturbation model as a simple model for variation in the positions of the landmarks around their mean locations. This model also allows for covariation between the landmarks. In this model the $k \times m$ matrix of coordinates for the k m -dimensional landmarks for the i th specimen is given by

$$\mathbf{X}_i = \alpha_i(\boldsymbol{\mu} + \mathbf{E}_i)\boldsymbol{\Omega}_i + \mathbf{1}\omega_i^t$$

where α_i is a scale factor (size of the i th specimen), $\boldsymbol{\mu}$ is the mean shape, \mathbf{E}_i is a matrix of random errors (normally distributed with zero mean), $\boldsymbol{\Omega}_i$ is a $m \times m$ matrix describing the orientation of the i th specimen (reflections excluded), $\mathbf{1}$ is a vector of all ones, and ω_i is a vector specifying the location of the specimen in the digitizing plane (or solid). Parameters α_i , $\boldsymbol{\Omega}_i$, and ω_i are the so-called nuisance parameters because they encode information unrelated to shape variation. As previously noted, the estimates of shape variation must be independent of these parameters. Matrix \mathbf{E}_i (when strung out as a single column vector with mk elements) has a covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_k \otimes \boldsymbol{\Sigma}_m$, where $\boldsymbol{\Sigma}_k$ is the covariance matrix for the landmark points and $\boldsymbol{\Sigma}_m$ is the covariance matrix for the dimensions ([Rohlf 2000](#)). The symbol \otimes denotes the Kronecker product of two matrices. Till now we have investigated the simplest case of identical independent variation around each mean landmark position, i.e., $\boldsymbol{\Sigma}_k = \sigma^2\mathbf{I}_k$ and $\boldsymbol{\Sigma}_m = \mathbf{I}_m$. This is the type of variation one might expect from digitizing error. But Goodall's F test [1991](#), related to this model for shape variation, in addition to assuming independent samples and that \mathbf{E}_i follows the multivariate normal distribution, makes the further assumptions that $\boldsymbol{\Sigma}_k = \sigma^2\mathbf{I}$

Table 3.10 Simulations under $H_0: n_1 = n_2 = 10, \lambda^{1/3}, k = 6, m = 3, B = CMC = 1,000$

Test	Achieved α -level					
	0.01	0.05	0.10	0.20	0.30	0.50
T''	0.012	0.056	0.102	0.204	0.295	0.493
G_T	0.000	0.052	0.088	0.195	0.290	0.489
$G_{d,T}$	0.004	0.038	0.103	0.195	0.288	0.475
G_F	0.011	0.055	0.099	0.197	0.292	0.504
$G_{d,F}$	0.011	0.050	0.098	0.196	0.293	0.496
$G_{MA,T}$	0.008	0.031	0.065	0.136	0.206	0.350
$G_{d,MA,T}$	0.008	0.027	0.073	0.143	0.216	0.345
$G_{MA,F}$	0.041	0.112	0.164	0.266	0.340	0.508
$G_{d,MA,F}$	0.043	0.122	0.185	0.305	0.402	0.550
Goodall	0.025	0.069	0.119	0.201	0.284	0.465
T^2	0.033	0.098	0.159	0.240	0.331	0.470
$G_{(\mu, \mathbb{M}e), F}$	0.012	0.055	0.099	0.190	0.288	0.466
$G_{d,(\mu, \mathbb{M}e), F}$	0.011	0.054	0.098	0.191	0.300	0.479
$G_{(\mu, \mu^2), F}$	0.010	0.048	0.109	0.195	0.286	0.509
$G_{d,(\mu, \mu^2), F}$	0.009	0.043	0.096	0.192	0.287	0.485

and $\Sigma_m = \mathbf{I}$ for both samples. Under the assumed model, type I error rate is under control when σ^2 is small.

The model of equal and isotropic variation is a fairly restrictive assumption that is often violated in biological data sets, which often show clearly patterned variation (Klingenberg et al. 2002). However, when the isotropic normal model holds, more powerful tests result, especially when sample sizes are small. Furthermore, we remark that in the usual T^2 tests, power is necessarily low since many degrees of freedom are used in estimating the covariance matrix.

For sake of space, we present simulation results only for the case in which the number of 3D landmarks k is equal to 6 and $n_1 = n_2 = 10$ (see Tables 3.10–3.13).

In all the simulations under H_0 , when using global test with Fisher combining function, MA procedure and domain information, type I error rate was too large, thus invalidating inferential conclusions. Focussing on $\alpha = 0.05$, tests with nominal α -level out of control are highlighted in gray in Tables 3.10–3.11. For example, in Table 3.10, when $\alpha = 0.05$, we have that $G_{MA,F}$ has a corresponding achieved $\alpha = 0.112$ and $G_{d,MA,F}$ has a corresponding achieved $\alpha = 0.122$. We remind that we were considering location (mean) and distributional (Anderson–Darling’s statistic) aspects. As mentioned in the previous section, GPA superimposition may modify dependency structures, thus altering the final distribution. Fisher combining function is more sensitive to MA procedure than Tippett combining function. If we change the aspects, e.g., if we consider mean μ and median $\mathbb{M}e$ or mean and second moments μ^2 , the corresponding $G_{MA,F}$ and $G_{d,MA,F}$ are able to control the nominal α -level. These results highlight that GPA affects the initial distribution of the data, hence a particular care is needed when a MA procedure is performed.

Table 3.11 Simulations under $H_0: n_1 = n_2 = 10, \lambda^{1/10}, k = 6, m = 3, B = CMC = 1,000$

Test	Achieved α -level					
	0.01	0.05	0.10	0.20	0.30	0.50
T''	0.011	0.043	0.089	0.189	0.286	0.504
G_T	0.000	0.040	0.087	0.183	0.282	0.481
$G_{d,T}$	0.007	0.041	0.092	0.187	0.293	0.494
G_F	0.011	0.046	0.090	0.187	0.284	0.500
$G_{d,F}$	0.008	0.045	0.088	0.187	0.300	0.498
$G_{MA,T}$	0.007	0.030	0.065	0.138	0.198	0.338
$G_{d,MA,T}$	0.010	0.030	0.077	0.141	0.197	0.349
$G_{MA,F}$	0.040	0.104	0.165	0.259	0.348	0.503
$G_{d,MA,F}$	0.039	0.130	0.200	0.309	0.417	0.571
Goodall	0.015	0.059	0.111	0.209	0.302	0.501
T^2	0.033	0.090	0.153	0.256	0.332	0.466
$G_{(\mu, Mle),F}$	0.009	0.053	0.098	0.199	0.304	0.509
$G_{d,(\mu, Mle),F}$	0.008	0.052	0.105	0.222	0.311	0.507
$G_{(\mu, \mu^2),F}$	0.013	0.062	0.125	0.223	0.324	0.524
$G_{d,(\mu, \mu^2),F}$	0.008	0.060	0.115	0.208	0.314	0.488

Table 3.12 Simulations under $H_1: n_1 = n_2 = 10, \lambda^{1/3}, k = 6, m = 3, B = CMC = 1000$

Test	Power					
	0.01	0.05	0.10	0.20	0.30	0.50
T''	0.337	0.621	0.757	0.888	0.930	0.981
G_T	0.000	0.538	0.709	0.855	0.923	0.978
$G_{d,T}$	0.123	0.478	0.683	0.828	0.902	0.971
G_F	0.299	0.564	0.709	0.858	0.916	0.973
$G_{d,F}$	0.286	0.547	0.679	0.819	0.898	0.962
$G_{MA,T}$	0.234	0.456	0.636	0.754	0.851	0.945
$G_{d,MA,T}$	0.255	0.401	0.600	0.755	0.824	0.923
Goodall	0.321	0.551	0.667	0.782	0.863	0.949
T^2	0.253	0.483	0.609	0.739	0.808	0.872
$G_{(\mu, Mle),F}$	0.284	0.571	0.711	0.853	0.918	0.969
$G_{d,(\mu, Mle),F}$	0.164	0.431	0.596	0.797	0.880	0.952
$G_{(\mu, \mu^2),F}$	0.224	0.467	0.612	0.740	0.833	0.922
$G_{d,(\mu, \mu^2),F}$	0.137	0.369	0.495	0.663	0.782	0.903

The test performing better are highlighted in gray (see Tables 3.12 and 3.13). Among more powerful tests, we find $G_{d,MA,T}$, Goodall's F , Hotelling's T^2 , $G_{(\mu, \mu^2),F}$, and $G_{d,(\mu, \mu^2),F}$. With reference to Goodall's F and Hotelling's T^2 parametric tests, they are not valid. Actually when nominal α -level is equal to 5%, Goodall's achieved α -level is 6.9% and Hotelling's T^2 achieved α -level is 9.8% (see Table 3.10). Hence the evaluation of power behavior for these tests should be made with caution.

Table 3.13 Simulations under $H_1: n_1 = n_2 = 10, \lambda^{1/10}, k = 6, m = 3, B = CMC = 1000$

Test	Power					
	0.01	0.05	0.10	0.20	0.30	0.50
T''	0.086	0.242	0.370	0.547	0.674	0.820
G_T	0.000	0.162	0.284	0.474	0.604	0.781
$G_{d,T}$	0.039	0.200	0.332	0.495	0.628	0.795
G_F	0.088	0.239	0.362	0.550	0.672	0.817
$G_{d,F}$	0.097	0.260	0.380	0.547	0.667	0.816
$G_{MA,T}$	0.037	0.111	0.218	0.361	0.471	0.656
$G_{d,MA,T}$	0.048	0.133	0.248	0.398	0.507	0.673
Goodall	0.075	0.220	0.339	0.496	0.600	0.769
T^2	0.054	0.155	0.244	0.362	0.448	0.585
$G_{(\mu, \mathbb{M}le),F}$	0.097	0.257	0.369	0.524	0.631	0.782
$G_{d,(\mu, \mathbb{M}le),F}$	0.066	0.219	0.355	0.508	0.626	0.798
$G_{(\mu, \mu^2),F}$	0.066	0.193	0.298	0.435	0.557	0.741
$G_{d,(\mu, \mu^2),F}$	0.047	0.152	0.265	0.424	0.543	0.727

Moreover we recall that significance levels obtained by parametric F -tests may not be reliable. Permutation tests (Good 2000) provide an alternative that can be used for Procrustes analysis of Variance (ANOVA) even when distributions are not normal or sample size is small (Klingenberg and McIntyre 1998; Klingenberg et al. 2002).

3.3 Paired Data Problems: Study of Symmetric Structures

In shape analysis, paired data problem is commonly associated with the study of symmetric structures. Here we briefly summarize the terminology and the inferential techniques used for symmetric shapes. The most prominent type of symmetry in the organization of living organisms is bilateral symmetry. A 2D (or 3D) object is said to be bilaterally symmetric if its mirror image about some line or some plane is the same as the original form after relabeling some landmarks. This mirroring locus in general is called the midplane. In a perfect bilaterally symmetric shape there are two types of landmark. Some are paired, they don't lie on the midplane, but appear separately on left and right sides. Some other are unpaired and they lie on the midplane. Along with bilateral symmetry, we may mention other sorts of symmetry, like reflection symmetry with multiple axes (or planes) of symmetry, rotational symmetry, translational symmetry, and scaling symmetry. In the shape analysis of bilaterally symmetric structures, two categories of symmetry have been distinguished: *matching symmetry* and *object symmetry*. Object symmetry relates to the symmetry within a single object, such as a face, and hence it considers parts with internal left-right symmetry. In matching symmetry two separate structures exist as mirror images of each other, one on each body side (Klingenberg et al. 2002). Thus

it is concerned with symmetry between two corresponding objects, such as left and right hands (Mardia et al. 2000).

In other terms, matching symmetry concerns pairs of repeated structures that are separated from each other by a mirror plane. This mirror plane passes between the objects (outside of each). The two structures differ by a reflection and an appropriate translation. In presence of object symmetry, a single configuration is itself symmetric, as the reflection axis (or plane) passes through the configuration (e.g., the vertebrate skull). Object symmetry considers both the shape information from the left and right side, as in matching symmetry, and the additional information on the relative arrangement of the two connected halves (Savriama and Klingenberg 2006). In order to study matching symmetry, the landmark configurations from one side are reflected, then all the configurations are superimposed by GPA to produce an overall mean shape. Variations in the averages of the pairs of configurations embody the symmetric variation among individuals. The deviations of each configuration from the consensus provide an estimate of the asymmetry component.

For the analysis of object symmetry, the data set includes both the original landmark configurations and their reflected copies with the paired landmarks relabeled. A GPA is applied to all configurations to produce a single consensus, which is symmetric. The symmetric variation among individuals is measured from the averages of the original configuration and its reflected (appropriately relabelled) copy. Again the asymmetry is estimated by the deviations of each configuration from the consensus (Savriama and Klingenberg 2006). We refer to the isotropic case when the covariance matrix is a multiple of the identity; in the non-isotropic case, the covariance matrix of Procrustes coordinates can be any positive-semidefinite matrix with the appropriate null space. With reference to the inferential aspect, in order to test object symmetry in the isotropic case, we may conduct an analysis of variance (ANOVA) and then use F statistic test. Through the ANOVA test, the total sum of squares is decomposed into two terms: the square of what has been called *directional asymmetry* plus n times the so-called *fluctuating asymmetry* in the Procrustes metric.

Subtle asymmetries are small and completely random departures from bilateral symmetry. Fluctuating asymmetry is considered the most familiar of these asymmetries, providing a surprisingly convenient measure of developmental precision: the more precisely each side develops the greater the symmetry (Palmer and Strobeck 1997). It has become popular as a measure of environmental quality, stress, health, or fitness. Unfortunately because subtle asymmetries are often so small, they are exceedingly difficult to measure and analyze reliably. Conspicuous asymmetries are easily detected, either as asymmetrical structures on otherwise bilaterally symmetrical animals, or as whole-body asymmetries. They are classified as *antisymmetry* (also called random asymmetry) or *directional asymmetry*. Directional asymmetry (also called fixed asymmetry) arises either when one side is larger than the other on average, or the larger member of a bilateral pair tends to be on the same side. These two asymmetries provide information about the evolutionary history, suggesting how symmetry is broken during development (Palmer 1996).

In the non-isotropic case, we may use T^2 Hotelling's test and the approximation to Fisher's F distribution.

The same holds for matching symmetry. Obviously there is a difference in the degrees of freedom of the tests. In the isotropic case we may preform ANOVA test, while in the non-isotropic case we carry out Hotelling's T^2 . As usual, when the number of shape variables is greater than the most practical sample size, no formal T^2 can be computed and working under a permutation framework is recommended. In particular it is possible to use a permutation test for which the pivotal role of the Procrustes distance is retained but the distributional assumptions underlying the F under H_0 are relaxed. The reference distribution becomes a Monte Carlo permutation distribution where what is permuted is the assignment of one of the forms to the reflected state (Mardia et al. 2000).

Now we briefly present the nonparametric permutation solution to the problem of multivariate paired data observations. Let us assume that a q -dimensional non-degenerate real variable \mathbf{X} is observed in k different occasions, for instance at times (τ_1, \dots, τ_k) , on the same n statistical units considered in two different experimental situations, corresponding to two levels A_1 and A_2 of a treatment. Typically, observations at level A_1 correspond to baseline responses and those at level A_2 to after-treatment responses.

In shape analysis field, we usually deal with paired landmark data, hence the two levels A_1 and A_2 of a treatment correspond to the left and right coordinate of the same landmark point. We recall that, commonly, treatment effect is strictly related to the "stratification" variable, in particular it could be age or gender effect.

Coming back to the general case, the whole data set can be denoted by

$$\mathbf{X} = \{X_{hjit}, t = 1, \dots, k, i = 1, \dots, n, j = 1, 2, h = 1, \dots, q\}$$

For simplicity, let us assume that the response variables behave according to the following model:

$$X_{hjit} = \mu_h + \mu_{hit} + \delta_{ht} + \sigma_{ht}(\delta_{ht}) \cdot Z_{hjit},$$

$t = 1, \dots, k, i = 1, \dots, n, j = 1, 2, h = 1, \dots, q$, where μ_h represents a population constant for the h th variable, μ_{hit} represents a time effect on the h th variable at time t and specific to the i th individual; δ_{ht} represents treatment time effect at level j on the h th variable which, without loss of generality, are assumed to be $\delta_{h1t} = 0, \delta_{h2t} \leq$ (or \geq) $0, \forall (h, t)$; $\sigma_{ht}(\delta_{ht}) > 0$ represent population scale coefficients for variable h at time t , which are assumed to be invariant with respect to units but which may depend on treatment levels through effects δ_{ht} , provided that, when $\delta_{h2t} \neq 0$, stochastic dominance relationships $X_{h1t} \stackrel{d}{<} (or \stackrel{d}{>}) X_{h2t}, h = 1, \dots, q$, are satisfied; Z_{hjit} are q -variate random errors, which are assumed to be exchangeable with respect to treatment levels, independent with respect to units, with null mean vector, $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$, and with unspecified distribution $P \in \mathcal{P}$; in particular, these multivariate random errors may be dependent with respect to component variables and time through any kind of monotonic regression.

Here we assume that treatment effects are fixed or stochastic; in the present case, they are assumed to be independent of errors.

We are interested in testing for these effects irrespective of time, underlying dependences, and unknown distributions. Thus, the hypotheses under consideration are formalized as

$$H_0 : \left\{ \bigcap_{t=1}^k [\mathbf{X}_{1t} \stackrel{d}{=} \mathbf{X}_{2t}] \right\} = \left\{ \bigcap_{t=1}^k \bigcap_{h=1}^q [\delta_{h2t} - \delta_{h1t} = 0] \right\} = \left\{ \bigcap_{t=1}^k \bigcap_{h=1}^q H_{0ht} \right\}$$

against the alternatives of the form

$$H_1 : \left\{ \bigcup_{t=1}^k [\mathbf{X}_{1t} <\neq> \mathbf{X}_{2t}] \right\} = \left\{ \bigcup_{t=1}^k \bigcup_{h=1}^q H_{1ht} \right\},$$

in which at least one among H_{0ht} is not true. Observe that H_0 implies exchangeability of profile responses with respect to treatment levels, so that two q -dimensional profiles \mathbf{X}_{1t} and \mathbf{X}_{2t} , $t = 1, \dots, k$, are exchangeable within units. Note that in general case some of the sub-alternatives may be one-sided, or restricted, and others two-sided. Hence, we are operating in the context of multivariate restricted alternatives. Also note that for each component variable we assume that the differences of observations between the two treatment levels are informative. As a matter of fact, and assuming that the response model is adequate, differences behave as

$$Y_{hit} = X_{h2it} - X_{h1it} = \delta_{h-t} + \sigma_{ht}(\delta_{h2t}) \cdot Z_{h2it} - \sigma_{ht}(\delta_{h1t}) \cdot Z_{h1it},$$

$t = 1, \dots, k$, $i = 1, \dots, n$, $h = 1, \dots, q$, where it is shown that differences Y_{hit} depend only on treatment effects, exchangeable errors, and $\delta_{h-t} = \delta_{h2t} - \delta_{h1t} = \delta_{h2t}$. The two scale functions $\sigma_{ht}(\delta_{h2t})$ and $\sigma_{ht}(\delta_{h1t})$ are equal under H_0 . It is known that a direct solution to this testing problem, with restricted alternatives, is generally very difficult to obtain in a parametric framework, especially if the normality of P is not assumed and/or if the covariance matrix is unknown (Robertson et al. 1988). Conversely, nonparametric permutation methods can be applied even in the case of heterogeneous sample variance-covariance matrices, allowing for an almost exact solution to the multivariate Behrens-Fisher problem (Pesarin 1997).

Using NPC methods, this problem is processed in two phases. At the first stage kq partial permutation tests, each suitable for paired observations, are preformed. In the second stage, all the previously obtained tests are combined by means of NPC methodology.

Partial permutation tests have the form

$$T_{ht} = \phi_{ht}(\sum_i Y_{hit}), t = 1, \dots, k, h = 1, \dots, q,$$

where Y_{hit} are unit-by-unit and variable-by-variable observed differences; functions ϕ_{ht} correspond to the absolute value or to sign plus or minus according to whether

the h th sub-alternative H_{1ht} of interest is two-sided, \neq , or one-sided $>$ or $<$ respectively. It is worth noting that all these partial tests $\mathbf{T} = \{T_{ht}, t = 1, \dots, k, h = 1, \dots, q\}$ are marginally unbiased. Actually each sub-hypothesis H_{0ht} against H_{1ht} , being separately related to the h th component variable, may be considered as if it were univariate. As a consequence, the NPC methodology may be applied and, due to the assumed exchangeability in H_0 , the multivariate permutation distribution of \mathbf{T} is generated by the random attribution of individual data vectors to A_1 and A_2 . This implies that permutations are within individuals and with respect to treatment levels A_1 and A_2 . Thus, there are two permutations for each individual, and the cardinality of the permutation sample space $\mathcal{X}_{/\mathbf{Y}}$, where $\mathbf{Y} = \{Y_{hit}, t = 1, \dots, k, i = 1, \dots, n, h = 1, \dots, q\}$, is 2^n . In practice, for each unit, the permutation approach considers an equally likely random choice of the sign to be attributed to the vector of differences $\{Y_{hit}, t = 1, \dots, k, i = 1, \dots, n, h = 1, \dots, q\}$. Note that the random signs are invariant within units with respect to h and t and are independent with respect to units. This guarantees that the dependence relations within variables are preserved. When $k = q = 1$ and responses are homoscedastic and normally distributed, this testing problem becomes the classical Student t for paired data. Moreover, multivariate paired observations testing problems may be viewed as multivariate testing of symmetry. Hence in shape analysis, we first consider all the differences between right and left coordinates of each landmark point. Then, once obtained partial p -values for the coordinates (coordinate level), we combine these p -values in order to obtain information on landmarks (landmark level). Finally we consider domains and aspects, if present, as well as the global combination of partial p -values.

References

- Brombin C (2009) A nonparametric permutation approach to statistical shape analysis, ph.D. thesis. Padova, Italy: University of Padova
- Brombin C, Salmasso L, Villanova C (2009) 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
- Frost SR, Marcus LF, Bookstein FL, Reddy DP, Delson E (2003) Cranial allometry, phylogeography, and systematics of large-bodied papionins (primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. *The Anatomical Record (Part A)* 275A:1048–1072
- 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
- Klingenberg CP, McIntyre GS (1998) Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with procrustes methods. *Evolution* 52:1363–1375
- Klingenberg CP, Barluenga M, Meyer A (2002) Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution* 56:1909–1920
- Mardia KV, Bookstein FL, Moreton IJ (2000) Statistical assessment of bilateral symmetry of shapes. *Biometrika* 87:285–300

- Palmer AR (1996) From symmetry to asymmetry: Phylogenetic patterns of asymmetry variation in animals and their evolutionary significance. *Proceedings of the National Academy of Sciences (USA)* 93:14,279–14,286
- Palmer AR, Strobeck C (1997) Fluctuating asymmetry and developmental stability: Heritability of observable variation vs. heritability of inferred cause. *Journal of Evolutionary Biology* 10:39–49
- Pesarin F (1997) An almost exact solution for the multivariate behrens-fisher problem. *Metron* 55(3–4):85–100
- 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
- Robertson T, Wright FT, Dykstra RL (1988) *Order Restricted Statistical Inference*. Wiley, Chichester
- Rohlf FJ (2000) Statistical power comparisons among alternative morphometric methods. *American Journal of Physical Anthropology* 111:463–478
- Rohlf FJ (2008) TPSTri, Explore shapes of triangles, version 1.25. Department of Ecology and Evolution, State University of New York at Stony Brook
- Savriama Y, Klingenberg CP (2006) Geometric morphometrics of complex symmetric structures: Shape analysis of symmetry and asymmetry with procrustes methods. In: Barber S, Baxter PD, Mardia KV, Walls RE (eds) *Interdisciplinary Statistics and Bioinformatics*, Leeds University Press, pp 158–161, ISBN 0-85316-252-2

Chapter 4

Power Behavior of Permutation Tests with High Dimensional Data

“The statistical method is more than an array of techniques. The statistical method is a Mode of Thought; it is Sharpened Thinking; it is Power”.

Paper presented at meeting of the International Statistical Institute, September 1953.

W.E. Deming

Abstract In this chapter we present main results of an extensive simulation study aiming to evaluate the power of multivariate permutation tests combination-based. In particular we show that, for a given and fixed number of subjects, when the number of variables and the associated noncentrality parameter both diverge, then the power of multivariate NPC tests converges to one. This holds true even when the number of variables is larger than the permutation sample space. These results allow us to introduce and then extend to shape analysis the notion of *finite-sample consistency*.

Keywords Informative variables • Multivariate permutation tests • Permutation sample space • Power function

4.1 High Dimensional Data with Small Sample Sizes

In statistical shape analysis, like in several application fields, e.g., longitudinal analysis (Diggle et al. 1994), analysis of microarrays and genomics (Salmaso and Solari, 2005, 2006), analysis of brain images (Friman and Westin 2005), and functional data analysis (Ferraty and Vieu 2006), it may happen that the sample sizes are fixed and the number of observed variables is much larger than sample sizes.

As already said, in the parametric framework, the most natural way to compare two mean shapes is by using the Hotelling's T^2 test. Despite its widespread use, this test presumes independent samples, that the shape coordinates (e.g., Kendall tangent space, Bookstein shape coordinates, or Rao and Suryawanshi shape variables, 1996; 1998) follow a multivariate normal distribution, and that the samples are drawn from populations with the same covariance matrix. These assumptions could be very demanding (Dryden and Mardia 1998; Blair et al. 1994). As an example, in T^2 -test using Rao and Suryawanshi shape variables (1996), the requirement of the equality of variances can be a problem since the expected covariance matrix depends upon the mean shape. In addition, when the number of landmarks k is larger than 3, larger sample sizes are required than for the other tests since the dimension of shape space is larger than for the other methods (Rohlf 2000).

Therefore, traditional statistical analysis tools designed for Euclidean spaces have to be reformulated (Terriberry et al. 2005).

Alternative inferential procedures are those based on a permutation approach.

In Chap. 2, we have introduced an Hotelling's T^2 permutation counterpart within the nonparametric combination approach. Simulation results displayed in Tables 2.1–2.3 show that the power for the suggested test statistic increases when increasing the number of the processed variables or the noncentrality parameter δ , even when the number of variables is larger than the cardinality of permutation sample space \mathcal{X}/\mathbf{X} . In light of these results, we have performed a simulation study to evaluate the power of multivariate permutation tests combination-based (Pesarin 2001; Pesarin and Salmaso 2010). In the next section we show that, for a given and fixed number of subjects, when the number of variables and the associated noncentrality parameter both diverge, then the power of multivariate permutation tests based on nonparametric combining functions converges to one. This is still true in the case in which the number of variables is larger than the permutation sample space (Brombin 2009; Brombin et al. 2008; Alfieri et al. 2012).

These results allow us to introduce and then extend to shape analysis the notion of *finite-sample consistency* for NPC tests, illustrated in Chap. 5. Hence, it is possible to obtain powerful tests in a nonparametric framework by increasing the number of informative variables while the number of cases is held fixed. As a result, in shape analysis, even in presence of few available specimens, many informative landmarks and semilandmarks coordinates may be allocated, thus allowing for a good accuracy in the description of the shapes.

4.2 Simulation Study and Results

We start by presenting the problem in a general case. Extensions to the specific field of shape analysis are straightforward. Let us consider the two independent sample case and assume that the response variables behave according to the following multivariate model:

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

$i = 1, \dots, n_j, j = 1, 2, h = 1, \dots, k$, where 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 $\delta_{h1}=0, \delta_{h2} \leq (or \geq) 0$, and Z_{hji} are random errors assumed to be exchangeable with respect to treatment levels, independent with respect to units, with zero mean vector ($\mathbb{E}(\mathbf{Z}) = 0$), and finite second moments. Hence h , in a shape analysis framework, would indicate the 2D or 3D landmark coordinates. Let $\bar{X}_{hj}, j = 1, 2$, be the sample mean for the h -th variable. The symbol $*$, if present, denotes in this case a permutation of the original data.

Different multivariate distributions have been considered: normal $\mathcal{N}(0, 1)$, Cauchy $\mathcal{C}y(0, 1)$, Student's $\mathcal{S}t(2)$ with 2 d.f. and Pareto $\mathcal{P}a(1, 1)$ distributions. We note that, because of the chosen parameters, here we deal with "particular" distributions. Actually Cauchy $\mathcal{C}y(0, 1)$ has no mean and infinite variance, Student t with 2 d.f. has finite mean and infinite variance, and Pareto $\mathcal{P}a(1, 1)$ has infinite mean and infinite variance.

The notion of unconditional finite sample consistency, defined for divergent fixed effects δ , is different from the common notion of (unconditional) consistency of a test, which considers the behavior of rejection rate for given δ when $\min(n_1, n_2)$ diverges. It is known that in order to attain permutation (unconditional) consistency it is required that random deviates Z have finite second moment (Lehmann 1986; Romano 1990). Here we only require measurability, so that random deviates Z are not required to be provided with finite moments of integer order ≥ 1 .

We focus on the two independent sample case, in the particular case in which only three specimens are available in each group. Once more we are interested in investigating what happens to the power when the number of variables is larger than the cardinality of the permutation sample space. We recall that with $n_1 = n_2 = 3$ the cardinality of the permutation sample space is given by $\binom{6}{3} = 20$, hence we wish to analyze the power behavior using all the possible permutations and recalling that in this case the minimum attainable α -level is $1/20$ for two-sided tests. Figure 4.1 shows simulation results in the directional case, when the underlying distribution is multivariate normal. In the case of nondirectional (two-sided) alternatives we consider the test statistic given by $T^* = \sum_h (\bar{X}_{h1}^* - \bar{X}_{h2}^*)^2$, where $\bar{X}_{hj}^*, j = 1, 2$ are permutation sample means, while in the presence of directional alternatives we simply consider the permutationally equivalent test statistic given by $T^* = \sum_h (\bar{X}_{h2}^*)$. Focussing on the directional case

- in Fig. 4.2 are shown simulation results when the underlying distribution is the multivariate Cauchy $\mathcal{C}y(0, 1)$,
- in Fig. 4.3 are displayed simulation results when data are distributed according to a Student t with 2 d.f., denoted by $\mathcal{S}t(2)$,
- in Fig. 4.4 are shown simulation results when data follows a Pareto distribution $\mathcal{P}a(1, 1)$, with a fixed parameter of shape equal to 1.

Complete simulation study has been published elsewhere in Brombin et al. (2008). Both in the nondirectional, not shown here, and directional cases, under the multivariate normal distribution, the power increases when increasing the number of the processed variables or the value of the noncentrality parameter δ . We also

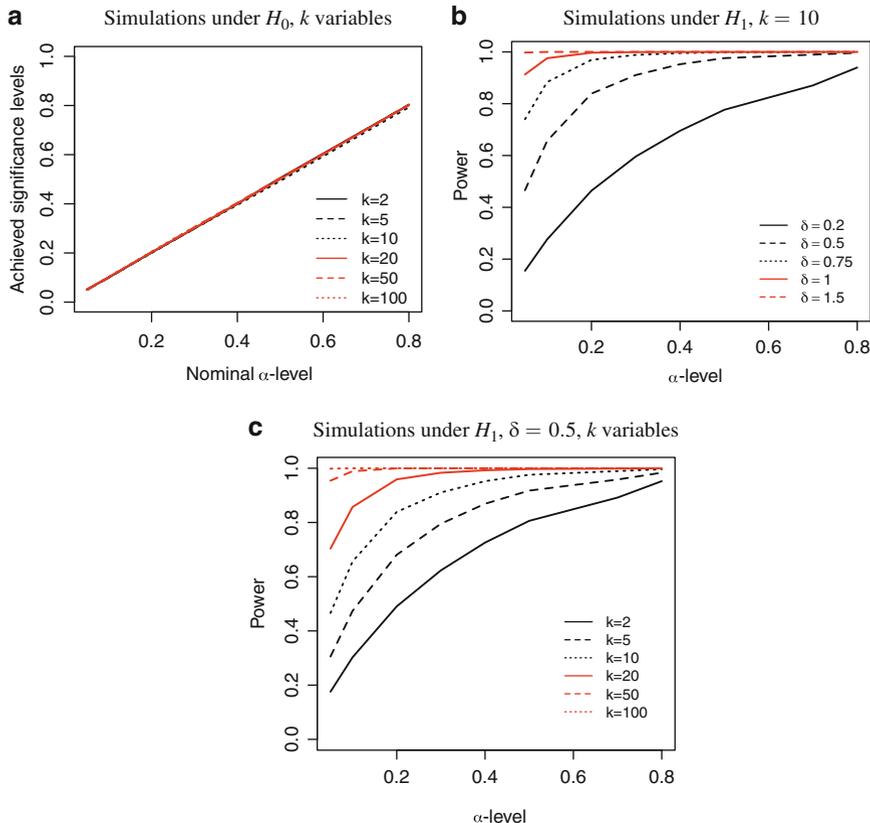


Fig. 4.1 Simulation results under Multivariate Normal Distribution $\mathcal{N}(0, 1)$, directional alternatives (a)–(c)

have found that, under the Cauchy distribution, the power holds steady increasing the number of covariates, increases when increasing the standardized non-centrality parameter (producing evidence for its consistency) and converges to 1 when diverging the non-centrality parameter. Under the $St(2)$ distribution, it is possible to show that the test is consistent even if the data distribution does not admit finite variance. If we consider a Pareto distribution $\mathcal{Pa}(1, 1)$, simulation results show that power could not increase when increasing the number of covariates but increases when increasing the value of standardized δ , converging to 1 when δ diverges.

As said before, Cauchy $\mathcal{Cy}(0, 1)$ has no mean and infinite variance and Pareto $\mathcal{Pa}(1, 1)$ has infinite mean and infinite variance. It is to be emphasized that for fixed (n_1, n_2) , with random deviates distributed according to either Cauchy $\mathcal{Cy}(0, \sigma)$ or Pareto $\mathcal{Pa}(\theta, \sigma)$, the latter with shape parameter $0 < \theta \leq 1$, when δ is fixed and the number of variables diverges both are not consistent, because in this case the law of large numbers does not apply.

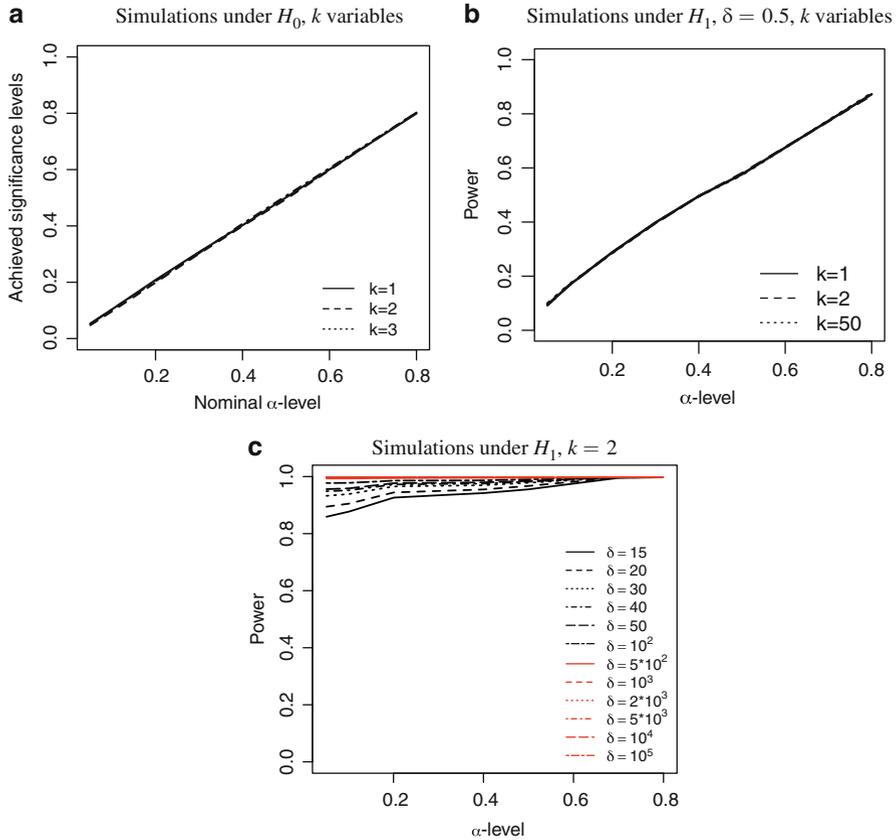


Fig. 4.2 Simulation results under Cauchy $\mathcal{C}_Y(0, 1)$ distribution, directional alternatives (a)–(c)

4.3 Final Remarks

Very often, researchers do not have access to large samples. More likely they have to analyze large data sets with few cases. Asymptotic theory cannot be applied. Moreover, due to limited resources or limited size of the population under study, they cannot increase sample size.

Standard parametric approaches assume larger samples than they could ever obtain in practice (Hoyle 1999). Moreover, assumptions underlying classic parametric tests, e.g., normality and homoscedasticity, are rarely met when analyzing real data. Departures from these assumptions can critically reduce the power of standard tests and regression methods. Adequate sample size is another of the assumptions underlying parametric tests. Whenever assumptions underlying parametric approaches are not satisfied, statistical methodologies that offer the flexibility of large sample strategies without requiring prohibitively large samples should be applied. Modern

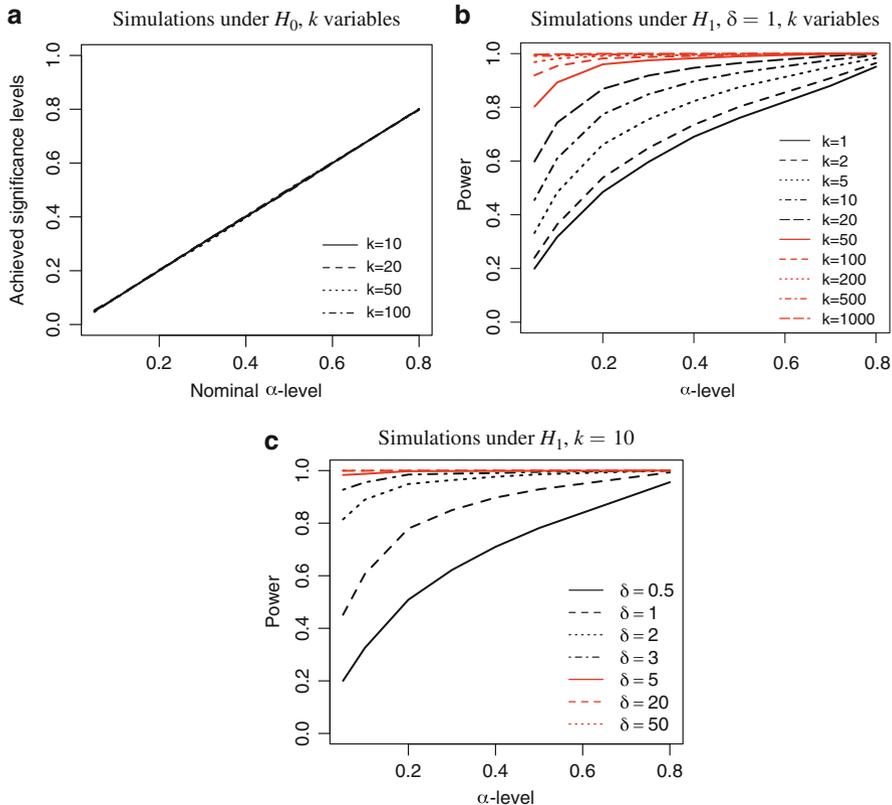


Fig. 4.3 Simulation results under Student's $\mathcal{S}(2)$ distribution, directional alternatives (a)–(c)

robust statistical methodologies may represent a valid alternative approach to classic parametric methods (Erceg-Hurn and Mirosevich 2008).

Through the simulation study, we have showed that power of multivariate permutation tests, derived from NPC methodology (Pesarin 2001; Pesarin and Salmaso 2010; Brombin et al. 2008), increases when the number of processed variables increases, provided that the noncentrality parameter δ increases, even when the number of variables is larger than the permutation sample space. These findings allow to conclude that it is possible to obtain powerful tests in a nonparametric framework by increasing the number of informative variables while leaving the number of statistical units fixed. Hence the suggested nonparametric tests provide efficient and robust solutions in multivariate small-sample-size problems.

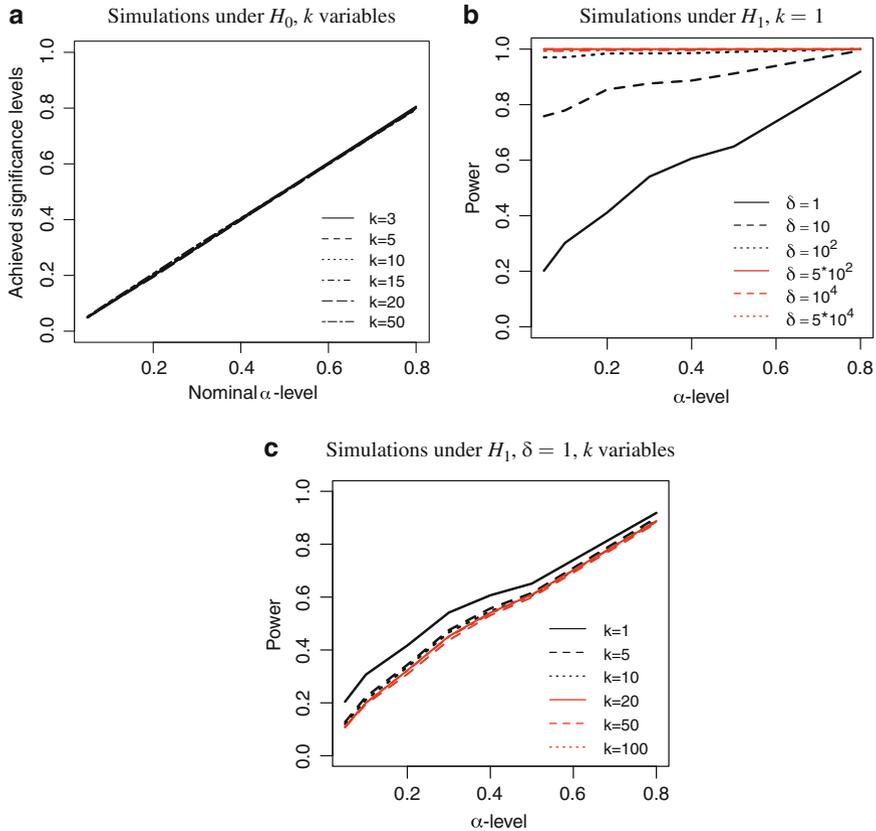


Fig. 4.4 Simulation results under Pareto $\mathcal{P}a(1, 1)$ distribution, directional alternatives (a)–(c)

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
- 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
- Brombin C (2009) A nonparametric permutation approach to statistical shape analysis, ph.D. thesis. Padova, Italy: University of Padova
- 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
- Diggle PJ, Liang KY, Zeger SL (1994) *Analysis of Longitudinal Data*. Oxford University Press, New York

- Dryden IL, Mardia KV (1998) *Statistical Shape Analysis*. John Wiley & Sons, London
- Erceg-Hurn DM, Mirosevich VM (2008) Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist* 63:591–601
- Ferraty F, Vieu P (2006) *NonParametric Functional Data Analysis: Theory and Practice*. Springer, New York
- Friman O, Westin CF (2005) Resampling fMRI time series. *NeuroImage* 25:859–867
- Hoyle RH (1999) *Statistical Strategies for Small Sample Research*. SAGE Publications, Inc
- Lehmann EL (1986) *Testing Statistical Hypotheses*. Wiley, New York
- 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
- Rao C, Suryawanshi S (1996) Statistical analysis of shape of objects based on landmark data. *Proceedings of the National Academy of Sciences of the United States of America* 93:12,132–12,136
- Rao C, Suryawanshi S (1998) Statistical analysis of shape through triangulation of landmarks: a study of sexual dimorphism in hominids. *Proceedings of the National Academy of Sciences of the United States of America* 95:4121–4125
- Rohlf FJ (2000) Statistical power comparisons among alternative morphometric methods. *American Journal of Physical Anthropology* 111:463–478
- Romano JP (1990) On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association* 85:686–692
- Salmaso L, Solari A (2005) Multiple Aspect Testing for case-control designs. *Metrika* 12:1–10
- Salmaso L, Solari A (2006) Nonparametric iterated combined tests for genetic differentiation. *Computational Statistics & Data Analysis* 50:1105–1112
- 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

Chapter 5

Finite-Sample Consistency of Combination-Based Tests in Shape Analysis

“Most of the fundamental ideas of science are essentially simple, and may, as a rule, be expressed in a language comprehensible to everyone.”

The Evolution of Physics (1938)

Albert Einstein

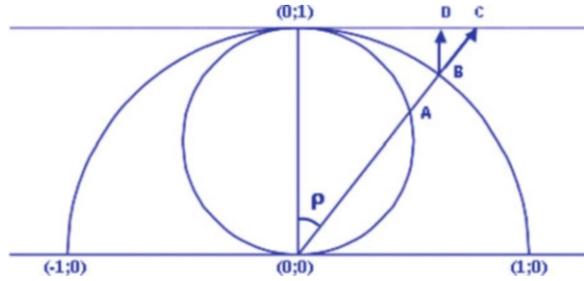
Abstract In this chapter we extend and discuss the notion of finite-sample consistency for permutation tests combination-based to the shape analysis field. Sufficient conditions are given in order that the rejection rate converges to one, for fixed sample sizes at any attainable α -value, when the number of variables diverges, provided that the noncentrality induced by test statistics also diverges. On the basis of these findings, we emphasize that the proposed tests provide efficient solutions to multivariate small sample problems, like those encountered in the shape analysis field. We illustrate finite-sample consistency property of NPC tests by means of a toy example.

Keywords Conditional and unconditional finite sample consistency • Convergence in probability • Tangent space • Unconditional consistency

5.1 How to Obtain a Tangent Space: A Brief Overview

Goodall's F test and Hotelling's T^2 test using Kendall tangent space coordinates are both based on using a generalized least-squares Procrustes analysis (or GLS superimposition, also called GPA) to compute the average shape for the entire dataset. Each specimen is then fit to this overall mean. Rohlf (1999) shows that triangles (i.e., shapes described by three landmarks in 2D) corresponding to these aligned specimens lie on the surface of a unit hemisphere of the same dimensionality

Fig. 5.1 Kendall's shape space for *triangles* (redrawn from Rohlf 1999). See also, for further details on shape spaces, Rohlf (2000), Slice (2001), Zelditch et al. (2012)



as Kendall's shape space. When shape variation is small, the distribution of points on this hemisphere can be satisfactorily approximated by an orthogonal projection onto an Euclidean tangent plane (if the overall mean is used as the point of tangency). Kent (1994) calls this space Kendall tangent space but Dryden and Mardia (1992) call it Kent tangent space. In Fig. 5.1, redrawn from Rohlf (1999), it is shown a diagram of a cross-section of a construction of Kendall's shape space for triangles (circle with a radius of $1/2$), hemisphere of pre-shapes aligned to the reference (half circle with a radius of 1), and tangent space (tangent line). Procrustes distance is the angle ρ in radians. Point A represents the position of a shape in Kendall's shape space and B is the corresponding position in the hemisphere (yielding Procrustes tangent space coordinates). Point C is the stereographic projection of point A onto the tangent space (yielding stereographic shape coordinates) and D is the orthogonal projection of point B onto tangent space (yielding Kendall tangent space coordinates).

Statistical methods are required to take into account the non-Euclidean geometry of Kendall's shape space for both two- and three-dimensional landmarks. In case of small shape variation it is possible to make a good linear approximation to the space and then use standard multivariate methods (Kent 1994). The resulting space is of the same dimensionality as the shape space and may be viewed as tangent to it.

The point of tangency corresponds to the reference shape (usually taken as an average shape). The projections of the points corresponding to the observed shapes are used for subsequent statistical analyses.

Spaces tangent to Kendall's shape space have been constructed in two rather different ways in morphometrics. A stereographic projection has been used to map points from the surface of the shape space sphere to a tangent space. Stereographic projection is a standard tool for mapping points on the complex plane into a one-to-one correspondence with points on a sphere. The projection is the intersection of tangent space with a line that goes from the point antipodal to the reference through the point being projected (point C). The coordinates of stereographic projections are called shape coordinates. Shapes close to the reference will map to points close to the origin, and the point antipodal to the reference maps to infinity.

A tangent space may also be constructed by a projection of the hemisphere of pre-shapes aligned with respect to the average shape onto the space perpendicular

to the direction corresponding to the reference. In particular Kendall tangent space coordinates are based on Procrustes tangent space coordinates and correspond to point D in Fig. 5.1.

5.2 Main Theorems and General Characterization of Finite Sample Consistency

We will show that, under mild conditions, the power function of permutation tests based on associative statistics monotonically increases when increasing the number of the processed variables, provided that the induced noncentrality parameter δ increases, even when the number of variables is larger than the permutation sample space. In particular, for any added variable the power does not decrease if each variable makes larger noncentrality. Specifically, we will show that, for a given and fixed number of subjects, when the number of variables k (typically in shape analysis we handle $h = 1, \dots, km$ variables, describing k landmarks in m dimensions) and the associated noncentrality parameter δ both diverge, then the power of multivariate combination-based permutation tests converges to one. These results confirm and extend those presented by Blair et al. (1994), allowing us to introduce the notion of “finite-sample consistency” for combination-based permutation tests (Pesarin 2001; Pesarin and Salmaso 2010; Brombin 2009). Sufficient conditions are given in order that the rejection rate converges to one, for fixed sample sizes at any attainable α -value, when the number of variables diverges, provided that the noncentrality induced by test statistics also diverges.

Such findings look very relevant to solve multivariate small sample problems (like those encountered in shape analysis field) since they demonstrate that it is possible to obtain powerful tests in a nonparametric framework by increasing the number of informative variables while the number of cases is held fixed.

Definition The *configuration* is the set of landmarks on a particular object. The *configuration matrix* X is the $k \times m$ matrix of Cartesian coordinates of the k landmark in m dimensions. The configuration space is the space of all possible landmark coordinates.

Usually in the applications there are $k \geq 3$ landmarks in $m = 2$ or $m = 3$ dimensions and the configuration space is typically \mathbb{R}^{km} .

Assume to have $n = n_1 + n_2$ individuals and consider two independent random samples of configurations X_1, \dots, X_{n_1} and Y_1, \dots, Y_{n_2} from independent populations with mean shapes $[\mu_1]$ and $[\mu_2]$.

Let v_1, \dots, v_{n_1} and w_1, \dots, w_{n_2} be the coordinates (each a $k \times m$ matrix) of aligned specimens obtained through a generalized Procrustes analysis (GPA). We remind the reader that this procedure is performed to estimate a mean shape and to align the specimens to it. These aligned specimens are then used for the computation of their tangent space projections and possibly partial warp scores that are useful for subsequent statistical analyses.

As shown in Fig. 5.1 the space to visualize aligned specimens is the surface of a hyper-hemisphere (a hemisphere if $k = 3$) with the mean shape corresponding to its pole. This space has a radius of 1. Kendall tangent space represents an orthogonal projection of this space onto a plane tangent to its pole. For triangles this tangent space is a unit disk. We refer the reader to Rohlf (1999) for more details about these relationships.

We start recalling the notation used in Chap. 2 (Sect. 2.3), in the case in which GPA superimposition step is included. Let V be the $n_1 \times (k \times m)$ matrix of aligned specimens in tangent space in the first group. Similarly W is the $n_2 \times (k \times m)$ matrix of aligned specimens in tangent space, representing subjects belonging to the second group. Finally we define $U = \begin{pmatrix} V \\ W \end{pmatrix}$ the $n \times (k \times m)$ matrix of aligned specimens in tangent space, i.e., our data set, where $n = n_1 + n_2$.

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

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

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

With reference to the scale coefficients σ_h , we observe that these parameters could be very useful since they reflect the “intrinsic” biases in the registration of landmarks. Actually there are landmark points readily available, hence easier to be captured than others by the operator or machine. As a consequence, they are less variable in their location. Hence landmark coordinates in the first group differ from

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

For the sake of simplicity, we will consider testing problems for stochastic dominance alternatives, generated by treatments with nonnegative random shift effects δ even if the focus on stochastic dominance alternatives could be questioned since in shape analysis it is more likely to formulate landmark specific alternatives, thus considering a proper set of side assumption specific to the problem, instead of assuming the dominance of a shape onto another in all the points. Anyway the extension to specific different sets of alternatives for different groups (or domains) of landmarks as well as extensions to negative random effects and two-sided alternatives are straightforward. Therefore the specific hypotheses may be expressed as

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

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

Without loss of generality we can model the data set as $U(\delta) = \{Z_1 + \delta, Z_2\}$, where $Z = (Z_1, Z_2)$ have the role of random deviates whose distribution is generally unknown.

We start considering associative partial test statistics, defined as

$$T_h^*(\delta) = \frac{1}{n_1} \sum_{i=1}^{n_1} \varphi_h [U_{h1i}^*(\delta)] - \frac{1}{n_2} \sum_{i=n_1+1}^n \varphi_h [U_{h2i}^*(\delta)],$$

where φ_h is any non-degenerate measurable non-decreasing function of the data, potentially dependent on the variable under study. For example, T_h^* may correspond to the comparison of sampling means, carried out coordinate by coordinate.

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 nonparametric combination are obviously the same of those presented in Chap. 2 (Sect. 2.3). Hence all permutation partial tests are marginally unbiased, consistent, and significant for large values.

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

Again, we refer the reader to Chap. 2 (Sect. 2.3) for details regarding the application of NPC methodology in shape analysis context.

As already said, we consider the global test T'' obtained after combining at the first stage with respect to m and then with respect to k (of course, the sequence may be reversed). For the sake of simplicity, we may assume to use associative partial tests

and direct combining function. In particular, because of the use of direct combining functions, if we reverse the sequence, thus combining at first with respect to k and then with respect to m , we obtain exactly the same result.

As mentioned above, in [Pesarin \(2001\)](#), [Pesarin and Salmaso \(2010\)](#) it is proved that if at least one partial permutation test $T_h, h = 1, \dots, km$ is weakly consistent for H_{0h} against H_{1h} respectively, then $T'' = \psi\{\lambda_1, \dots, \lambda_{km}\}, \forall \psi$ in the class of combining functions \mathcal{C} , is weakly consistent combined test for

$$H_0 : \bigcap_{h=1}^{km} \{H_{0h}\} \quad \text{vs.} \quad H_1 : \bigcup_h^{km} \{H_{1h}\}.$$

In virtue of this theorem, we will focus our attention to the unidimensional case, i.e., $h = 1$ and $T_1 = T$, since if we are able to prove that at least one partial permutation test is weakly consistent for H_{0h} against H_{1h} , then we can state that the global test T'' , obtained after combining with respect to the k landmarks and the m dimensions, is weakly consistent too.

Afterwards, we will indicate the n -dimensional sample space \mathcal{U} with \mathcal{U}^n and with \mathcal{U}_U^n the conditional reference space associated with U , containing all the permutations of U .

At first we will study the behavior of conditional (permutation) rejection rate when sample sizes (n_1, n_2) and non-degenerate random deviates $Z = (Z_1, Z_2)$ are held fixed, while the fixed effect δ goes to the infinity, according to whatever monotonic sequence $\{\delta_v, v \geq 1\}$. Then we examine the unconditional (population) rejection rate when i.i.d. random deviates Z do vary in the sample space \mathcal{X}^n according to the n -dimensional distribution P_Z . The extension from fixed to random effects will be presented too.

We limited our attention to the notion of weak consistency (or in probability), i.e., for divergent values of non-centrality parameter induced by the test statistic, the rejection probability of test is of one for any fixed $\alpha > 0$. The almost sure version (strong or with probability one), although of great mathematical importance, in the permutation context presents a limited relevance.

In [Pesarin \(2001\)](#), [Pesarin and Salmaso \(2010\)](#), and [Hoeffding \(1952\)](#) it is stated that conditional and unconditional power functions of any associative test statistics both do not decrease as the effect increases. Similar behavior is true also for random effects Δ .

Let us start with a Lemma concerning the conditional finite sample consistency of T , a stepping stone to the results presented afterwards.

Lemma 5.2.1. *Suppose that:*

- (i) T is any associative test statistic for one-sided hypotheses;
- (ii) sample sizes (n_1, n_2) and the set of real deviates $Z = \{Z_1, Z_2\} \in \mathcal{U}^n$ are fixed;
- (iii) the data set is $U(\delta) = (Z_1 + \delta, Z_2)$, where $(Z_1, Z_2) \in \mathcal{U}^n$ are i.i.d. measurable real random deviates and δ is the vector of nonnegative fixed effects;

(iv) fixed effects δ diverge to the infinity according to whatever monotonic sequence $\{\delta_v, v \geq 1\}$, the elements of which are such that $\delta_v \leq \delta_{v'}$ for any pair $v < v'$.

If conditions (i)–(iv) are satisfied, then the permutation (conditional) rejection rate of T converges to 1 for all α -values not smaller than the minimum attainable α_a ; thus, T is conditional finite-sample consistent.

Proof. For any chosen $\delta > 0$, let us consider the observed data set $U(\delta) = (Z_1 + \delta, Z_2)$, where δ represents the vector of effects corresponding to the given set of deviates Z . We indicate with $\mathcal{T}_{U(\delta)}$ the permutation support induced by the test statistic T when applied to the data set $U(\delta)$, i.e., $\mathcal{T}_{U(\delta)} : \{T^*(\delta) = T(U^*(\delta)) : U^*(\delta) \in \mathcal{U}_{U(\delta)}^n\}$. It is possible to find a value δ_z of δ such that the related observed value is right-extremal for the induced permutation support, i.e.,

$$T^o(U(\delta_z)) = \max_{\mathcal{T}_{U(\delta_z)}} \{T^*(\delta_z) : U^*(\delta_z) \in \mathcal{U}_{U(\delta_z)}^n\}.$$

This value δ_z can be determined by observing that a sufficient condition for right-extremal property of T^o is that $\min_{n_1}(Z_{1i} + \delta_z) > \max_{n_2}(Z_{2i})$. Due to the non-decreasing monotonicity of φ_h , we can also write

$$\frac{1}{n_1} \sum_i \varphi_h(Z_{1i} + \delta_z) > \frac{1}{n_2} \sum_i \varphi_h(Z_{2i}).$$

In this way, $T^o(\mathbf{U}(\delta_z))$ is right-extremal because, for all permutations $\mathbf{U}^*(\delta_z) \neq \mathbf{U}(\delta_z)$, $T^o(\mathbf{X}^*(\delta_z)) < T^o(\mathbf{X}(\delta_z))$.

The rejection rate relative to the minimum attainable α -value α_a (that is equal to $1/\binom{n}{n_1}$ for one-sided alternatives and to $2/\binom{n}{n_1}$ for two-sided alternatives), in force of monotonic behavior with respect to δ , reaches 1 for all $\delta > \delta_z$, since $T(U^*(\delta_z)) < T^o(U(\delta_z))$ for all permutations $U^*(\delta_z) \in \mathcal{U}_{U(\delta_z)}^n$ such that $U^*(\delta_z) \neq U(\delta_z)$. Hence, due to the monotonicity property with respect to α , it is of 1 also $\forall \alpha > \alpha_a$.

The conditional power function of T , denoted by

$$\Pr\{\lambda(U(\delta)) \leq \alpha | \mathcal{U}_{U(\delta)}^n\},$$

is of 1 for all $\delta \geq \delta_z$ and $\alpha \geq \alpha_a$, thus has 1 as a limit. ■

Theorem 5.2.1. *Suppose that:*

- (i) T is any associative test statistic for one-sided hypotheses;
- (ii) sample sizes (n_1, n_2) are fixed and finite;
- (iii) the data set is $U(\delta) = (Z_1 + \delta, Z_2)$, where $(Z_1, Z_2) \in \mathcal{U}^n$ are i.i.d. measurable real random deviates and δ is the vector of nonnegative fixed effects;
- (iv) fixed effects δ diverge to the infinity according to the monotonic sequence $\{\delta_v, v \geq 1\}$ as in Lemma 5.2.1.

If conditions (i) to (iv) are satisfied, then the permutation unconditional rejection rate of test T converges to 1 for all α -values not smaller than the minimum attainable α_a ; thus, T is weak unconditional finite-sample consistent.

Proof. We indicate with $P_Z(z) = \Pr\{Z \leq z\}$ the distribution of vector Z . Since of random deviates Z are required to be provided with measurability, we get that $\lim_{z \downarrow -\infty} \Pr(Z \leq z) = 0$ and $\lim_{z \uparrow +\infty} \Pr(Z \leq z) = 1$.

According to the Lemma 5.2.1, a sufficient condition for the observed value $T^o(U(\delta))$ being right-extremal in the induced permutation support $\mathcal{T}_{U(\delta)}$ is that $\min_{n_1}(Z_{1i} + \delta) > \max_{n_2}(Z_{2i})$. The unconditional probability of this event, as random deviates Z are i.i.d., is given by

$$\Pr\left\{\min_{n_1}(Z_{1i} + \delta) > \max_{n_2}(Z_{2i})\right\} = \int_{\mathcal{U}} \{[1 - P_Z(t - \delta)]^{n_1}\} d[P_Z(t)]^{n_2}.$$

The limit of this probability, as δ tends to infinity, is equal to 1, since (n_1, n_2) are fixed and finite and, in force of the Lebesgue's monotone convergence theorem, the associated sequence of probability measures $\{P_Z(t - \delta_v), v \geq 1\}$ converges to zero monotonically for any t .

Hence the probability of finding a set $Z \in \mathcal{U}^n$ for which there does not exist a finite value δ_Z such that $\min_{n_1}(Z_{1i} + \delta_Z) > \max_{n_2}(Z_{2i})$ converges to zero monotonically as δ diverges. This, by taking also account of Lemma 5.2.1, implies that the unconditional rejection rate

$$W_\alpha(\delta) = \int_{\mathcal{U}} \Pr\{\lambda(U(\delta)) \leq \alpha | \mathcal{U}_{U(\delta)}^n\} dP_Z(z),$$

converges to 1 for all $\alpha \geq \alpha_a$, as δ tends to the infinity. ■

Theorem 5.2.2. *Suppose that random deviates Z and effects δ are such that:*

- (i) *there exists a function $\rho(\delta) > 0$ of effects δ the limit of which is 0 as δ goes to the infinity;*
- (ii) *T is any associative test statistic for one-sided hypotheses;*
- (iii) *the data set is obtained by considering the transformation $Y(\delta) = \rho(\delta)U(\delta)$;*
- (iv) *$\lim_{\delta \uparrow \infty} \delta \rho(\delta) = \tilde{\delta} > 0$, and $\lim_{\delta \uparrow \infty} \Pr\{\rho(\delta) \cdot |Z| > \varepsilon\} = 0, \forall \varepsilon > 0$;*
- (v) *and further conditions are the same as in Theorem 5.2.1.*

If conditions (i)–(v) hold then the unconditional rejection rate converges to 1 for all α -values not smaller than the minimum attainable α_a ; thus, T is weak unconditional finite-sample consistent.

Proof. At first, we remark that data $Y(\delta) = \rho(\delta)[Z_1 + \delta, Z_2]$, as δ goes to the infinity, collapse in distribution toward $[\tilde{\delta}, 0]$. Then, for any fixed set of random deviates (Z_1, Z_2) , $T(Y(\delta))$ is right extreme in the induced permutation support when $\min_{n_1}[(Z_{1i} + \delta)\rho(\delta)] > \max_{n_2}[Z_{2i}\rho(\delta)]$.

We also notice that $\rho(\delta)$ is positive, hence the event defined by this relation is equivalent to $\min_{n_1}[Z_{1i} + \delta] > \max_{n_2}[Z_{2i}]$, in the sense that the latter is true if and only if the former is true. From proof of Theorem 5.2.1, we have that

$$\begin{aligned} \Pr \left\{ \min_{n_1} [(Z_{1i} + \delta)\rho(\delta)] > \max_{n_2} [Z_{2i}\rho(\delta)] \right\} &= \Pr \left\{ \min_{n_1} (Z_{1i} + \delta) > \max_{n_2} (Z_{2i}) \right\} \\ &= \int_{\mathcal{Q}} \{ [1 - P_Z[t - \delta]^{n_1}] \} d [P_Z(t)]^{n_2}. \end{aligned}$$

The limit of this probability, as δ tends to infinity, is equal to 1, since the associated sequence of probabilities $\{P_Z[t - \delta_v], v \geq 1\}$ monotonically converges to zero and (n_1, n_2) are fixed and finite.

According to Theorem 5.2.1, the related rejection rate converges to 1 for all $\alpha \geq \alpha_a$. T is weak unconditional finite-sample consistent. ■

Results obtained in Lemma 5.2.1, concerning the conditional finite-sample consistency of T , and in Theorem 5.2.1, concerning the weak unconditional finite-sample consistency of T , even in the presence of a function $\rho(\delta) > 0$ of effect δ (Theorem 5.2.2), can be extended to the case of random effects Δ .

Theorem 5.2.3. *Suppose that:*

- (i) T is any associative test statistic for one-sided hypotheses;
- (ii) sample sizes (n_1, n_2) are fixed and finite;
- (iii) the data set is $U(\Delta) = (Z_1 + \Delta, Z_2)$, where $(Z_1, Z_2) \in \mathcal{Q}^n$ are i.i.d. measurable real random deviates and Δ are random effects;
- (iv) random effects Δ diverge according to the monotonic sequence according to whatever sequence $\{\Delta_v, v \geq 1\}$, whose elements are stochastically non-decreasing, i.e., $\Delta_v \stackrel{d}{\leq} \Delta_{v+1}, \forall v \geq 1$;
- (v) $\lim_{v \uparrow \infty} \Pr\{\Delta_v > u\} \rightarrow 1$ for every finite u .

If conditions (i)–(v) are satisfied, then the permutation unconditional rejection rate of test T converges to 1 for all α -values not smaller than the minimum attainable α_a ; thus, T is weak unconditional finite-sample consistent.

Proof. In order to apply the Lebesgue's monotone convergence theorem it is sufficient that

$$P_Z(t - \Delta'' \leq u) \stackrel{d}{\leq} P_Z(t - \Delta' \leq u), \forall u,$$

whenever $\Delta' \stackrel{d}{\leq} \Delta''$. So the associated sequence of probabilities $\{P_Z[t - \Delta_v], v \geq 1\}$ monotonically converges to zero. ■

The divergence of random effects Δ can be realized by processing a divergent number k of quantitative variables (landmarks or semilandmarks).

It is not required that the k variables are independent, actually they can be dependent in any way. What is important is that the distribution induced by $T(U(0))$ is measurable and that of $T(U(\delta))$ diverges at least in probability. Hence these results are very useful in a multidimensional field as shape analysis.

The notion of unconditional finite-sample consistency, defined for divergent fixed effects δ , is different from the common notion of (unconditional) consistency of a test, which considers the behavior of rejection rate for given δ when $\min(n_1, n_2)$ diverges. It is known that, in order to attain permutation (unconditional) consistency it is required that random deviates Z have finite second moment (Hoeffding 1952). Here we only require measurability, so that random deviates Z are not required to be provided with finite moments of integer order equal to or greater than 1.

Theorem 5.2.4. *Suppose a two-sample problem, for one-sided alternatives with the data set $U(\delta) = (\delta + \sigma Z_1, \sigma Z_2)$, is such that:*

- (i) *the permutation test statistic T is associative and assumed to be weak unconditional finite sample consistent;*
- (ii) *conditions stated in Theorem 5.2.3 are satisfied;*
- (iii) *unidimensional random deviates Z are provided with null mean value (i.e., $\mathbf{E}(Z) = 0$);*
- (iv) *two sample sizes (n_1, n_2) satisfy the relation $(n_1 = vn'_1, n_2 = vn'_2)$, so that they can diverge according to the sequence $\{(vn'_1, vn'_2), v \geq 1\}$.*

Then for any given $\delta > 0$ the unconditional rejection probability of T converges to 1 as v diverges to the infinity; thus, T is weak unconditional consistent in accordance with the common notion of consistency.

Proof. Since the fixed effect δ is a unknown constant and sample sizes diverge, the common notion of consistency may be directly applied to T .

Hence let us organize the unidimensional data set $U(\delta)$ with 1 column and $n = n_1 + n_2$ rows, in a matrix $U'(\delta)$ with Q columns ($r = 1, \dots, Q$) and $n' = n'_1 + n'_2$ rows. Of course, as v diverges also $\min(n_1, n_2)$ diverges.

Let $T(U'(\delta))$ indicate the test statistic T applied to the data set $U'(\delta)$.

As the conditions of Theorem 5.2.1 and/or of Theorem 5.2.2 are satisfied by assumption, $T(U'(\delta))$ is unconditionally finite sample consistent.

Moreover, for any $v \geq 1$, the observed value of T applied to $U'(\delta)$ is given by $T(U'(\delta)) = \sum_{i \leq n'_1} \sum_{r \leq Q} \frac{U'_{r1i}(\delta)}{vn'_1}$ and applied to $U(\delta)$ is $T(U(\delta)) = \sum_{i \leq n_1} \frac{U_{1i}(\delta)}{n_1}$.

Certainly $T(U'(\delta)) = T(U(\delta))$.

Furthermore, we may write $T(U(\delta)) = T(U(0)) + \delta/\sigma = T(U'(\delta))$, emphasizing that two form have the same null distribution and the same non-centrality parameter which does not vary as v diverges. In contrast the null component $T(U(0))$, as v diverges, collapses almost surely toward zero by the strong law of large numbers. We recall that, by assumption, the random deviates Z admits finite first moment. Thus, in force of Theorem 5.2.2, the rejection probability for both ways converges to 1, $\forall \delta > 0$.

This allows us to state that weak unconditional finite sample consistency implies weak unconditional consistency, in accordance with the common notion of consistency, for all $\alpha \geq \alpha_a$. ■

An important observation must be done as regards the permutation sample space. In fact when processing the n -rows unidimensional data set $U(\delta)$ the permutation sample space has $\binom{n}{n_1}$ elements, and when processing the data rearranged according to the n' -rows Q -dimensional data set U' it has $\binom{n'}{n'_1}$ elements.

The two ways of considering permutation testing, given the same non-centrality, have the same unconditional power and so both are consistent for all $\alpha \geq \alpha_a = 1/\binom{n'}{n'_1}$. However, they are not completely equivalent in inferential terms. In order to prove their complete equivalence, we have to prove that both are consistent for all $\alpha > 0$ and that convergence should be obtained for any kind of sequences such that $\min(n_1, n_2)$ diverges.

We have proved that a unconditional δ -consistent associative T is also unconditionally consistent for all $\alpha \geq \alpha_a$ when the sequence of sample sizes is $\{(vn'_1, vn'_2), v \geq 1\}$. In practice, if we require consistency at least for $\alpha > \alpha^\circ$ and sample sizes are according to $\{(vn'_1, vn'_2), v \geq 1\}$, then we may find a pair of sample sizes (n'_1, n'_2) such that $\alpha^\circ > 1/\binom{n'}{n'_1}$. And so two ways are equivalent at least for all $\alpha \geq \alpha^\circ$.

Since for any arbitrarily chosen α° we may find a pair (n'_1, n'_2) such that $\alpha^\circ > \alpha_a$, then we may conclude that unconditional inferential conclusions associated with two ways are always coincident, provided that sample sizes are according to the sequence $\{(vn'_1, vn'_2), v \geq 1\}$. Hence, if deviates Z are provided with null mean value, any unconditional finite-sample consistent associative test statistic is unconditionally consistent at whatever α -value at least when sample sizes diverge according to the sequence $\{(vn'_1, vn'_2), v \geq 1\}$.

5.3 A Toy Example

Using the tpsDig2 program for digitizing landmarks and outlines for geometric morphometric analyses (Rohlf 2007), we have chosen $k = 98$ points along the contour and inside of a mosquito's wing, shown in Fig. 5.2. We have used the image of the left wing of a female *Aedes canadensis* (a woodland pool mosquito), available in the tpsDig2 program Examples.

For sake of simplicity, we have considered all the points (represented by the red bullets in the Fig. 5.2) as landmarks. Actually there are not true landmark points, at least they could be considered semilandmarks. Anyways we have decided to process them as true landmark points, since our goal is to investigate what happens to the power of permutation tests combination-based when the number of informative variables (landmarks) increases, while the number of cases is held fixed.

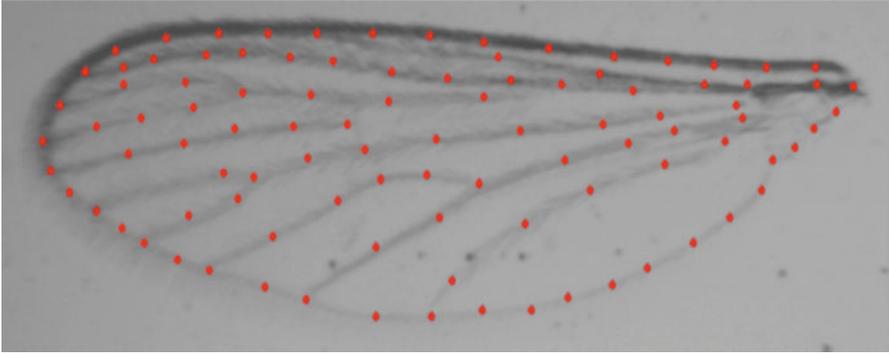


Fig. 5.2 $k = 98$ points registered in the left wing of female *Aedes canadensis*

In order to evaluate power behavior of nonparametric permutation tests combination-based when increasing the number of the processed variables or the value of the noncentrality parameter δ , thus studying the finite-sample consistency, we have carried out a simulation study. In particular, we have generated two independent samples from a multivariate normal distribution, in the particular case in which only five specimens are available in each group (small sample sizes). At the beginning we planned to generate two independent samples from a multivariate normal distribution, in the particular case in which only three specimens are available in each group. Unfortunately we found that in correspondence of nominal α levels 0.01 and 0.05, the power of the test was equal to 0, “jumping” to 1 in correspondence of $\alpha = 0.20$.

We guess that this behavior is due to the Procrustes superimposition process that probably has a more considerable stretching or shortening strength in the presence of small sample sizes. Let us assume that our samples are made of configurations of $k = 98$ landmarks in $m = 2$ dimensions, characterized by slightly different means.

Since they are not true landmarks, there is no rule in selecting points. Landmarks 1 and 2 correspond to the baseline, i.e., they represent the length of the wing, the other points have been chosen following the clockwise direction, with the unique intent of reproducing the main wing structures, thus drawing the image contour.

We have used the points digitized in Fig. 5.2 as an hypothetical configuration mean, before performing the superimposition. Therefore it contains the raw x and y coordinates. This mean will be used for generating data in the first group.

Data in the second group differs for that in the first one, according to a random percentage of variation represented by the parameter Δ . For example, in Table 5.2, the label “ Δ effect up 15%” means that Δ randomly varies in the interval $[0, 0.15]$ and then it is interpreted in terms of percentage change.

In Tables 5.1 and 5.2, in order to draw the attention of the reader, we have highlighted in bold the case $k = 10$.

Table 5.1 Controlling achieved α level and evaluating power: $n_1 = n_2 = 5$, $m = 2$, $B = MC = 1,000$

Nominal α	$\Delta=0$					
	$k = 2$	$k = 5$	$k = 10$	$k = 20$	$k = 50$	$k = 98$
0.01	0.004	0.008	0.006	0.002	0.006	0.011
0.05	0.035	0.047	0.044	0.041	0.048	0.054
0.10	0.065	0.092	0.096	0.096	0.098	0.112
0.20	0.132	0.182	0.199	0.192	0.197	0.208
0.30	0.217	0.285	0.297	0.296	0.302	0.292
0.50	0.383	0.487	0.494	0.490	0.492	0.487

In Table 5.1 we present simulation results under the null hypothesis that $\Delta = 0$. Type I error rate is under control. As covariance matrix, we have chosen a diagonal matrix with $\sigma^2 = 0.25$, i.e., we have considered homogeneous, independent, spherical variation at each landmark.

Simulation settings consider configurations made of $k = 2, 5, 10, 20, 50$ and $k = 98$ landmarks, with Δ effects up to 1%, 5%, 10%, 15%, and 20%. We have included the superimposition step (2D GPA), hence when carrying out nonparametric permutation tests we have used shape coordinates obtained after filtering out location, scale, and rotational effects from the original data.

With reference to the NPC procedure, in the first stage we have combined with respect to the coordinates (thus combining x and y coordinates for each landmark and obtaining k partial tests and their associated p -values). In the second and last step we have combined with respect to the landmarks (thus obtaining the global p -value). We have used the direct combining function in both steps.

Through this toy example it is possible to “appreciate” the notion of weak unconditional finite sample consistency for random effects. In fact, examining the results displayed in Table 5.2, we can see that, for a given and fixed number of subjects ($n_1 = n_2 = 5$), when the number of landmarks k and the random effects Δ both diverge, then the power of multivariate permutation tests based on Pesarin’s nonparametric combining functions converges quickly to one (Pesarin 2001; Pesarin and Salmaso 2010; Brombin 2009). It is also noteworthy that when $k = 2$ it seems that we are under H_0 . In this case, we are just considering the distance between points 1 and 2, corresponding respectively to the most extreme point on the left and on the right of the Fig. 5.2 (i.e., the baseline). Hence Procrustes superimposition process involve just a shortening step in order to obtain completely matching configurations.

Table 5.2 Evaluating power: $n_1 = n_2 = 5, m = 2, B = MC = 1,000$

Δ effect up 1 %						
α -level	$k = 2$	$k = 5$	$k = 10$	$k = 20$	$k = 50$	$k = 98$
0.01	0.000	0.139	0.081	0.400	0.634	0.712
0.05	0.031	0.576	0.415	0.908	0.998	1.000
0.10	0.057	0.790	0.613	0.989	1.000	1.000
0.20	0.133	0.933	0.808	0.999	1.000	1.000
0.30	0.206	0.974	0.900	1.000	1.000	1.000
0.50	0.358	0.998	0.968	1.000	1.000	1.000
Δ effect up 5 %						
0.01	0.004	0.627	0.556	0.583	0.749	0.723
0.05	0.025	0.988	0.988	0.993	1.000	1.000
0.10	0.045	1.000	1.000	0.999	1.000	1.000
0.20	0.123	1.000	1.000	1.000	1.000	1.000
0.30	0.205	1.000	1.000	1.000	1.000	1.000
0.50	0.372	1.000	1.000	1.000	1.000	1.000
Δ effect up 10 %						
0.01	0.003	0.433	0.666	0.723	0.740	0.730
0.05	0.027	0.927	0.998	1.000	1.000	1.000
0.10	0.062	0.991	1.000	1.000	1.000	1.000
0.20	0.131	1.000	1.000	1.000	1.000	1.000
0.30	0.215	1.000	1.000	1.000	1.000	1.000
0.50	0.355	1.000	1.000	1.000	1.000	1.000
Δ effect up 15 %						
0.01	0.002	0.525	0.707	0.722	0.715	0.717
0.05	0.027	0.953	1.000	1.000	1.000	1.000
0.10	0.058	1.000	1.000	1.000	1.000	1.000
0.20	0.131	1.000	1.000	1.000	1.000	1.000
0.30	0.193	1.000	1.000	1.000	1.000	1.000
0.50	0.367	1.000	1.000	1.000	1.000	1.000
Δ effect up 20 %						
0.01	0.004	0.734	0.730	0.732	0.721	0.711
0.05	0.031	1.000	1.000	1.000	1.000	1.000
0.10	0.062	1.000	1.000	1.000	1.000	1.000
0.20	0.133	1.000	1.000	1.000	1.000	1.000
0.30	0.204	1.000	1.000	1.000	1.000	1.000
0.50	0.378	1.000	1.000	1.000	1.000	1.000

References

Blair RC, Higgins JJ, Kamiski 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

Brombin C (2009) A nonparametric permutation approach to statistical shape analysis, ph.D. thesis. Padova, Italy: University of Padova

- Dryden IL, Mardia KV (1992) Size and shape analysis of landmark data. *Biometrika* 79:57–68
- Hoeffding W (1952) The large-sample power of tests based on permutations of observations. *Annals of Mathematical Statistics* 23:169–192
- Kent JT (1994) The complex bingham distribution and shape analysis. *Journal of the Royal Statistical Society: Series B* 56:285–299
- 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 (1999) Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification* 16:197–225
- Rohlf FJ (2000) On the use of shape spaces to compare morphometric methods. *Hystrix - Italian Journal of Mammalogy (NS)* 11:8–24
- Rohlf FJ (2007) tpsDig2, digitize landmarks and outlines, version 2.12. Department of Ecology and Evolution, State University of New York at Stony Brook
- Slice DE (2001) Landmark coordinates aligned by procrustes analysis do not lie in kendall's shape space. *Systematic Biology* 50:141–149
- Zelditch ML, Swiderski DL, Sheets HD (2012) *Geometric Morphometrics for Biologists, Second Edition: A Primer*. Academic Press, Elsevier Science Publishing Co Inc, United States

Chapter 6

Applications to Real Case Studies

“All great scientists have, in a certain sense, been great artists; the man with no imagination may collect facts, but he cannot make great discoveries”.

The Grammar of Science (1892), 37.

Karl Pearson

Abstract This chapter is mainly devoted to practical applications. In particular we present an application concerning the facial expression of emotion along with a case study aimed at analyzing aortic valve morphology. We also introduce two innovative topics: biometric morphing and nonparametric iterated combination for paired data. In practice, the result of the application of the NPC tests described in the previous chapters depends also on the choice of the combining function. Every combining function has its own characteristics and can be preferable to others in specific situations. As a consequence, results may differ slightly depending on which combining function is used. Sometimes, especially with categorical variables, different combining functions may lead to different results in terms of statistical significance. To remove these differences and assess the useful properties of many different types of combinations, the iterated combining procedure can be applied.

Keywords Biometric morphing • Expression of emotion • Iterated combination

6.1 Biometric Morphing and Facial Expression of Emotions

Biometric morphing is an innovative technique allowing pictures of a population of patients to be graphically represented by a single image that describes the average morphology of a precise anatomic region of interest. It combines procedures typical of statistical shape analysis and image processing ([Pahuta et al. 2009](#)).

Morphing is an effective image-processing tool that transforms (or morphs) one image into another through a seamless transition. Thin-plate spline geometric morphometrics (TPSGM) allow to quantify actual shape variation, thus taking into account the geometry of image deformation.

Biometric morphing is the application of TPSGM to morphing and resulting images are called *morphs*. Average results, induced by the therapy, may be visualized using this procedure, following the guidelines given by [Pahuta et al. \(2009\)](#). Here, we provide a brief summary of the algorithm which involves three steps: ([Brombin et al. 2011](#); [Alferi et al. 2012](#)).

1. Choose and digitize landmarks and semilandmarks tpsDig software has been used). Semilandmarks were interactively selected at equidistant intervals along anatomic contours or curves between landmarks, and imported into tpsUtil semilandmark definition software toolkit.
2. Perform Generalized Procrustes Analysis (GPA) to align patients. Pre- and post-therapy samples are separately processed (tpsRelw software has been used).
3. Unwarp the images on the basis of the transformation from original registered points to target GPA points. Average the unwarped images in order to generate an average (tpsSuper software has been used).

All the tps softwares are available online at <http://life.bio.sunysb.edu/morph/>.

In this chapter we apply biometric morphing to facial expression. We considered a subsample of 23 subjects performing two basic expressions, anger and disgust, along with their neutral facial expression. These images are contained in the MUG database.

Actually, the MUG database contains many sequences of an adequate number of subjects for the development and the evaluation of facial expression recognition systems that use posed expressions of the six basic emotions (i.e., anger, disgust, fear, happiness, sadness, surprise). In contrast to other similar databases, the MUG database is appropriate for the statistical evaluation of subject-independent and subject-dependent recognition systems using either images or image sequences. All the sequences are categorically labeled. Additionally, landmark points annotation is available for a significant number of images depicting the basic expressions of various subjects.

In particular, this database consists of image sequences of 86 subjects performing the six basic expressions. About 35 women and 51 men all of Caucasian origin, between 20 and 35 years of age, participated in the database. Men are with or without beard. There are no occlusions except for a few hair falling on the face.

Images were acquired under highly controlled conditions, as regards to patient position, camera position, background, illumination, light diffusion, and occlusions. More details can be found in [Aifanti et al. \(2010\)](#). A short tutorial about the basic emotions was given to the subjects. The subjects were informed about how the six facial expressions are performed according to the “emotion prototypes” as defined in the Investigator’s Guide in the FACS manual. The aim was to avoid erroneous expressions, that is expressions that do not actually correspond to their label. After

Fig. 6.1 Locations of 26 landmarks used in our investigation

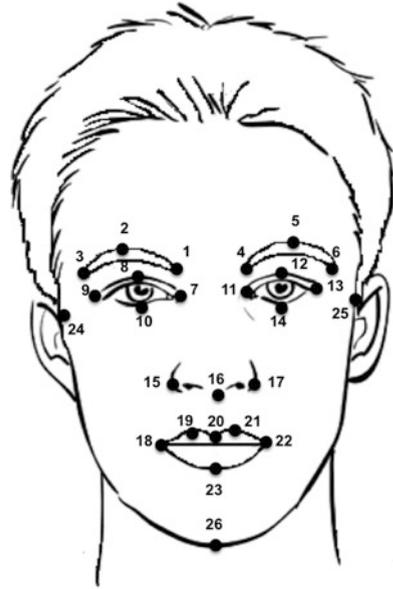


Table 6.1 The set of 26 landmarks used in the research

No.	Landmark description
3–6	Right/left lateral-ciliary points located above the most lateral aspect of the eyebrow
2–5	Right/left supra-ciliary points located above the most superior aspect of the eyebrow
1–4	Right/left inter-ciliary points located above the medial aspect of the eyebrow
7–11	Point at inner right/left side of the eye
8–12	Highest point on higher margin right/left eye
9–13	Point at outer right/left side of the eye
10–14	Lowest point on lower margin right/left eye
15–17	Right and left lateral alar points located on the lateral alar rims;
16	Nasal tip/Midpoint of the nose;
18–22	Right/left commissure points located on the commissure
19	Highest point on right side of lip
20	Midpoint on upper lip
21	Highest point on left side of lip
19–21	Right and left upper lip points located on the peak of Cupid's bow
23	Mid-lower lip point
24–25	Attachment of the right/left ear lobe to the cheek
26	Mid point of chin

the subjects had learned the different ways that the six expressions are performed, they freely chose to imitate one of them. About 80 landmark facial points were manually annotated on several images and the landmark point annotation for 401 images of 26 subjects is publicly available. As mentioned above, we focused on 23 subjects and on a subsample of 26 landmarks, manually annotated (see selected landmarks in Fig. 6.1 and Table 6.1).

The link between the subjective perception of emotions and their expression to the outer world has generated a great deal of clinical and research interest in analyzing expressed emotions.

In the field of 2D and 3D facial expression analysis, several imaging techniques and statistical methods have been developed to recognize/classify/quantify expression of emotion. In psychological literature, “emotion” has been defined as an umbrella concept that includes various different components such as affective, cognitive, behavioral, expressive, and physiological changes (Lambie and Marcel 2002; Panksepp 2005). More generally, emotions have been defined as subjective experiences, associated with mood, temperament, personality, and disposition and have been studied extensively in several fields, such as philosophy, psychology, sociology, and computer sciences.

Facial expression is one of the most powerful means for humans to communicate their emotions, cognitive states, intentions, and opinions to each other (Ekman 1982). Face can express emotion sooner than people verbalize or even realize their feelings.

Recently, research efforts have been directed toward the recognition of complex and spontaneous emotional phenomena (e.g., boredom or lack of attention, frustration, and stress) rather than on the recognition of deliberately displayed prototypical expressions of six basic emotions (see Sandbach et al. 2012, Nicolaou et al. 2011, Gunes and Pantic 2010, Zeng et al. 2009).

Some graphical representations obtained applying biometric morphing techniques are shown in Fig. 6.2, where consensus configurations are shown along with average results. Differences in shape between emotions are visible to a naked eye. However, we are interested in quantifying shape changes using inferential procedures.

Inferential methods in shape analysis are parametric in nature and may not be very powerful unless a large number of cases is available (Brombin 2009; Brombin et al. 2008; Brombin and Salmaso 2009). On the other hand, permutation tests represent an appealing alternative since they are distribution-free, allow for quite efficient solutions when the number of cases is lower than the number of covariates, they may be tailored for sensitivity to specific treatment alternatives and provide one-sided as well as two-sided tests of hypotheses (Blair et al. 1994). For this reason an extension of the NonParametric Combination (NPC) methodology to statistical shape analysis has been proposed in Brombin (2009), where the two-independent-sample case has been discussed extensively (Brombin and Salmaso 2009).

We have to deal with repeated measurements. In particular, if we want to compare angry and neutral expression or disgusted and neutral expression, we have to perform a nonparametric two dependent samples test, using as data the coordinates of aligned landmarks to statistically quantify these differences in mean. For each coordinate, we may define a suitable test and get a partial p -value. Then, combining these partial tests, it is possible to obtain a p -value for each landmark. Results are shown in Table 6.2. The p -values of the combined tests are adjusted with the close testing method to control the multiplicity of the test (p -FWE). Significance of the global p -values allows to reject the global null hypothesis of “no change” in shape induced by a certain emotion (Tables 6.3–6.6).

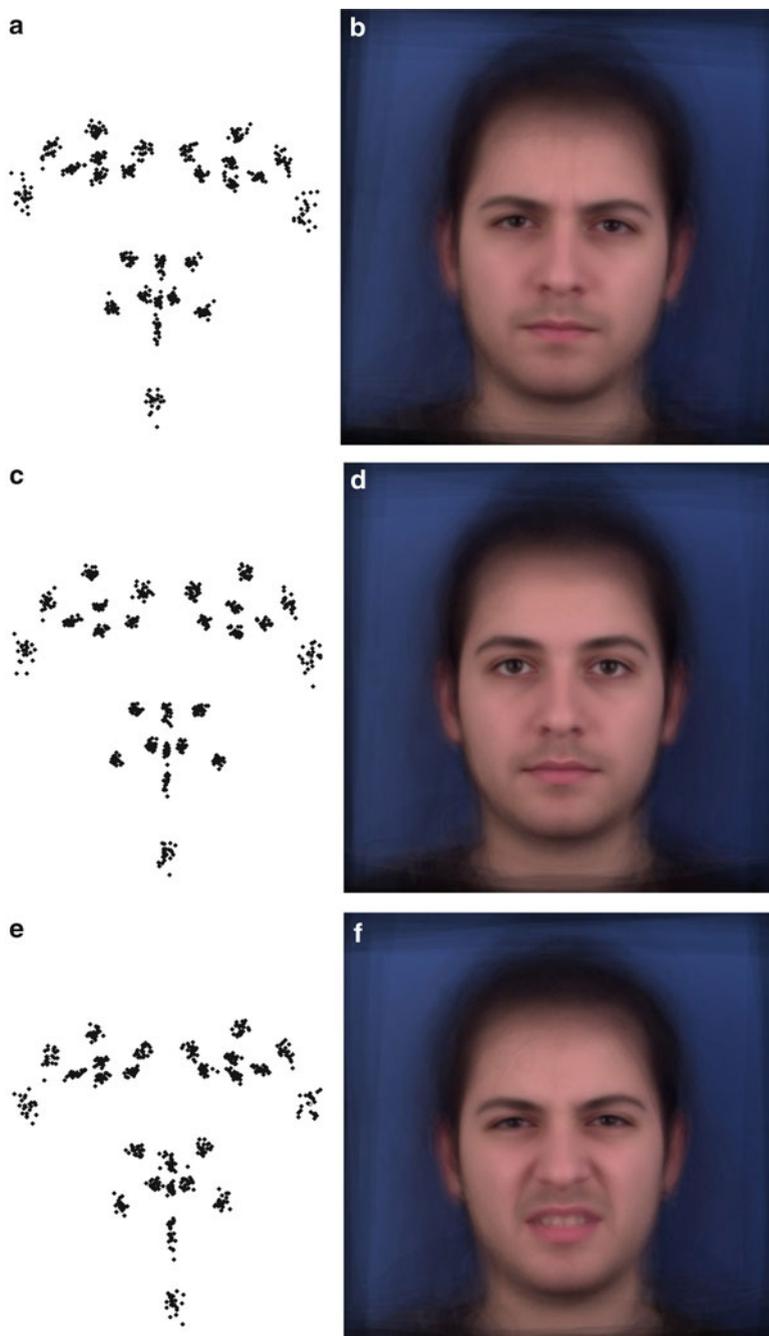


Fig. 6.2 Consensus distribution of 23 face images of the MUG dataset, when subjects are performing angry (a), neutral (c), and disgust (e) facial expression. Average results (morphs) of angry (b), neutral (d), and disgusted (f) subjects

Table 6.2 Comparison between angry and neutral facial expression 6.2(a) and disgusted and neutral expression 6.2(b)

(a)			(b)		
No.	p-value	p-FWE	No.	p-value	p-FWE
1	0.0001	0.0013	1	0.0001	0.0011
2	0.0001	0.0013	2	0.0002	0.0020
3	0.0001	0.0013	3	0.0001	0.0011
4	0.0001	0.0013	4	0.0001	0.0011
5	0.0002	0.0020	5	0.0001	0.0011
6	0.0001	0.0013	6	0.0016	0.0093
7	0.0003	0.0024	7	0.0003	0.0025
8	0.0001	0.0013	8	0.0001	0.0011
9	0.0001	0.0013	9	0.0035	0.0111
10	0.0001	0.0013	10	0.0001	0.0011
11	0.0001	0.0013	11	0.0001	0.0011
12	0.0001	0.0013	12	0.0025	0.0111
13	0.0398	0.0743	13	0.0001	0.0011
14	0.0001	0.0013	14	0.0005	0.0032
15	0.0004	0.0027	15	0.0003	0.0025
16	0.0014	0.0042	16	0.0105	0.0208
17	0.0004	0.0027	17	0.0001	0.0011
18	0.2349	0.2349	18	0.1981	0.1981
19	0.0001	0.0013	19	0.0001	0.0011
20	0.0002	0.0020	20	0.0001	0.0011
21	0.0002	0.0020	21	0.0001	0.0011
22	0.0002	0.0020	22	0.0001	0.0011
23	0.0002	0.0020	23	0.0001	0.0011
24	0.0005	0.0027	24	0.0001	0.0011
25	0.0001	0.0013	25	0.0001	0.0011
26	0.0009	0.0036	26	0.0023	0.0111
Global		0.0013	Global		0.0011

6.2 Some Remarks on Iterated Combination

For any given data set \mathbf{X} , different combining functions due to different rejection regions may of course give slightly different overall p -values, although, due to their consistency, they are asymptotically equivalent in the alternative. However, in order to reduce this influence, we may *iterate* the combination procedure by applying more than one combining function ψ_1, \dots, ψ_s , $2 \leq s$, to the same partial tests, and then combine the resulting second-order p -values $(\lambda_1'', \dots, \lambda_s'')$ into a third order of combination by means of one combining function, $\psi_l(\lambda_1'', \dots, \lambda_s'')$ say. From a series of Monte Carlo studies, provided that the second-order combination functions have different rejection regions, we obtained that the third-order p -values λ_l''' are almost invariant with respect to the choice of ψ_l within the class \mathcal{C} . Of course, this procedure may be iterated into a fourth order, and so on. Table 6.7 illustrates the iterated combination algorithm (Pesarin 2001; Pesarin and Salmaso 2010; Alfieri et al. 2012).

Table 6.3 Comparison between angry and neutral facial expression

No.	p -Iter Fisher	p -raw Fisher	p -FWE Liptak	p -Iter Liptak	p -raw Liptak	p -FWE Lancaster	p -Iter Lancaster	p -raw Lancaster	p -FWE
Land. 1	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 2	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 3	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 4	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 5	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 6	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 7	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 8	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 9	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 10	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 11	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 12	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 13	0.0316	0.0332	0.0319	0.0314	0.0331	0.0318	0.0320	0.0346	0.0320
Land. 14	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 15	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 16	0.0025	0.0023	0.0025	0.0025	0.0028	0.0025	0.0025	0.0024	0.0025
Land. 17	0.0002	0.0001	0.0002	0.0002	0.0001	0.0002	0.0002	0.0001	0.0002
Land. 18	0.2259	0.2287	0.2266	0.2223	0.2191	0.2248	0.2268	0.2330	0.2268
Land. 19	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 20	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 21	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 22	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 23	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 24	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 25	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 26	0.0004	0.0002	0.0004	0.0004	0.0009	0.0004	0.0004	0.0002	0.0004

Iterated, raw, and FWE adjusted p -values. $B=10,000$

Table 6.4 Comparison between angry and neutral facial expression

No.	p -Iter Fisher	p -raw Fisher	p -FWE	p -Iter Liptak	p -raw Liptak	p -FWE	p -Iter Tippett	p -raw Tippett	p -FWE
Land. 1	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 2	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 3	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 4	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 5	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 6	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 7	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 8	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 9	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 10	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 11	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 12	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 13	0.0291	0.0326	0.0330	0.0281	0.0343	0.0322	0.0474	0.0301	0.0474
Land. 14	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 15	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 16	0.0018	0.0019	0.0021	0.0017	0.0019	0.0020	0.0035	0.0036	0.0035
Land. 17	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0010	0.0002
Land. 18	0.2375	0.2303	0.2377	0.2377	0.2124	0.2377	0.2311	0.2146	0.2364
Land. 19	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 20	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 21	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 22	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 23	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 24	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 25	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 26	0.0002	0.0002	0.0002	0.0002	0.0001	0.0002	0.0003	0.0004	0.0003

Iterated, raw, and FWE adjusted p -values. $B=10,000$

Table 6.5 Comparison between disgusted and neutral facial expression

No.	p -Iter Fisher	p -raw Fisher	p -FWE Liptak	p -Iter Liptak	p -raw Liptak	p -FWE Lancaster	p -Iter Lancaster	p -raw Lancaster	p -FWE
Land. 1	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 2	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 3	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 4	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 5	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 6	0.0023	0.0015	0.0023	0.0023	0.0093	0.0023	0.0023	0.0112	0.0023
Land. 7	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 8	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 9	0.0015	0.0016	0.0015	0.0015	0.0011	0.0015	0.0015	0.0017	0.0015
Land. 10	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 11	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 12	0.0029	0.0024	0.0029	0.0029	0.0052	0.0029	0.0029	0.0024	0.0029
Land. 13	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 14	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 15	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 16	0.0054	0.0064	0.0054	0.0054	0.0043	0.0054	0.0054	0.0082	0.0054
Land. 17	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 18	0.1919	0.1846	0.1924	0.1894	0.2235	0.1913	0.1926	0.1743	0.1926
Land. 19	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 20	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 21	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 22	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 23	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 24	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 25	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 26	0.0022	0.0024	0.0022	0.0022	0.0021	0.0022	0.0022	0.0031	0.0022

Iterated, raw, and FWE adjusted p -values. $B=10,000$

Table 6.6 Comparison between disgusted and neutral facial expression

No.	p -Iter Fisher	p -raw Fisher	p -FWE Liptak	p -Iter Liptak	p -raw Liptak	p -FWE Tippett	p -Iter Tippett	p -raw Tippett	p -FWE
Land. 1	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 2	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 3	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 4	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 5	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 6	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 7	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 8	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 9	0.0012	0.0013	0.0015	0.0011	0.0009	0.0012	0.0016	0.0026	0.0016
Land. 10	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0002	0.0001
Land. 11	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 12	0.0039	0.0044	0.0051	0.0039	0.0078	0.0051	0.0050	0.0030	0.0051
Land. 13	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 14	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 15	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 16	0.0048	0.0047	0.0049	0.0049	0.0021	0.0049	0.0042	0.0486	0.0046
Land. 17	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 18	0.1768	0.1893	0.1903	0.1734	0.2241	0.1884	0.1924	0.1409	0.1924
Land. 19	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 20	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 21	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 22	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 23	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 24	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0006	0.0001
Land. 25	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001
Land. 26	0.0030	0.0030	0.0036	0.0027	0.0026	0.0034	0.0049	0.0050	0.0049

Iterated, raw, and FWE adjusted p -values. $B=10,000$

Table 6.7 The iterated combination algorithm

Ψ_1^o	Ψ_{11}^*	\cdots	Ψ_{1b}^*	\cdots	Ψ_{1B}^*
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Ψ_s^o	Ψ_{s1}^*	\cdots	Ψ_{sb}^*	\cdots	Ψ_{sB}^*
↓					
$\hat{\lambda}_1^{''}$	$\hat{L}_{11}^{''*}$	\cdots	$\hat{L}_{1b}^{''*}$	\cdots	$\hat{L}_{1B}^{''*}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\hat{\lambda}_s^{''}$	$\hat{L}_{s1}^{''*}$	\cdots	$\hat{L}_{sb}^{''*}$	\cdots	$\hat{L}_{sB}^{''*}$
↓					
$T_1^{'''o}$	$T_{11}^{'''*}$	\cdots	$T_{1b}^{'''*}$	\cdots	$T_{1B}^{'''*}$

Hence, the result of the application of the combination-based permutation test described in the previous section depends also on which combining function ψ is used. Every combining function has its own characteristics and can be preferable to others in specific situations. As such, results may differ slightly depending on which combining function is used. Sometimes, especially with categorical variables, different combining functions may lead to different results in terms of statistical significance. To remove these differences and assess the useful properties of many different types of combinations, the iterated combining procedure can be applied. The most commonly used combining functions are:

- Tippett’s combining function: $\psi_T = \max_{j=1,\dots,k}(1 - \lambda_j)$,
- Liptak’s combining function: $\psi_L = \sum_{j=1}^k \phi^{-1}(1 - \lambda_j)$,
- Fisher’s combining function: $\psi_F = -2 \sum_{j=1}^k \log(\lambda_j)$,
- Lancaster’s combining function: $\psi_G = \sum_{j=1}^k \Gamma_{r,a}^{-1}(1 - \lambda_j)$,

where $\phi(\cdot)$ represents the normal cumulative distribution function (CDF) and $\Gamma_{r,a}^{-1}$ is the inverse CDF of a central gamma distribution with known scale parameter a and r degrees of freedom.

All four functions satisfy the properties described in the previous section and can be used in the NPC procedure. The main difference relates to the rejection regions. In problems where some (but not many) sub-alternatives are expected to be true, Tippett’s combining function is preferable; when all the sub-alternatives or most of them are expected to be true, Liptak’s combining function should be used; Fisher’s combining function is good in all the intermediate situations (also known as “omnibus” combination); also Lancaster’s combining function has a similar behavior. Often the information about the number of true alternative sub-hypotheses is not known, hence the iterated combination represents a reasonable, practical, relatively efficient, and robust solution.

The iterated nonparametric combination can be performed using the following steps:

1. Calculate the second-order partial tests T'' with some selected combining functions, for example three out of the four above described, obtaining the observed

- values $\{T_{0,L}^{\prime\prime}, T_{0,G}^{\prime\prime}, T_{0,F}^{\prime\prime}\}$ and the permutation distributions $\{\mathbf{T}_L^{\prime\prime,*}, \mathbf{T}_G^{\prime\prime,*}, \mathbf{T}_F^{\prime\prime,*}\}$, as already pointed out in the first phase of the testing procedure at steps (i)–(iii);
2. Calculate the significance level functions of each test obtaining the three-variate test statistic ${}^1\hat{\mathbf{L}}(\mathbf{z}|\mathbf{X}) = \{{}^1\hat{L}_L(z|\mathbf{X}), {}^1\hat{L}_G(z|\mathbf{X}), {}^1\hat{L}_F(z|\mathbf{X})\}$ and the corresponding p -values ${}^1\hat{\lambda} = \{{}^1\hat{\lambda}_L, {}^1\hat{\lambda}_G, {}^1\hat{\lambda}_F\}$ (according to point (iv) of the first phase of the testing procedure);
 3. The three test statistics of the previous step are combined according to step (i) and (ii) of the second phase of the testing procedure; all three combining functions are applied so that three results (one for each combination) are achieved, thus obtaining ${}^2\hat{\lambda} = \{{}^2\hat{\lambda}_L, {}^2\hat{\lambda}_G, {}^2\hat{\lambda}_F\}$;
 4. Repeat t times step 3 with the combinations of the p -values ${}^{t-1}\hat{\lambda} = \{{}^{t-1}\hat{\lambda}_L, {}^{t-1}\hat{\lambda}_G, {}^{t-1}\hat{\lambda}_F\}$, $t \geq 3$, obtaining ${}^t\hat{\lambda} = \{{}^t\hat{\lambda}_L, {}^t\hat{\lambda}_G, {}^t\hat{\lambda}_F\}$, until the p -values of the t -th iteration are “very close” to those of the $(t-1)$ -th iteration or reasonably similar to each other according to a stopping rule, for example:
 - (a) After a specific number t_0 of iterations the procedure stops;
 - (b) Euclidean distance between the vectors of p -values of two consecutive iterations:

$$\|{}^t\hat{\lambda} - {}^{t-1}\hat{\lambda}\| \leq \varepsilon \text{ for a given } \varepsilon > 0 \text{ such that } \varepsilon \text{ is very close to zero;}$$
 - (c) “Deviance type” rule:

$$\sqrt{({}^t\hat{\lambda}_L - {}^{t-1}\hat{\lambda}_L)^2 + ({}^t\hat{\lambda}_G - {}^{t-1}\hat{\lambda}_G)^2 + ({}^t\hat{\lambda}_F - {}^{t-1}\hat{\lambda}_F)^2} \leq \varepsilon \text{ for a given } \varepsilon > 0, \text{ where}$$

$${}^{t-1}\bar{\lambda} \text{ is the sample mean of } {}^{t-1}\hat{\lambda}_L, {}^{t-1}\hat{\lambda}_G \text{ and } {}^{t-1}\hat{\lambda}_F.$$

A similar iterated solution was proposed for tests on genetic differentiation and independent samples and proved its good behavior under H_0 . Such a procedure for independent samples guarantees the rejection of the global hypothesis with probability α when all null (local) hypotheses are true (Salmaso and Solari 2006).

6.3 On Morphology of Aortic Valve

6.3.1 Introduction

In another application we have examined aortic valve shape. Data come from a real case study performed by Dr. Carla Villanova at “Casa di cura Villa Maria” in Padova (Italy). Preliminary results are given. Full results have been published elsewhere in Brombin et al. (2009).

Data consist of 16 echocardiograms, i.e., 16 2D pictures. As known echocardiography is one of the most widely used diagnostic tests for heart disease, since it is noninvasive and provides helpful information, e.g., size and shape of the heart, its pumping capacity, the location, and extent of any damage to its tissues. Moreover, it produces accurate assessment of the velocity of blood and cardiac tissue at any arbitrary point using pulsed or continuous wave doppler ultrasound. This allows us to evaluate cardiac valve areas and function, any abnormal communications between

the left and right side of the heart, possible valvular regurgitation, and calculation of the cardiac output. Our database contains complete patients information (e.g., age, gender, body mass index (BMI), systolic and diastolic blood pressure, and cardiac frequency (CF)). In particular, there are 9 men and 7 women. Mean age is 59.69 ± 17.25 years (7 patients are younger than 60 years old, 9 are sixties or older). Seven patients are overweighted or belong to the obese Class I (i.e., their BMI ranges from 25 to 35) and 9 are normal, i.e., BMI is comprised between 18.5 and 25.

Information concerning cardiovascular risk factors (e.g., familiarity with heart disease, smoking habits, hypertension, presence or not of dyslipidemia, diabetes, high cholesterol and/or triglycerides, and the practice of regular physical activity) were also recorded. Unfortunately some of these variables contain missing values. Measurements of M-mode left atrial maximum diameter and of the aortic valve have also been collected. Telediastolic and telesystolic volumes and ejection fraction (EF) of the 2D left ventricle have been calculated. Two-dimensional right ventricle surface area and shortening fraction (SF) have been determined in telediastole and telesystole. Mitral doppler variables (e.g., velocity of early filling wave (E), velocity of late filling wave due to atrial contraction (A), deceleration time, and regurgitation) and aortic doppler variables (e.g., proximal and distal velocity, regurgitation) have also been measured. Traditionally, the evaluation of the aortic valve status is based on morphometric measurements. Here we propose a landmark (and semilandmark)-based approach. At first we define the design of experiment. We have chosen 4 landmarks (red points in Fig. 6.3b), related to the measurements taken at the sinuses of Valsalva and ascending aorta (see dashed lines in Fig. 6.3a). Moreover we have digitized 20 semilandmarks (blue points) and 4 curves (curve 1 includes points 7–11; curve 2 points 12–16; curve 3 is made of points 17–21; and curve 4 of points 22–26).

Semilandmarks fail to be true landmarks in the fact that they do not enjoy homology property. They lie on homologous curves yet their exact position along these (usually smooth) curves is unclear. As a part of the superimposition procedure, the semilandmarks are allowed to slide along their curves in order to minimize the Procrustes distance from the actual landmark configuration to the sample average configuration. In Fig. 6.3 are shown consensus and all subjects are represented as points (c) or vectors indicating the variability at each point (d). In addition we have recorded 2 *artificial* landmarks (green points). We call them artificial since their utility is only related to a feature of tpsRelw program in particular with reference to the creation of a sliders file. Actually, slider files are used to define which semilandmarks should be allowed to slide along an estimated curve during the GPA superimposition. The points can be positioned so as to minimize the distance between the adjusted position and the corresponding point in the consensus or they can be positioned so as to minimize the bending energy required for a deformation of the consensus to the selected specimen. The program allows one to draw links between any triplets of landmarks. The middle landmark of a triplet is then considered a semilandmark. In order to define a curve made up of many semilandmarks, simply you have to define a series of overlapping triplets of points.

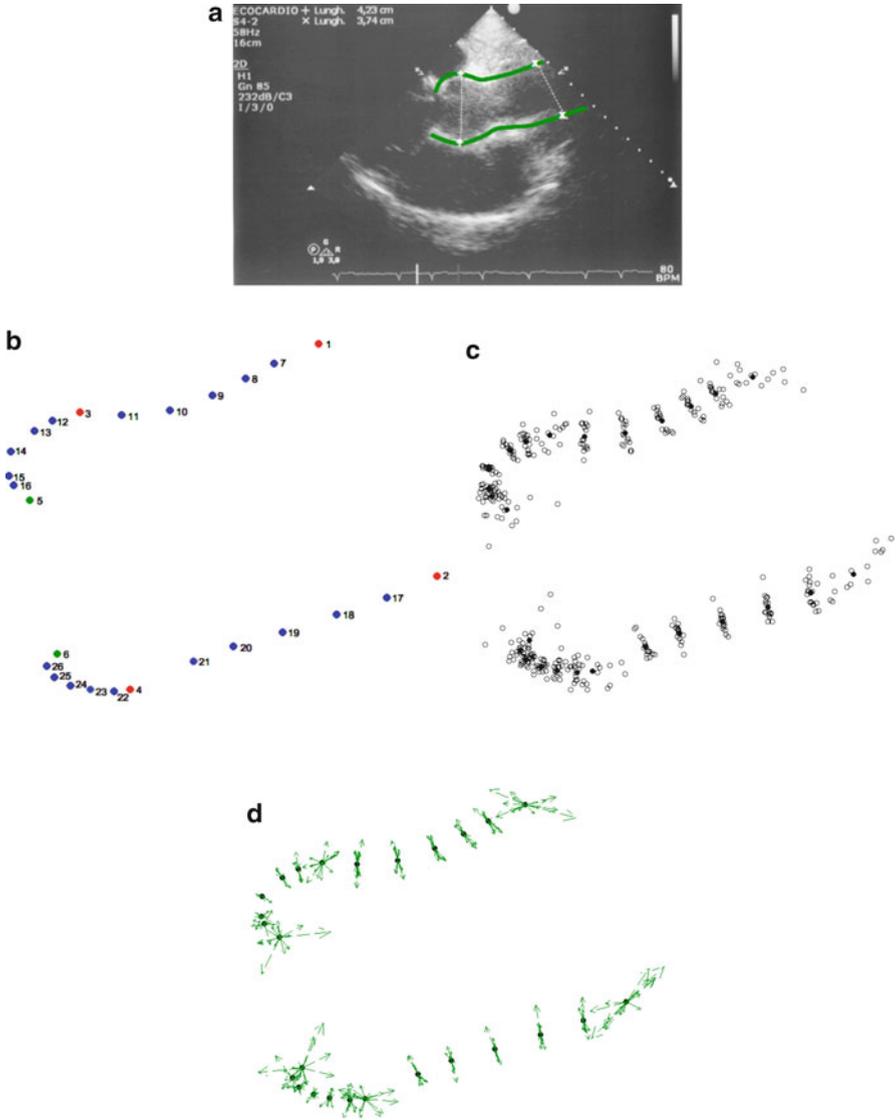


Fig. 6.3 Echocardiogram, distances, and outlines (a). Consensus configuration (b). Consensus and all subject represented as points (c) or vectors indicating the variability at each point (d)

While one can draw these links in any way that makes sense, a point can only be defined once as a semilandmark, i.e., it can only be used once as the middle point of a triplet. For details, we refer the reader to the Tpsrelw guide. Adding these points we may get more semilandmarks, thus obtaining a better description of the whole shape. TpsRelw, as previously said, provides a plot of the relative warp scores

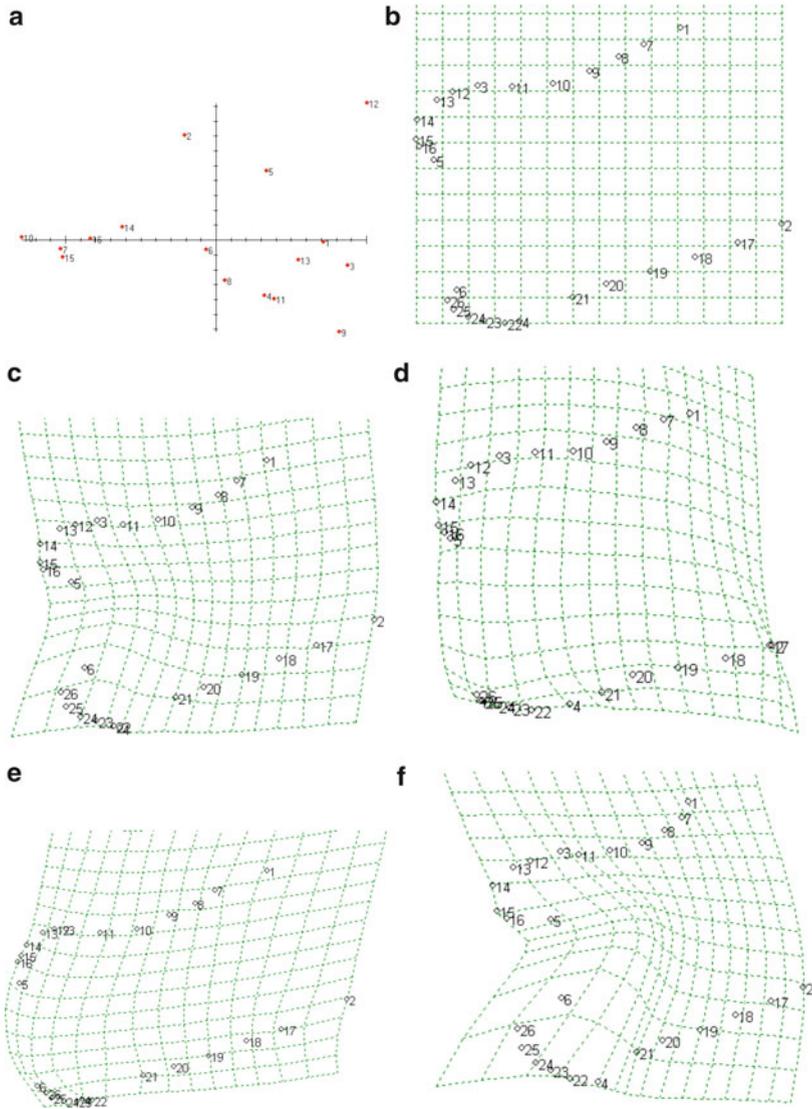


Fig. 6.4 Output from Tpsrelw. Plot of the relative warp scores matrix (a). Deformation grid related to the consensus (b) and changes in shape when moving from the center to patient 2 (c), or to patient 9 (d), 10 (e), and 12 (f)

matrix also showing the position of each specimen with respect to the first and second partial warps (see for example Fig. 6.4a). It also allows us to explore the deformations associated with different position in this ordination. For example, in Fig. 6.4b we show the deformation corresponding to the position of the consensus. In Fig. 6.4c, we show the deformation produced when moving from the center to

patient 2. We can see a contraction of the grid: points 5 and 6 shift inside, thus producing an increase in distances between points 5–16 and points 6–26, while the distance between points 11 and 21 decreases. Again, in Fig. 6.4d, we display the deformation corresponding to the position of patient 9. We can see that there is a global enlargement and lengthening of the shape, with a loss of the roundness of the curve 4. In Fig. 6.4e we display the deformation in shape corresponding to patient 10. There is a global shape prolongation of the shape and a loss of roundness of the curve 4. Moreover it could be noticed an increase in the distances between points 1–7 and points 2–17. To conclude, in Fig. 6.4f we show the deformation corresponding to the position of patient 12. We can see a remarkable contraction of the shape. Points 5 and 6 are responsible for this variation in shape: points 5 and 6 shift inside, thus producing an increase in distances between points 5–16 and points 6–26.

6.3.2 Inferential Results

We have used the plot of the relative warp scores matrix to define two groups. The first group includes patients 2, 6, 7, 10, 14, 15, and 16 ($n_1 = 7$). The second includes patients 1, 3, 4, 5, 8, 9, 11, 12, and 13 ($n_2 = 9$). Five of the seven overweighted or obese patients in our sample are allocated in this group. We have carried out a two independent sample test, using NPC methodology, in order to see where are located significant shape differences among these two groups. As a remark, our data are now the 2D coordinates of landmarks and semilandmarks after sliding and after superimposition, minimizing the Procrustes distance. Here we deal with $k = 26$ points in $m = 2$ dimensions. As usual we break the problem up into two stages, considering both the coordinate and the landmark level (and, if present, the domain level too). In particular we formulate partial test statistics for one-sided hypotheses and then we consider the global test obtained after combining at the first stage with respect to m and in the second stage with respect to k .

Actually x and y landmark and semilandmark coordinates could be considered the sub-hypotheses of the problem, thus providing a set of partial tests. Hence, combining these partial tests we can get a p -value for each 2D landmark and semilandmark. We also have computed p -values for each curve (corresponding to a domain), as well as a global p -value (see results in Table 6.8). We have found that the two groups are significantly different in all the six landmarks, in curves 1, 2, and 3 and globally. We recall that p -values associated with the curves are also global p -values. Actually these p -values are obtained after combining all the landmark coordinates of the points included in the curve itself (e.g., p -value associated with curve 1 is obtained as a combination of the x and y coordinates of points 7, 8, 9, 10, and 11). In Table 6.8, we call “global” the p -value we get after combining all the previously obtained p -values. Here we have also used a closed testing procedure controlling the familywise error rate (FWE).

Table 6.8 Results

	p-value
Landmark 1	0.0028
Landmark 2	0.0001
Landmark 3	0.0006
Landmark 4	0.0010
Landmark 5	0.0218
Landmark 6	0.0553
Curve 1	0.0005
Curve 2	0.0316
Curve 3	0.0077
Curve 4	0.6246
Global	0.0023

This kind of analysis cannot be carried out in a parametric framework, since standard Hotelling's T^2 is approximately distributed according to an F_{M, n_1+n_2-M-1} , where $M = km - m - m(m-1)/2 - 1$ is the dimension of the tangent space. In this case $k = 26$, $m = 2$, and $n_1 + n_2 = 16$; hence we should calculate $F_{48, -33}$, which is impossible.

In groups defined using information on BMI, age, and gender, no significant differences among patients have been found. This is probably due to the very small sample size.

6.4 Final Remarks

NPC tests, due to their nonparametric nature, may be computed even when the number of covariates exceeds the number of cases. With reference to the problem of small sample sizes, we recall that the results obtained within the NPC framework can be extended to the corresponding reference population. In [Pesarin \(2002\)](#) it is proved that it is possible to extend the permutation conditional to unconditional or population inferences since permutation tests are provided with similarity and conditional unbiasedness properties. Actually in the parametric field, this extension is possible when the data set is randomly selected by well-designed sampling procedures on well-defined population distributions, provided that their nuisance parameters have boundedly complete statistics in the null hypothesis or are provided with invariant statistics. In practice, this situation does not always occur and parametric inferential extensions might be wrong or even misleading. Permutation tests enable us for such extensions, at least in a weak sense, requiring that the similarity and conditional unbiasedness properties (sufficient and not necessary conditions) are jointly satisfied ([Pesarin 2002](#); [Ludbrook and Dudley 1998](#)). Moreover, we have shown how NPC methodology enables the researcher to give local assessment using a combination with domains. We feel confident that

developing geometric morphometrics techniques in a nonparametric permutation framework makes possible to obtain valid solutions for the high dimensional and small sample size problems.

References

- Aifanti N, Papachristou C, Delopoulos A (2010) The MUG Facial Expression Database. In: 11th Int. Workshop Image Analysis for Multimedia Interactive Services (WIAMIS)
- 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
- 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
- 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, Salmaso L, Villanova C (2009) 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
- Brombin C, Salmaso L, Ferronato G, Galzignato P (2011) Multi-aspect procedures for paired data with application to biometric morphing. *Communications in Statistics - Simulation and Computation* 40:1–12
- Ekman P (1982) *Emotion in the human face* (2nd ed.). New York: Cambridge University Press
- Gunes H, Pantic M (2010) Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions* 1:68–99
- Lambie JA, Marcel AJ (2002) Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review* 109:219–259
- Ludbrook J, Dudley H (1998) Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician* 52:127–132
- Nicolaou M, Gunes H, Pantic M (2011) Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing* 2:92–105
- Pahuta MA, Rohlf FJ, Antonyshyn OM (2009) Biometric morphing: A novel technique for the analysis of morphologic outcomes after facial therapy. *Annals of plastic surgery* 62:48–53
- Panksepp J (2005) Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition* 14:30–80
- Pesarin F (2001) *Multivariate Permutation tests: with application in Biostatistics*. John Wiley & Sons, Chichester-New York
- Pesarin F (2002) Extending permutation conditional inference to unconditional one. *Statistical Methods and Applications* 11:161–173
- Pesarin F, Salmaso L (2010) *Permutation Tests for Complex Data: Theory, Applications and Software*. Wiley

- Salmaso L, Solari A (2006) Nonparametric iterated combined tests for genetic differentiation. *Computational Statistics & Data Analysis* 50:1105–1112
- Sandbach G, Zafeiriou S, Pantic M, Yin L (2012) Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing* DOI:10.1016/j.imavis.2012.06.005
- Zeng Z, Pantic M, Roisman G, Huang T (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. *The IEEE Transactions on Pattern Analysis and Machine Intelligence* 31:39–58