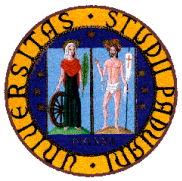


Università degli Studi di Padova
Dipartimento di Scienze Statistiche

Scuola di Dottorato in
Scienze Statistiche
Ciclo XXI

A nonparametric permutation approach to statistical shape analysis

Chiara Brombin

Direttore: Prof.ssa Alessandra Salvan

Supervisore: Prof. Luigi Salmaso

Co-supervisore: Prof. Fortunato Pesarin

2 Febbraio 2009

Acknowledgements

Somebody, who I will never forget for a lot of reasons, defined me a tenacious woman, adding that this is a quality and a limitation at the same time. A quality because it, always or very often, allows you to achieve your goals. This implies that you are competitive, ambitious and motivated to go over the top. But, when you are in trouble, you feel frustrated, sad, and people cannot stand you.

Well, this thesis is the result of three years of hard work, during which I have been accompanied and supported by many people and, beyond a shadow of a doubt, by my family.

Here I wish to express my gratitude to all of them, still apologizing for those periods in which I was intractable.

First of all, I would like to gratefully acknowledge prof. Luigi Salmaso and prof. Fortunato Pesarin for supervising my work during these three years, for their trust and constant encouragements.

I am grateful to Stanislav Katina, that I visited in Vienna in 2006, who introduced me to the shape analysis field.

I would also like to thank prof. James Rohlf, that I visited in 2007, and with whom I had many discussions on shape analysis. It was a pleasure to work with him at Stony Brook University, where I had my best research and life experience ever.

Special thanks go to Livio F., Dario and Livio C. for their helpful suggestions and for the time spent together during conferences.

I cannot forget in these acknowledgements my colleagues and friends of the Ph.D. program, Alberto, Fany and Manuela, as well as my friends Lucia, Carla, Elena, Giulia, Giovanni, Alessandro, Ezio, Nicola, Connie, Lois&Kate, Emily...and the list is very very long. All these friends have played an important role during this time, supporting me in heavy days and showing me the positive side of the situation.

To conclude, I am forever indebted to my parents and my brother for their support, confidence, understanding, endless patience and encouragement. I dedicate this work to them.

To my family

The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity.

Albert Einstein

Contents

Abstract

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 Ph.D. thesis we present an extension of NonParametric Combination (NPC) methodology (Pesarin, 2001) to 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, focussing on the two independent sample case, through an exhaustive comparative simulation study, we evaluate the behaviour of traditional tests along with nonparametric permutation tests using also Multiple Aspect (MA) procedures and domain combinations. The case of heterogeneous and dependent variation at each landmark is also investigated, along with the effects of superimposition on the power of NPC tests. 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. 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. 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. Along with simulation studies, we present two applications to real data sets concerning Mediterranean monk seal skulls and aortic valve morphology.

Riassunto

Nell'ultimo decennio la comunità statistica ha mostrato un crescente interesse per i problemi di shape analysis, con particolare riferimento allo sviluppo di tecniche inferenziali robuste. In questa tesi di dottorato presentiamo un'estensione della metodologia NPC per la combinazione non parametrica di test di permutazione dipendenti (Pesarin, 2001) nell'ambito della shape analysis. Inizialmente si introduce una revisione dei metodi inferenziali noti in letteratura, evidenziando alcune problematiche legate all'uso della statistica test T^2 di Hotelling. Focalizzandoci poi sul caso di due campioni indipendenti, tramite un esauriente studio di simulazione, abbiamo confrontato il comportamento, in termini di potenza, dei test parametrici tradizionali con quello dei test non parametrici proposti. Sono state utilizzate anche procedure di tipo multi aspetto (MA) e combinazioni per domini. È stato anche esaminato il caso in cui i landmark sono correlati tra loro. Inoltre è stato valutato l'impatto della superimposizione sulla potenza dei test NPC. I test di permutazione sono stati valutati in potenza e sotto H_0 nel caso in cui il numero di variabili processate è superiore alla cardinalità dello spazio di permutazione. Abbiamo inoltre effettuato uno studio di simulazione per valutare la potenza dei test multivariati NPC, evidenziando che la potenza di questi test cresce al crescere del numero di variabili processate, qualora apportino un aumento della non centralità, anche quando il numero di variabili è superiore alla cardinalità dello spazio di permutazione. Questi risultati preliminari ci hanno consentito di estendere la nozione di *finite-sample consistency* per i test NPC nell'ambito della shape analysis. Vengono fornite condizioni sufficienti tali per cui la potenza del test converge a uno, per ampiezze campionarie fissate ad ogni livello raggiungibile α, quando il numero di variabili diverge, posto che diverga anche la non centralità indotta dall'aumento del numero di variabili. Sulla base dei risultati ottenuti, possiamo affermare che i test NPC forniscono soluzioni efficienti per i problemi multivariati di shape analysis in presenza di bassa numerosità campionarie, problemi del resto frequenti nell'ambito della shape analysis. Oltre agli studi di simulazione, vengono presentati due casi studio, uno relativo allo studio della forma del cranio della foca monaca del Mediterraneo e l'altro relativo alla morfologia della valvola aortica.

Chapter 1

Introduction

1.1 Overview

Statistical shape analysis is considered a cross-disciplinary field, allowing for applications in biology, geology, medicine and many other sciences, since the theory and techniques are very flexible and potentially adaptable to any appropriate configuration matrices. 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 models for shape variation of entire configurations of points corresponding to the locations of morphological landmarks. Inferential methods known in the shape analysis literature make use of configurations of landmarks optimally superimposed using a least-squares procedure or analyze matrices of interlandmark distances.
For example, in the two independent sample case, a practical method for comparing the mean shapes in the two groups is to use the Procrustes tangent space coordinates and if data are concentrated, calculate the Mahalanobis distance and then the Hotellings T^2 test statistic. Under the assumption of isotropy, another simple approach is to work with statistics based on squared Procrustes distance and then consider the Goodall's F test statistic.
All the above mentioned tests are based on quite stringent assumptions, such as the equality of covariance matrices, the independency of variation within and among landmarks or the multinormality of the model describing landmarks.
As pointed out in Good (2000), the assumption of equal covariance matrices may be unreasonable in certain applications, the multinormal model in the tangent space may be doubted and sometimes there are few individuals and many landmarks, implying over-dimensioned spaces and loss of power for the Hotelling's T^2 test. Hence an alternative procedure is to consider a permutation approach. Further limitations of traditional inferential procedures have been highlighted in Terriberry et al. (2005). Actually, useful shape models contain parameters lying in non-Euclidean spaces. Some of these parameter parameters may have a large variance, may be highly correlated, or have completely different scales, thus invalidating analyses and final results. Hence traditional statistical tools designed for Euclidean spaces must be used with particular care when they are applicable. On the contrary, permutation

tests are appealing because they make no distributional assumptions, requiring only that the data in each group are exchangeable under the null hypothesis. In the paper by (Terriberry et al., 2005), they presented an application of NPC methodology in shape analysis, but the properties of the method itself are not further investigated. On the strength of these considerations, we suggest an extension of NonParametric Combination (NPC) methodology (Pesarin, 2001) to shape analysis. Actually, under very mild and reasonable conditions, the NPC method is found to be consistent, unbiased and extendible to unconditional inferences. We remark that in the parametric approach, this extension is possible when the data set is randomly selected by well-designed sampling procedures on well-defined population distributions. It is well known that in practice this situation very rarely occurs. When similarity and conditional unbiasedness properties are jointly satisfied, and if correctly applicable, permutation tests allow for inferential extensions at least in a weak sense (Pesarin, 2002; Ludbrook and Dudley, 1998).
We observe that permutation tests require homogeneous covariance matrices in order to guarantee exchangeability only under H_0 thus relaxing more stringent assumptions required by parametric tests, since they do not require homoscedasticity in the alternative.
At first we review inferential methods known in the shape analysis literature, highlighting some drawbacks of using Hotelling's T^2 test statistic. Then, focussing on the two independent sample case, through an exhaustive comparative simulation study, we evaluate the behaviour of traditional tests along with nonparametric permutation tests using also Multiple Aspect (MA) procedures and domain combinations.
In this nonparametric framework we also analyze the case of heterogeneous and dependent variation at each landmark.
Besides we examine the effect of superimposition on the power of NPC tests. Finally we introduce the notion of finite-sample consistency for combination-based permutation tests in shape analysis. 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.
In particular we show that it is possible to obtain powerful tests in a nonparametric framework by increasing the number of informative variables and not the number of cases. On the basis of these findings and due to their nonparametric nature, we may assert that the suggested tests provide efficient solutions, allowing to deal with datasets including many informative landmarks and few specimens. Along with simulation studies, we present two applications to real data sets concerning Mediterranean monk seal skulls and aortic valve morphology.

1.2 Contributions of the thesis

An overview of the original results obtained during the Ph.D. thesis development and presented in the thesis is listed below.

- At first we highlight that the proposed extension of the NPC method per-

forms better than traditional tests used in shape analysis in terms of power. Moreover, using the MA procedure and the information about domains, NPC enables the researcher 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.

- As expected, the encouraging results obtained in this first study have been confirmed even in the case of heterogeneous and dependent variation at each landmark (nonzero covariance).

- With reference to the drawbacks of using Hotelling's T^2 test, we have proposed a nonparametric permutation counterpart, stressing the case in which the number of variables is larger than the permutation sample space and finding it very powerful. On the basis of these results, we have performed a simulation study to evaluate the power of multivariate NPC tests, showing that the power for the suggested tests increases when increasing the number of the processed variables provided that the induced noncentrality parameter δ increases, even when the number of covariates is larger than the permutation sample space.

- These preliminary results allowed us to extend the notion of *finite-sample consistency* for permutation tests combination-based to the shape analysis field. 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 $h = 1, \ldots, km$ variables, describing k landmarks in m dimensions) and the associated noncentrality parameter δ, induced by the test statistic, both diverge, then the power function of multivariate NPC tests based on associative statistics converges to one.
  We demonstrate their conditional finite-sample consistency and weak unconditional finite-sample consistency, even in the case we are examining a function $\rho(\delta) > 0$ of effects δ or we are considering random effects Δ.
  Such findings look very relevant to solve multivariate small sample problems (like those encountered in shape analysis field) since they dimostrate 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.

- With reference to the first case study, data at hand consist of 17 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.
We propose to process these data applying geometric morphometric techniques and carrying out inference in a nonparametric permutation framework.
We have chosen 4 anatomical landmarks and 24 semilandmarks, also defining three domains. By means of the NPC approach we are able to assign without classification errors young and adult specimens.

- The second case study concerns aortic valve morphology. Preliminary results are given. Data at hand consists of 16 echocardiograms including complete patients information (e.g. age, gender, BMI, systolic and diastolic blood pressure, and cardiac frequency). Information concerning cardiovascular risk factors were also recorded. Several other variables provided by echocardiography (e.g. the evaluation of the velocity of blood and cardiac tissue, the assessment of cardiac valve areas and function, possible valvular regurgitation, and calculation of the cardiac output) were also available. We have digitized 4 landmarks, 20 semilandmarks, 4 curves and 2 artificial landmarks. We have analyzed relative warps and deformation grids. In particular, we have used the plot of the relative warp scores matrix to define two groups. Using NPC methodology, we have found significant differences between the two groups.

Chapter 2

A brief overview on statistical shape analysis

2.1 Some historical notes

It is difficult to imagine a time in history when people have not been fascinated by shapes. Our visual fine arts, such as painting and sculpture, have appeal across cultures and illustrate the universality of shapes or forms (Small, 1996). Usually the term *shape* is used to indicate the external form or appearance characteristic of an object, the outline of an area or figure.
Statistical shape analysis 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. 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).

Definition *Size-and-shape is all the geometrical information that remains when location and rotational effects are filtered out from an object.*

In order to have the same size-and-shape, two objects are required to be rigid-body transformations of each other (Dryden and Mardia, 1998).
Statistical shape analysis is considered a cross-disciplinary field characterized by flexible theory and techniques. Specific applications of shape analysis may be found in archaeology, architecture, biology, geography, geology, agriculture, genetics, medical imaging, security applications such as face recognition, entertainment industry (movies, games), computer-aided design and manufacturing and so on. David Kendall and Fred Bookstein are without doubt the pioneers in this field. But to be precise, even Galilei (1638) had shown an interest in shapes and in particular knew that bones in larger animals are not solely scaled up versions of those in smaller

animals, but there is a shape difference too. Actually a bone has become proportionally thicker so that it does not break under the increased weight of the heavier animal (Dryden and Mardia, 1998). In 1977, Kendall publieshed a brief note in which he introduced a new representation of shapes as elements of complex projective spaces. He stated that under an appropriate random clock, the shape of a set of independent particles diffusing according to a Browninan motion law could be regarded as a Browninan motion on complex projective space Small (1996). But at that time many statistician did not see the practical utility of this claim for their own applications. It was only in 1984 that full details of Kendall's theory of shape were published. Endowed with great elegance, this work contained some intriguing areas of research, thus drawing the attention of both the probabilists and the statisticians. Mardia et al. (1977) investigated the distribution of the shapes of triangles generated by certain processes, and in particular considered wheter towns in a plain are spread regularly with equal distances between nieghbouring towns. Dryden and Mardia peculiar interest in statistical shape analysis began in 1986, with an approach from Paul O' Higgings and David Johnson in the Department of Anatomy at the University of Leeds (UK), asking for advice about the analysis of the shape of some mouse vertebrae.
The same year (1986) Kendall was invited to be a discussant for a stimulating article by Bookstein in the journal of *Statistical Science*, then published in Volume 1 (Bookstein, 1986). In that occasion it was easy to perceive that there was a close connection between the elegant and deep mathematical work in shape theory from Kendall landmark paper (Kendall, 1984) and the practical application proposed in Bookstein's paper.
Even if Kendall and Bookstein had the same intuition with reference to the possibility of representing spaces on manifolds, their proposals were essentially original in the underlying theory and in the applications each researcher emphasized. 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, his applications were mainly addressed to archeological and astronomical sciences and he studied the shapes of random sets of points, such as are to be found in a Poisson scattering. 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), that initially developed the field of geometrical shape analysis from a biological point of view, Julian Huxley (1932), and later researchers in allometry and multivariate morphometrics (Small, 1996).

2.2 How to describe shapes

A substantial role in shape analysis research has been played by "landmark-based" analysis, where shapes are represented by a discrete sampling of the object contours (Dryden and Mardia, 1998; Small, 1996).

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 represents 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 (Dryden and Mardia, 1998).

- *Type I landmarks* (usually the easiest and the most reliable to locate) are mathematical points whose homology is reinforced by the strongest evidence, such as a local pattern of juxtaposition of tissue types or a small patch of some unusual histology.

- *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.

- *Type III landmarks* are the most difficult and the least reliable to locate. They occur at extremal points or constructed landmarks (e.g. maximal diameters and centroids) and have at least one deficient coordinate, for instance, either end of a longest diameter, or the bottom of a concavity. 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.
Along with landmarks, it is possible to collect semilandmarks points, that 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 lies 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 characterising 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 (midsagitttal curve);

Type 5: semilandmarks on surfaces;

Type 6: constructed semilandmarks.

In order to illustrate how landmarks and semilandmarks are chosen and then classified in real applications, we briefly introduce our case study on monk seal skulls.

2.2.1 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.
Figures 2.1-2.2 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;

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 continaution of the sutures until these lines intersect;
- canine base (denoted by *cb*), most mesial point on the outler 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 zygomatyc process.

In the left-lateral view we have chosen 5 midplane landmarks and 7 bilateral landmarks (see Table 2.1 and Figure 2.1).
In the frontal view we have chosen 4 midplane landmarks and 8 bilateral landmarks. Landmarks marked by grey 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 Figure 2.1 we show the landmarks and the curves we have chosen. The grey 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 (Figure 2.2).

2.3 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). 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). 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

Table 2.1: Anatomical landmarks

MIDSAGITTAL
Prosthion, pr
Nasospinale, ns
Rhinion, rhi
Nasion, n
Bregma, b
BILATERAL
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

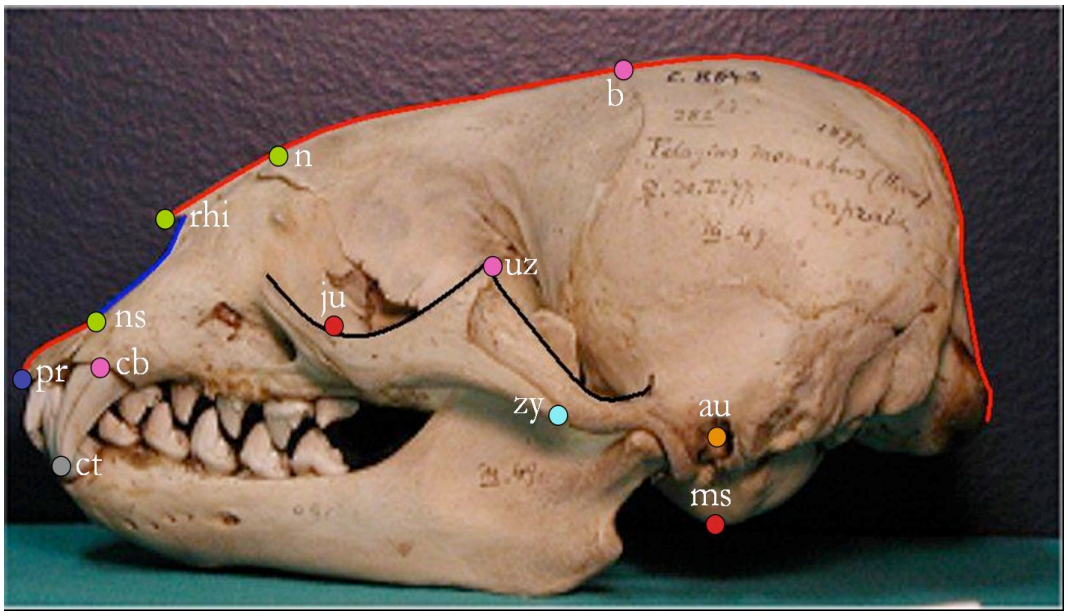


Figure 2.1: Left-lateral view (Legend: Type 1 ●, Type 2 ●, Type 3 ●, Type 4 ●, Type 5 ●, Type 6 ●, unknown Type ●, nasal curve ∼, midsagittal curve ∼, zygomatic curve ∼)

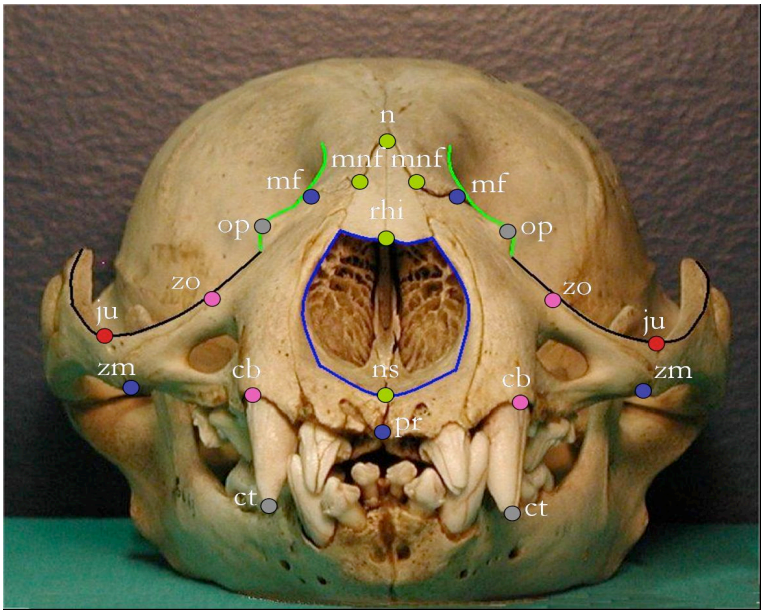


Figure 2.2: Frontal view (Legend: Type 1 •, Type 2 •, Type 3 • Type 4 •, unknown Type •, nasal curve ∼, zygomatic curve ∼, part of orbital curve ∼)

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. However the geometric origin of the measured variables is generally not exploited further. Moreover visualizing results through graphical representations of shape is very demanding because the geometric relationships among the variables (linear distances) are not preserved, thus loosing some aspects of shape.
On the other hand, 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 lenghts 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 which we mention Fred Bookstein and James Rohlf, proposed to combine traditional multivariate morphometrics and deformation analysis, calling this synthesis *geometric morphometrics*. 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).
The direct analysis of databases of landmark locations is not convenient because of the presence of nuisance parameters, such as position, orientation and size. In order 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 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 presence of semilandmarks, a newsworthy method it 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 software guide by Rohlf (2008a).
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 EDMA (Euclidean Distance Matrix Analysis) 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).

Chapter 3

Multi-aspect permutation tests in shape analysis with small sample size

3.1 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. Technical details on the above mentioned tests are provided in the Appendix 3.A.

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 distribu-

tion. 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 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 $\lambda_{\min}$, Hotelling T^2, Goodall F, 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 $\lambda_{\min}$, 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 $i = 1, \ldots, k$. Assume that $n^{1/2}\hat{\mathbf{M}}_i\mathbf{m}_0$ has an asymptotic complex normal distribution $\mathrm{CN}_{d-1}(0, \mathbf{G}_i), i = 1, \ldots, k$, 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 $\chi^2_{2(k-1)(d-1)}$ under the null hypothesis of equality of means across populations (Amaral et al., 2007).
Another statistic used in this paper is the James statistic (see Seber, 1984) that represents an effort to solve the multivariate Behrens-Fisher problem and it is given by

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

where $\mathbf{v}_i \sim N(\xi_1, \mathbf{\Sigma}_1)$ for $i = 1, \ldots, n_1$ and $\mathbf{w}_j \sim N(\xi_2, \mathbf{\Sigma}_2)$, for $j = 1, \ldots, n_2$ are the partial Procrustes tangent coordinates, $\mathbf{v}_i$ and $\mathbf{w}_i$ are mutually independent, $\bar{\mathbf{v}}$, $\bar{\mathbf{w}}$ and $\mathbf{S}_1$, $\mathbf{S}_2$ are the sample means and sample covariance matrices (with divisors n_1 and n_2) in each group. It is proved that $F_J \sim \chi^2_M$. 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-dimensioned 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). We observe that a key condition for applying permutation tests is the exchangeability of observations under the null hypothesis (Pesarin, 2001). 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.

3.2 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 $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ 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 $X = \{X_{hji},\ i = 1, \ldots, n,\ j = 1, 2,\ h = 1, \ldots, km\}$, 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 $(a_1^*, \ldots, a_n^*)$ a permutation of the labels $(1, \ldots, n)$, $X^* = \{X^*_{hji} = X_{hj}(a_i^*),\ i = 1, \ldots, n,\ j = 1, 2,\ h = 1, \ldots, km\}$ is the related permutation of X, so that $X^*_{h1} = \{X^*_{h1i} = X_{h1}(a_i^*), i = 1, \ldots, n_1, h = 1, \ldots, km\}$ and $X^*_{h2} = \{X^*_{h2i} = X_{h2}(a_i^*), i = n_1 + 1, \ldots, n, h = 1, \ldots, km\}$ are the two permuted samples,

respectively.
For simplicity, we may assume that the landmark coordinates in tangent space behave according to the following model:

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

$i = 1, \ldots, n$, $j = 1, 2$, $h = 1, \ldots, km$, 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 δ_{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}, $i = 1, \ldots, n$, $j = 1, 2$, $h = 1, \ldots, km$, indicates a permutation of the original data.
Therefore the specific hypotheses may be expressed as

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

where $\stackrel{d}{>}$ stands for distribution (or stochastic) dominance.
With $T^o_h(0)$ 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 $\mathbf{T} = \{T_h, h = 1, \ldots, km\}$ necessary for nonparametric combination 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,

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

   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).

Let $\lambda_h, h = 1, \ldots, km$ 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, $h = 1, \ldots, km$ 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 Figure 3.1).

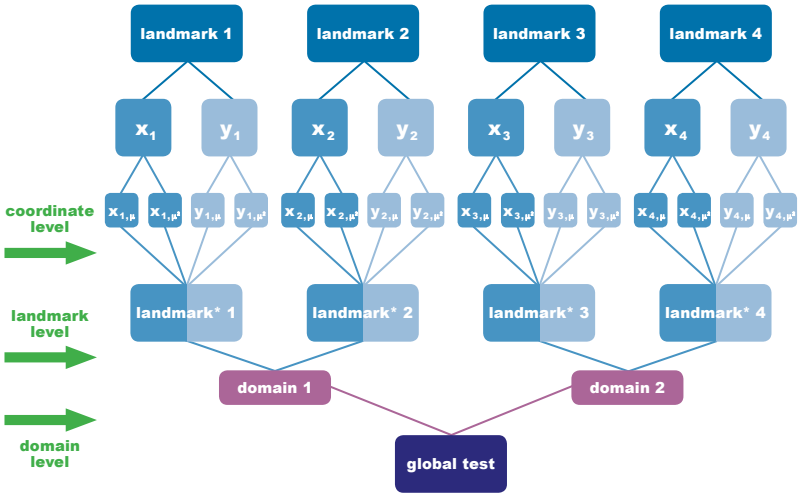


Figure 3.1: Different levels of combination.

3.2.1 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).

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

  (1.a) Calculate the vector of observed values of tests $\mathbf{T} : \mathbf{T}_o = \mathbf{T}(\mathbf{X})$.

(1.b)) 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 $(a_1^*, \ldots, a_n^*)$ of integers $(1, \ldots, n)$ and then by assignment of related individual data vectors to the proper group; thus, according to the unit-by-unit representation, $\mathbf{X}^* = \{\mathbf{X}(a_i^*), i = 1, \ldots, n; n_1, n_2\}$.

(1.c) Carry out B independent repetitions of step (b). The set of Conditional Monte Carlo (CMC) sampling results $\{\mathbf{T}_r^*, r = 1, \ldots, B\}$ is thus a random sampling from the permutation km-variate distribution of vector test statistics $\mathbf{T}$.

(1.d) The km-variate EDF $\hat{F}_B(\mathbf{z}|\mathbf{X}) = \left[\frac{1}{2} + \sum_r \mathbf{I}(\mathbf{T}_r^* \leq \mathbf{z})\right] /(B+1), \forall \mathbf{z} \in \mathcal{R}^{km}$, gives an estimate of the corresponding km-dimensional permutation distribution $F(\mathbf{z}|\mathbf{X})$ di $\mathbf{T}$. Moreover,

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

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 $\lambda^*{}_1$ is the permutation p-value corresponding to the x coordinate of landmark 1, $\lambda^*{}_2$ the permutation p-value corresponding to the y coordinate of landmark 1, $\lambda^*{}_3$ the permutation p-value corresponding to the x coordinate of landmark 2, $\lambda^*{}_4$ is the permutation p-value corresponding to the y coordinate of landmark 2 and so on (see Figure 3.2).

■ The second phase (*landmark level*) of the algorithm include the following steps.

(2.a) The km observed p-values are estimated from the data $\mathbf{X}$ by $\lambda_h = \hat{L}_h(T_{ho}|\mathbf{X})$, where $T_{ho} = T_h(\mathbf{X})$, $h = 1, \ldots, km$, represent the observed values of partial tests and $\hat{L}_h$ is the hth marginal significance level function, the latter being jointly estimated by the Conditional Monte Carlo (CMC) sampling method on data set $\mathbf{X}$, in accordance with step (1.d) above.

(2.b) 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 Figure 3.2. For example the observed statistic related to the first landmark (in 2D case), is given by

$$T_{1o}^{''} = \psi(\lambda_1, \lambda_2).$$

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

$$T''^{*}_{1r} = \psi(\lambda^{*}_{1r}, \lambda^{*}_{2r}),$$

where $\lambda^{*}_{1r} = \hat{L}_1(T^{*}_{1r}|\mathbf{X})$, $r = 1, \ldots, B$.
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.

(3.a) 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.

(3.b) The combined observed value of the third-order test is evaluated through the same CMC results of the second phase, and is given by

$$T'''_{Zo} = \psi(\lambda'_1, \ldots, \lambda'_Z).$$

corresponding to the first domain,

$$T'''_{Ao} = \psi(\lambda'_1, \ldots, \lambda'_A).$$

corresponding to the second domain, and

$$T'''_{Co} = \psi(\lambda'_1, \ldots, \lambda'_C).$$

corresponding to the third domain.

(3.c) The rth combined value of vector statistics is then calculated by

$$T'''^{*}_{Zr} = \psi(\lambda'^{*}_{1r}, \ldots, \lambda'^{*}_{Zr}),$$

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

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

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

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

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

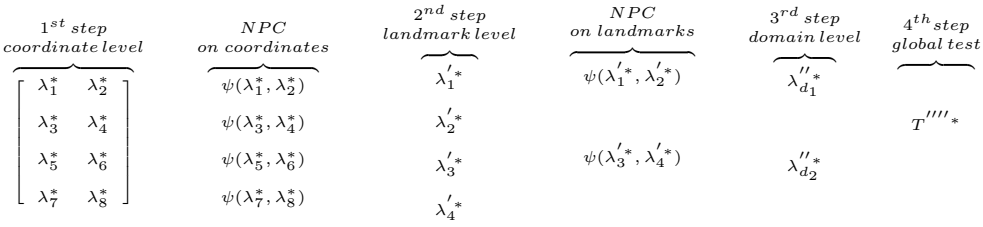


Figure 3.2: Algorithm for $k = 4$ landmarks in 2D and two domain combinations.

Hence at the end of this step we obtain different p-values corresponding to predefined domains. Figure 3.2 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.

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

$$T_o'''' = \psi(\lambda_1'^*, \lambda_2'^*, \lambda_Z''^*, \ldots, \lambda_A''^*, \ldots, \lambda_C''^*).$$

(4.b) The rth combined value of vector statistics (step ($\mathrm{S.d}_k$)) is then calculated by

$$T_r''''^* = \psi(\lambda_{1r}'^*, \lambda_{2r}'^*, \lambda_{Zr}''^*, \ldots, \lambda_{Ar}''^*, \ldots, \lambda_{Cr}''^*).$$

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

$$\lambda_\psi'''' = \textstyle\sum_r \mathbf{I}(T_r''''^* \geq T_o'''')/B.$$

(4.d) If $\lambda_\psi'''' \leq \alpha$, the global null hypothesis H_0 is rejected at significance level α.

3.2.2 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 $\mathbf{X}_j = \{X_{ji}, i = 1, \ldots, n_j\}$ be the data set of n_j elements related to the j-th sample or group, $j = 1, 2$. Let $\mathbf{X}_j^* = \{X_{ji}^*, i = 1, \ldots, n_j, j = 1, 2\}$ 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 $H_0 : \{X_1 \overset{d}{=} X_2\} = \{P_1 = P_2\}$ that the two groups have the same underlying distribution, against the global alternative hypothesis $H_1 : \{X_1 \overset{d}{<} X_2\}$ 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). H_0 may be broken down into

$$H_0 : \left\{ \bigcap_{i=1}^{K} H_{0i} \right\} \tag{3.1}$$

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

$$H_1 : \left\{ \bigcup_{i=1}^{K} H_{1i} \right\}. \tag{3.2}$$

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 be 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

$$\begin{aligned} H_{0l} &: \{[E(X_1) = E(X_2)] \cap [Me(X_1) = Me(X_2)]\} \\ H_{1l} &: \{[E(X_1) < E(X_2)] \cup [Me(X_1) < Me(X_2)]\}. \end{aligned} \tag{3.3}$$

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

$$\begin{aligned} H_{0d} &: \{F_1 = F_2\} \\ H_{1d} &: \{F_1 > F_2\} \end{aligned} \tag{3.4}$$

Applying the nonparametric combination 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

$$T^{*\prime\prime}_l = \max(1 - \lambda^*_\mu, 1 - \lambda^*_{Me}). \tag{3.5}$$

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

$$T^{*\prime\prime}_d = \max(1 - \lambda^*_{KS}, 1 - \lambda^*_{AD}). \tag{3.6}$$

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

$$T^{*\prime\prime\prime}_{MA} = \psi(\lambda^{*\prime\prime}_l, \lambda^{*\prime\prime}_d), \tag{3.7}$$

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, Mahalanobis, etc. For the selection of a combining function, see the practical guidelines set out in Pesarin (2001). We have mainly used Fisher omnibus combining function, calculated as

$$T''_F = \sum_i \log(\lambda_i).$$

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 continuos, 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 Figure 3.2, is calculated by considering also the aspects (see Figure 3.3).

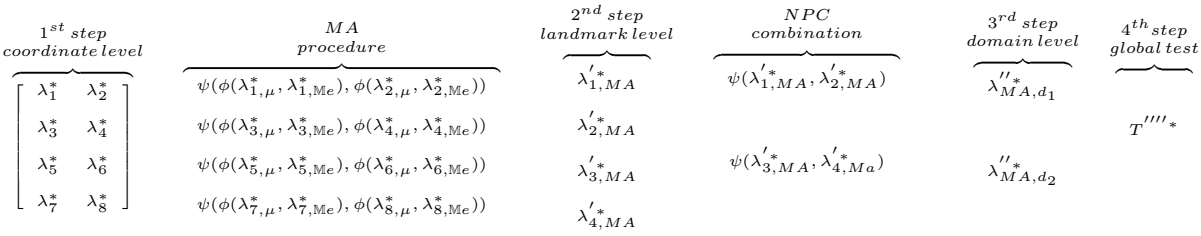


Figure 3.3: Algorithm for $k = 4$ landmarks in 2D, MA procedure (mean and median aspects) and domain combination. ψ and ϕ are suitable combining functions.

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 Figure 3.1). 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.

3.2.3 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 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 *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 to 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 Experimentwise 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 4 landmarks. Hence, with reference to the Figure 3.1 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 Figure 3.4 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, some other are only instrumental and are not in-

vestigated. 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 Figure 3.4). 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 $\max(0.0067, 0.0220, 0.0285, 0.0285, 0.0570, 0.0580, 0.0600, 0.0618) = 0.0618$.

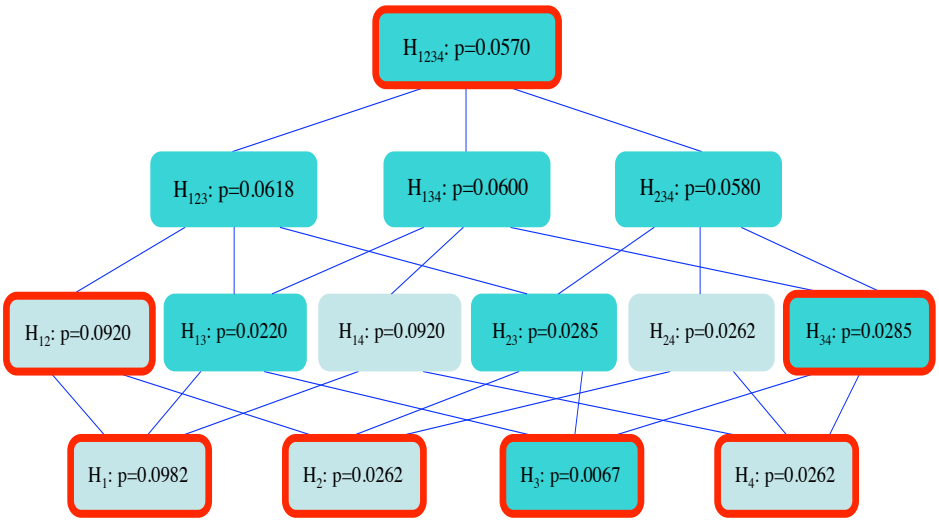


Figure 3.4: Illustration of the closed testing procedure, focussing on landmark level ($k = 4$).

3.3 Comparative simulation study

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:

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

$i = 1, \ldots, n_j$, $j = 1, 2$, $h = 1, \ldots, k$, 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). 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 Chapter 4.
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 3.1) with the same noncentrality parameter δ, even when the number of covariates (k) is larger than the permutation sample space (see results in Table 3.2 and Table 3.3).
We remark that Hotelling's T^2 test considered in Table 3.1 is computed using raw coordinates and not shape variables. Moreover, when $n_1 = n_2 = 10$ and $k = 19$, the 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 Chapter 5, thus introducing and extending the notion of "finite-sample consistency (FSC)", widely discussed in Chapter 6.

Table 3.1: Simulations under H_1 ($n_1 = n_2 = 10$, $\mu = 0$, $\delta = 0.40$, $B = CMC = 1000$)

		α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=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 3.2: Simulations under H_1 ($n_1 = n_2 = 3$, $\mu = 0$, $\delta = 0.40$, $B = MC = 1000$)

		α=0.10	α=0.20	α=0.30	α=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

Let us assume that our samples are made of configurations of $k = 8$ landmarks in $m = 2$ dimensions characterized by slightly different means. Suppose to deal with

Table 3.3: Simulations under H_1 ($n_1 = n_2 = 3$, $\mu = 0$, $\delta = 1$, $B = MC = 1000$)

		α=0.10	α=0.20	α=0.30	α=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

male and female skull configurations of a particular animal *ad hoc* created, representing the two independent sample (see Table 3.4). Since sample means slightly differ from each other in 6 over 16 coordinates (emphasized in red in Table 3.4), we have generated data using such configuration in order to evaluate the power of the competing tests.

Table 3.4: Hypothetical mean configurations

Domain	#	Lnd. name	Male		Female	
			x	y	x	y
1	1	nasion	65.00	223.00	65.00	**222.85**
1	2	basion	54.00	−40.00	**53.75**	−40.00
2	3	staphylion	0.00	0.00	0.00	0.00
2	4	prosthion	0.00	35.00	0.00	**34.50**
2	5	nariale	19.00	121.00	**18.90**	121.00
3	6	bregma	70.00	203.00	70.00	203.00
3	7	lambda	110.00	112.00	**109.95**	112.00
3	8	opisthion	104.00	17.00	104.00	**16.88**

In all the simulations we have set the number B of permutations equals to 1000 and the number CMC of Conditional Monte Carlo iterations equals to 1000. Throughout the simulation study, we have selected either Liptak and Fisher combining functions. But of course it is possible to take any other useful combining function into consideration, e.g. Tippett, Lancaster, Mahalanobis, etc. For the selection of a combining function, see the practical guidelines suggested by Pesarin (2001). Focussing on the two independent sample case, we have carried out the simulation study in the same conditions of homogeneous, independent, spherical variation at each landmark, as described in Rohlf (2000). In particular, we have compared, in terms of statistical power, traditional approaches for the statistical analysis of shape like Hotelling's T^2 test using approximate tangent space coordinates and Bookstein shape coordinates, Goodall's F test, `EDMA-I` (Lele and Richtsmeier, 1991) and `EDMA-II` tests (Lele and Cole, 1995; 1996), T^2-test using Rao and Suryawanshi shape variables, Rao-d (1996) and Rao-a (1998)(only for $n_1 = n_2 = 50$) to the following nonparametric permutation tests: (i) nonparametric permutation Hotelling's T^2, (ii) nonparametric permutation global tests with and without domains using

Liptak and Fisher combining functions, (iii) nonparametric permutation MA tests with and without domains, considering location and scale aspects and using Liptak and Fisher combining functions.
Let n_j, $j = 1, 2$, be the sample size in the two samples and $X_{hji}, i = 1, \ldots, n_j, j = 1, 2, h = 1, \ldots, k$ be the coordinates which are distributed according to a multivariate normal distribution.
Three domains have been considered, i.e. baseline (nasion and basion), face (staphylion, prosthion and nariale) and braincase (bregma, lambda and opisthion), have been examined. 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 (hence obtained after combining partial tests on chosen domains). Fisher, Liptak are the possible combining functions used and MA, if present, denotes the Multi Aspect procedure previously discussed. With regard to the comparative simulation study, 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: $n_1 = n_2 = 50$, $\sigma^2 = 0.25$. Simulation results are shown in Table 3.5 and Tables 3.11-3.13).
In Tables 3.5-3.10 we also display the simulation results for the tests proposed in the paper by Amaral et al. (2007), where p-values are obtained using both resampling methods (bootstrap or permutation test). Simulations are carried out both under the alternative and the null hypothesis. With 'lambda.pvalue' we denote the p-value for λ_{min} test based on resampling, while with 'lambda.table.pvalue' the p-value for λ_{min} test based on the asymptotic χ^2 distribution (large n_1, n_2). 'H.pvalue' indicates the p-value for the Hotelling T^2 test based on resampling and 'H.table.pvalue' indicates the p-value for the Hotelling T^2 test based on the null F distribution, assuming normality and equal covariance matrices. 'J.pvalue' stands for the p-value associated to the Hotelling T^2 test based on resampling while 'J.table.pvalue' stands for the p-value associated to the Hotelling T^2 test based on the null F distribution, assuming normality and unequal covariance matrices. Furthermore, 'G.pvalue' denotes the p-value for the Goodall test based on resampling and 'G.table.pvalue' the p-value for the Goodall test based on the null F distribution, assuming normality and equal isotropic covariance matrices. The symbol * indicates that the bootstrap resampling is carried out in each group; if * is not present, then permutation resampling is carried without replacement from the pooled samples. In order to calculate the four above mentioned test statistics ($\lambda_{\min}$, Hotelling T^2, Goodall F, James F_J tests) we used the `resampletest` function in the R (R Development Core Team, 2008) package `shapes`.
For these tests we report the results for small sample size ($n_1 = n_2 = 10$, $\sigma^2 = 0.25$) as well as for larger sample sizes, i.e. $n_1 = n_2 = 20$, $n_1 = n_2 = 50$ and $\sigma^2 = 0.25$. In Table 3.5 and Tables 3.11-3.13, the tests performing better are highlighted in bold and in glow yellow. In particular we show that the Hotelling's T^2 permutation counterpart, the Goodall's F test, EDMA I, the global nonparametric test using Fisher combining function, in its standard, domain and MA versions, are powerful in almost all situations.
As shown in Table 3.5, 'lambda.pvalue' bootstrap test, 'lambda.table.pvalue', 'J.table.

pvalue', 'G.pvalue' tests and 'G.table.pvalue' permutation test seem to be powerful at first glance, but examining simulation results under the null hypothesis (Table 3.6), some of these tests (in particular 'lambda.table.pvalue' and 'J.table.pvalue' tests), again highlighted in bold and in glow yellow, have type I error rates which are out of control. Moreover, 'H.pvalue' and 'J.pvalue' bootstrap tests, highlighted in orange in Table 3.5 cannot be estimated (power function is constantly equal to zero) Their achieved significance levels are also equal to 0 (see Table 3.6). Amaral et al. (2007) state that this behaviour is not due to the bootstraap method itself but to the low sample size, that is very small compared to the number of landmarks. They suggested to try the method with larger samples (e.g. n=20 or 30 for each group). In Tables 3.7-3.10 we show results for the cases $n_1 = n_2 = 20$ and $n_1 = n_2 = 50$. Hence, increasing the sample size, in the case $n_1 = n_2 = 20$, under H_1, 'lambda.pvalue' bootstrap test, 'lambda.table.pvalue' and 'J.table.pvalue' tests, 'G.pvalue' and 'G.table.pvalue' bootstrap tests have a good performance (they are highlighted in bold and glow yellow, see Table 3.7). Focussing on nominal α-level equal to 5% we show that 'lambda.table.pvalue' and 'J.table.pvalue' tests, again highlighted in bold and in glow yellow, have type I error rates which are out of control (see Table 3.8). In the same table, 'H.pvalue' and 'J.pvalue' bootstrap tests, highlighted in orange, are very conservative (but their achieved significance levels are now different from 0). Finally, in the case $n_1 = n_2 = 50$, 'lambda.pvalue', 'H.pvalue' and 'J.pvalue' bootstrap tests, 'lambda.table.pvalue' and 'J.table.pvalue' tests have a good performance (they are highlighted in bold and glow yellow, see Table 3.9). Focussing on nominal level $\alpha = 0.05$, in Table 3.10, 'lambda.table.pvalue' tests and 'J.table.pvalue' permutation tests, highlighted in bold and in glow yellow, have type I error rates which are out of control (see Table 3.7). 'H.pvalue' and 'J.pvalue' bootstrap tests are still conservative. We can conclude that the tests proposed by Amaral et al. (2007) need larger sample sizes than the nonparametric permutation methods. Our simulation study emphasize not only the good behaviour in power but also the flexibility of the nonparametric permutation solution in shape analysis, since it allows to carry out a shape analysis even in presence of small sample sizes and a large number of shape variables. We wish to highlight that NPC tests control the Type I error (see Tables 3.14-3.17).
In the simulation study we have considered α levels varying from 1% to 50% in order to explore power behavior, thus following a procedure similar to that leading to the construction of Receiver Operating Characteristic (ROC) curves.

Table 3.5: 1st simulation ($n_1 = n_2 = 10$, B=CMC=1000, $\sigma^2 = 0.25$, *=bootstrap test)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
T^2, Approx Proc Tg	0.026	0.105	0.200	0.367	0.495	0.730
T^2, perm	**0.099**	**0.270**	**0.392**	**0.555**	**0.658**	**0.834**
T^2, Bookstein SC	0.022	0.098	0.186	0.336	0.528	0.750
Goodall F-test	**0.106**	**0.250**	**0.398**	**0.520**	**0.622**	**0.812**
EDMA-I	**0.058**	**0.134**	**0.266**	**0.428**	**0.572**	**0.836**
EDMA-II	0.014	0.028	0.040	0.130	0.246	0.432
G (Liptak)	0.049	0.171	0.281	0.461	0.581	0.766
G_d (Liptak)	0.040	0.168	0.275	0.436	0.567	0.743
G (Fisher)	**0.072**	**0.233**	**0.365**	**0.544**	**0.650**	**0.810**
G_d (Fisher)	0.067	0.216	0.343	0.512	0.619	0.787
G (Liptak, MA)	0.057	0.176	0.303	0.458	0.586	0.749
G_d (Liptak, MA)	0.052	0.177	0.289	0.442	0.566	0.739
G (Fisher, MA)	**0.075**	**0.235**	**0.365**	**0.533**	**0.659**	**0.801**
G_d (Fisher, MA)	0.065	0.211	0.344	0.508	0.630	0.780
lambda.pvalue	0.068	0.214	0.310	0.488	0.604	0.766
	0.028*	**0.146***	**0.292***	**0.474***	**0.624***	**0.826***
lambda.table.pvalue	**0.286**	**0.466**	**0.562**	**0.716**	**0.788**	**0.878**
	0.294*	**0.462***	**0.574***	**0.688***	**0.778***	**0.898***
H.pvalue	0.028	0.116	0.206	0.354	0.518	0.714
	0.000*	**0.000***	**0.000***	**0.000***	**0.000***	**0.000***
H.table.pvalue	0.028	0.118	0.218	0.364	0.504	0.714
	0.024*	0.100*	0.186*	0.340*	0.492*	0.702*
J.pvalue	0.028	0.116	0.206	0.354	0.518	0.714
	0.000*	**0.000***	**0.000***	**0.000***	**0.000***	**0.000***
J.table.pvalue	**0.394**	**0.580**	**0.664**	**0.778**	**0.820**	**0.904**
	0.364*	**0.552***	**0.650***	**0.750***	**0.812***	**0.908***
G.pvalue	**0.094**	**0.248**	**0.386**	**0.540**	**0.650**	**0.820**
	0.014*	**0.100***	**0.226***	**0.424***	**0.598***	**0.808***
G.table.pvalue	**0.096**	**0.250**	**0.388**	**0.540**	**0.656**	**0.814**
	0.072*	0.220*	0.328*	0.488*	0.616*	0.794*

Table 3.6: Simulation considering tests proposed by Amaral et al. (2007) under H_0: $n_1 = n_2 = 10$, $\sigma^2 = 0.25$ (*=bootstrap test)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
lambda.pvalue	0.012	0.040	0.094	0.192	0.286	0.498
	0.004*	0.036*	**0.078***	0.200*	0.316*	0.576*
lambda.table.pvalue	**0.076**	**0.184**	**0.264**	**0.398**	**0.492**	**0.668**
	0.076*	**0.192***	**0.250***	**0.388***	**0.492***	**0.678***
H.pvalue	0.006	0.046	0.080	0.188	0.288	0.456
	0.000*	**0.000***	**0.000***	**0.000***	**0.000***	**0.000***
H.table.pvalue	0.01	0.048	0.088	0.198	0.286	0.460
	0.008*	0.058*	0.108*	0.210*	0.310*	0.510*
J.pvalue	0.006	0.046	0.080	0.188	0.288	0.456
	0.000*	**0.000***	**0.000***	**0.000***	**0.000***	**0.000***
J.table.pvalue	**0.216**	**0.334**	**0.422**	**0.534**	**0.632**	**0.776**
	0.236*	**0.364***	**0.460***	**0.578***	**0.656***	**0.774***
G.pvalue	0.010	0.046	0.082	0.184	0.286	0.504
	0.002*	0.014*	0.046*	0.128*	0.274*	0.546*
G.table.pvalue	0.012	0.036	0.084	0.184	0.296	0.492
	0.008*	0.042*	0.078*	0.170*	0.292*	0.488*

Table 3.7: Increasing sample size in tests proposed in Amaral et al. (2007) under H_1: $n_1 = 20, n_2 = 20$, $\sigma^2 = 0.25$(*=bootstrap test)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
lambda.pvalue	0.234	0.504	0.638	0.802	0.858	0.926
	0.248*	**0.560***	**0.680***	**0.814***	**0.898***	**0.962***
lambda.table.pvalue	**0.546**	**0.758**	**0.836**	**0.896**	**0.924**	**0.960**
	0.546*	**0.734***	**0.814***	**0.904***	**0.938***	**0.980***
H.pvalue	0.170	0.418	0.548	0.706	0.818	0.924
	0.010*	0.116*	0.280*	0.516*	0.696*	0.880*
H.table.pvalue	0.176	0.418	0.546	0.710	0.816	0.924
	0.170*	0.398*	0.542*	0.732*	0.832*	0.926*
J.pvalue	0.170	0.418	0.548	0.706	0.818	0.924
	0.010*	0.116*	0.280*	0.516*	0.696*	0.880*
J.table.pvalue	**0.336**	**0.544**	**0.662**	**0.796**	**0.876**	**0.942**
	0.320*	**0.534***	**0.678***	**0.808***	**0.872***	**0.944***
G.pvalue	0.286	0.526	0.662	0.800	0.880	0.936
	0.142*	**0.444***	**0.620***	**0.778***	**0.866***	**0.954***
G.table.pvalue	0.302	0.540	0.662	0.798	0.882	0.938
	0.288*	**0.542***	**0.678***	**0.794***	**0.866***	**0.956***

Table 3.8: Increasing sample size in tests proposed in Amaral et al. (2007) under H_0: $n_1 = 20, n_2 = 20$, $\sigma^2 = 0.25$(*=bootstrap test)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
lambda.pvalue	0.006	0.034	0.090	0.180	0.274	0.496
	0.008*	0.046*	0.102*	0.250*	0.388*	0.606*
lambda.table.pvalue	**0.052**	**0.146**	**0.232**	**0.342**	**0.480**	**0.684**
	0.042*	**0.134***	**0.236***	**0.374***	**0.504***	**0.660***
H.pvalue	0.010	0.048	0.102	0.178	0.276	0.508
	0.000*	**0.006***	**0.020***	**0.072***	**0.174***	**0.386***
H.table.pvalue	0.010	0.048	0.098	0.180	0.278	0.500
	0.006*	0.042*	0.086*	0.196*	0.306*	0.488*
J.pvalue	0.010	0.048	0.102	0.178	0.276	0.508
	0.000*	**0.006***	**0.020***	**0.072***	**0.174***	**0.386***
J.table.pvalue	**0.034**	**0.098**	**0.148**	**0.256**	**0.378**	**0.576**
	0.024*	**0.086***	**0.156***	**0.282***	**0.382***	**0.558***
G.pvalue	0.008	0.032	0.078	0.180	0.288	0.496
	0.000*	0.016*	0.054*	0.158*	0.294*	0.514*
G.table.pvalue	0.006	0.030	0.078	0.178	0.286	0.492
	0.004*	0.026*	0.076*	0.180*	0.310*	0.490*

Table 3.9: Increasing sample size in tests proposed in Amaral et al. (2007) under H_1: $n_1 = 50, n_2 = 50$, $\sigma^2 = 0.25$(*=bootstrap test)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
lambda.pvalue	0.834	0.938	0.968	0.988	0.996	1.000
	0.736*	**0.906***	**0.964***	**0.984***	**0.992***	**0.998***
lambda.table.pvalue	**0.898**	**0.968**	**0.984**	**0.990**	**1.000**	**1.000**
	0.882*	**0.960***	**0.984***	**0.992***	**0.994***	**1.000***
H.pvalue	0.776	0.924	0.960	0.982	0.994	1.000
	0.686*	**0.888***	**0.952***	**0.976***	**0.992***	**0.996***
H.table.pvalue	0.786	0.934	0.964	0.982	0.992	1.000
	0.754*	0.910*	0.960*	0.986*	0.992*	0.998*
J.pvalue	0.776	0.924	0.960	0.982	0.994	1.000
	0.686*	**0.888***	**0.952***	**0.976***	**0.992***	**0.996***
J.table.pvalue	**0.810**	**0.938**	**0.964**	**0.984**	**0.994**	**1.000**
	0.782*	**0.916***	**0.966***	**0.988***	**0.992***	**0.998***
G.pvalue	0.848	0.954	0.978	0.990	0.998	1.000
	0.782*	0.944*	0.970*	0.988*	0.992*	1.000*
G.table.pvalue	0.854	0.962	0.982	0.992	0.998	1.000
	0.838*	0.952*	0.972*	0.988*	0.992*	1.000*

Table 3.10: Increasing sample size in tests proposed in Amaral et al. (2007) under H_0: $n_1 = 50, n_2 = 50$, $\sigma^2 = 0.25$(*=bootstrap test)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
lambda.pvalue	0.020	0.058	0.104	0.178	0.298	0.512
	0.010*	0.042*	0.092*	0.186*	0.302*	0.498*
lambda.table.pvalue	**0.026**	**0.090**	**0.144**	**0.252**	**0.366**	**0.568**
	0.020*	**0.088***	**0.144***	**0.262***	**0.374***	**0.552***
H.pvalue	0.022	0.062	0.112	0.190	0.280	0.496
	0.008*	0.040*	0.074*	0.182*	0.272*	0.474*
H.table.pvalue	0.022	0.064	0.108	0.186	0.278	0.494
	0.008*	0.050*	0.088*	0.194*	0.294*	0.494*
J.pvalue	0.022	0.062	0.112	0.190	0.280	0.496
	0.008*	0.040*	0.074*	0.182*	0.272*	0.474*
J.table.pvalue	**0.022**	**0.070**	**0.118**	**0.196**	**0.288**	**0.504**
	0.008*	0.052*	0.096*	0.202*	0.302*	0.502*
G.pvalue	0.016	0.056	0.098	0.188	0.306	0.496
	0.010*	0.036*	0.082*	0.192*	0.318*	0.496*
G.table.pvalue	0.020	0.058	0.096	0.188	0.314	0.502
	0.012*	0.040*	0.088*	0.194*	0.316*	0.494*

Table 3.11: 2nd simulation ($n_1 = n_2 = 10$, B=CMC=1000, $\sigma^2 = 0.50$)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
T^2, Approx Proc Tg	0.009	0.071	0.144	0.291	0.420	0.624
T^2, perm	**0.045**	**0.160**	**0.263**	**0.401**	**0.524**	**0.694**
T^2, Bookstein SC	0.012	0.068	0.158	0.302	0.434	0.578
Goodall F-test	**0.052**	**0.174**	**0.226**	**0.392**	**0.538**	**0.696**
EDMA-I	**0.034**	**0.136**	**0.178**	**0.376**	**0.482**	**0.682**
EDMA-II	0.016	0.024	0.050	0.106	0.212	0.430
G (Liptak)	0.027	0.108	0.186	0.339	0.443	0.648
G_d (Liptak)	0.020	0.097	0.183	0.333	0.438	0.630
G (Fisher)	**0.034**	**0.134**	**0.227**	**0.369**	**0.498**	**0.689**
G_d (Fisher)	0.036	0.120	0.210	0.356	0.485	0.675
G (Liptak, MA)	0.022	0.109	0.206	0.345	0.446	0.651
G_d (Liptak, MA)	0.021	0.098	0.193	0.334	0.438	0.647
G (Fisher, MA)	**0.032**	**0.142**	**0.242**	**0.372**	**0.493**	**0.695**
G_d (Fisher, MA)	0.028	0.126	0.221	0.356	0.491	0.678

Table 3.12: 3rd simulation: $n_1 = 50, n_2 = 20$, $\sigma^2 = 0.25$

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
T^2, Approx Proc Tg	0.372	0.630	0.757	0.849	0.905	0.971
Goodall's F	0.475	0.708	0.786	0.886	0.933	0.972
T^2, perm	**0.511**	**0.789**	**0.869**	**0.940**	**0.968**	**0.987**
G (Tippett)	**0.000**	**0.787**	**0.870**	**0.932**	**0.959**	**0.987**
G_d (Tippett)	**0.088**	**0.761**	**0.854**	**0.923**	**0.955**	**0.985**
G (Fisher)	**0.385**	**0.677**	**0.797**	**0.907**	**0.953**	**0.990**
G_d (Fisher)	**0.337**	**0.634**	**0.771**	**0.887**	**0.940**	**0.984**
G (Tippett, MA)	0.488	0.690	0.811	0.889	0.931	0.971
G_d (Tippett, MA)	0.476	0.686	0.791	0.878	0.920	0.965
G (Fisher, MA)	**0.564**	**0.775**	**0.863**	**0.920**	**0.954**	**0.982**
G_d (Fisher, MA)	**0.549**	**0.767**	**0.853**	**0.911**	**0.948**	**0.982**

Table 3.13: 4th simulation ($n_1 = n_2 = 50$, B=CMC=1000, $\sigma^2 = 0.25$)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
T^2, Approx Proc Tg	**0.785**	**0.936**	**0.967**	**0.988**	**0.997**	**0.999**
Goodall's F	**0.869**	**0.954**	**0.981**	**0.990**	**0.996**	**0.999**
T^2, perm	**0.938**	**0.985**	**0.992**	**0.996**	**0.998**	**0.999**
T^2, Bookstein sh. c.	**0.798**	**0.910**	**0.970**	**0.992**	**0.996**	**0.998**
Goodall F-test	**0.880**	**0.926**	**0.982**	**0.996**	**0.996**	**0.998**
T^2, Rao-d	**0.476**	**0.686**	**0.808**	**0.922**	**0.964**	**0.998**
T^2, Rao-a	**0.800**	**0.914**	**0.966**	**0.994**	**0.996**	**0.998**
EDMA-I	**0.504**	**0.720**	**0.862**	**0.966**	**0.972**	**0.998**
EDMA-II	0.014	0.038	0.060	0.180	0.268	0.484
G (Liptak)	0.354	0.644	0.759	0.867	0.916	0.972
G_d (Liptak)	0.332	0.615	0.752	0.856	0.912	0.966
G (Fisher)	0.677	0.873	0.944	0.976	0.989	1.000
G_d (Fisher)	**0.641**	**0.855**	**0.922**	**0.970**	**0.985**	**0.998**
G (Liptak, MA)	0.359	0.630	0.762	0.877	0.920	0.979
G_d (Liptak, MA)	0.340	0.619	0.755	0.864	0.909	0.973
G (Fisher, MA)	**0.678**	**0.883**	**0.939**	**0.977**	**0.991**	**0.999**
G_d (Fisher, MA)	**0.651**	**0.863**	**0.925**	**0.971**	**0.985**	**0.999**

Table 3.14: Simulation considering only permutation tests under H_0: $n_1 = n_2 = 10$, $\sigma^2 = 0.25$

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=0.50
G (Liptak)	0.012	**0.053**	0.100	0.191	0.304	0.506
G_d (Liptak)	0.008	**0.058**	0.105	0.200	0.298	0.530
G (Fisher)	0.007	**0.042**	0.100	0.198	0.317	0.513
G_d (Fisher)	0.009	**0.047**	0.111	0.218	0.314	0.515
G (Liptak, MA)	0.011	**0.053**	0.106	0.208	0.308	0.512
G_d (Liptak, MA)	0.010	**0.053**	0.120	0.211	0.296	0.518
G (Fisher, MA)	0.007	**0.047**	0.104	0.210	0.315	0.513
G_d (Fisher, MA)	0.007	**0.039**	0.106	0.210	0.320	0.512

Table 3.15: Simulation considering only permutation tests under H_0: $n_1 = n_2 = 20$, $\sigma^2 = 0.25$

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=0.50
G (Liptak)	0.011	**0.036**	0.081	0.166	0.288	0.484
G_d (Liptak)	0.011	**0.038**	0.085	0.168	0.274	0.490
G (Fisher)	0.005	**0.041**	0.086	0.182	0.271	0.494
G_d (Fisher)	0.004	**0.043**	0.090	0.183	0.280	0.497
G (Liptak, MA)	0.011	**0.039**	0.079	0.175	0.265	0.483
G_d (Liptak, MA)	0.010	**0.037**	0.080	0.174	0.260	0.490
G (Fisher, MA)	0.005	**0.039**	0.076	0.176	0.281	0.494
G_d (Fisher, MA)	0.003	**0.041**	0.085	0.179	0.279	0.485

Table 3.16: Simulation considering only permutation tests under H_0: $n_1 = 50, n_2 = 20$, $\sigma^2 = 0.25$

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=0.50
G (Liptak)	0.009	**0.040**	0.091	0.183	0.280	0.483
G_d (Liptak)	0.007	**0.039**	0.082	0.181	0.291	0.477
G (Fisher)	0.006	**0.040**	0.094	0.181	0.294	0.489
G_d (Fisher)	0.008	**0.042**	0.083	0.184	0.281	0.496
G (Liptak, MA)	0.006	**0.041**	0.083	0.183	0.287	0.479
G_d (Liptak, MA)	0.007	**0.044**	0.084	0.183	0.283	0.482
G (Fisher, MA)	0.007	**0.041**	0.093	0.176	0.283	0.481
G_d (Fisher, MA)	0.007	**0.044**	0.082	0.168	0.283	0.476

Table 3.17: Simulation considering only permutation tests under H_0: $n_1 = n_2 = 50$, $\sigma^2 = 0.25$

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=0.50
G (Liptak)	0.007	**0.042**	0.085	0.188	0.278	0.485
G_d (Liptak)	0.007	**0.041**	0.084	0.186	0.284	0.488
G (Fisher)	0.009	**0.043**	0.089	0.198	0.291	0.487
G_d (Fisher)	0.008	**0.046**	0.088	0.203	0.284	0.496
G (Liptak, MA)	0.007	**0.041**	0.079	0.191	0.274	0.475
G_d (Liptak, MA)	0.010	**0.042**	0.091	0.190	0.283	0.483
G (Fisher, MA)	0.013	**0.047**	0.081	0.179	0.291	0.491
G_d (Fisher, MA)	0.012	**0.044**	0.092	0.191	0.292	0.495

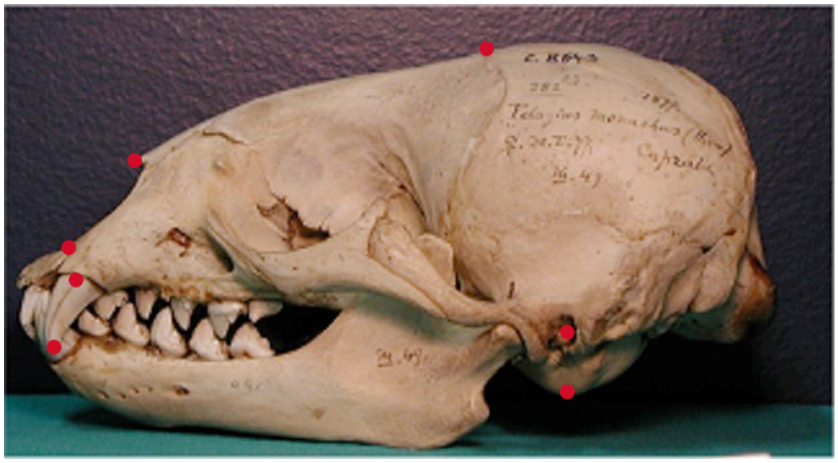


Figure 3.5: Landmarks: au, ms, b, rhi, ns, pr, cb and ct.

3.4 An application

We recall the Mediterranean monk seal skulls case-study. Information about sex and age class category of the specimens are summarized in Table 3.18, where 'nk' stands for those animals without any classification for sex and/or age.
As we showed before, combining shape analysis with nonparametric permutation techniques could be useful when dealing with small samples.

Table 3.18: Monachus monachus skull data

	Sex		
Age category	M	F	nk
3	-	-	2
4	1	1	-
7	-	2	2
8	-	2	-
9	2	-	-
nk	1	-	4

We have applied the NPC methodology previously discussed, in order to detect differences between age class categories. In such a way, we can obtain a sample of 12 specimens, 4 of those are considered young (i.e. the age class category is less or equal to 4) and 8 are adult (i.e. the age class category is greater than 4). We have chosen 8 anatomical landmarks in 2D. In particular auriculare (au), mastoidale (ms), bregma (b), rhinion (rhi), nasospinale (ns), prosthion (pr), canine base (cb) and canine tip (ct) as it is shown in Figure 3.5.

We have also defined two domains: the *midsaggittal and nasal* (b,rhi,ns,pr) and the *canine tooth* (cb,ct) If we carried out the traditional Goodall's F test we obtain a p-value equals to zero and the reference distribution is an F distribution with

M and $(n_1 + n_2 - 2)M$ d.f., where $M = 2k - 4$, i.e. $F_{12,120}$. Even reducing the number of relevant landmark, it is not possible to calculate Hotelling T^2 test because the reference distribution is an F distribution with M and $n_1 + n_2 - M - 1$ d.f., that is $F_{12,-1}$. Using the NPC methodology we have obtained the following p-values: $p = 0.0027$ for the auriculare landmark point, $p = 0.0027$ for the mastoidale, $p = 0.0027$ of for the first domain and $p = 0.0191$ for the second domain. The global p-value is equal to $p = 0.0027$. Hence even if we cannot find sexual dimorphism in our sample, we are able to find a significant difference between young and adult specimens studying landmarks and combinations with domains other then evaluating the global test.

3.5 Conclusions

The proposed nonparametric permutation approach performs better than the traditional tests used in shape analysis in terms of power. Due to their nonparametric nature, the suggested tests may be computed even when the number of covariates exceeds the number of cases. Moreover they do not rely on a well-defined distributional model, they have the obvious advantage of not requiring the assumption of homogeneity of variance, they deal with one-sided as well as two-sided tests of hypotheses and allow a more flexible analysis even in terms of the nature of the variables (continuous, categorical or mixed) involved in the analysis. With reference to the problem of small sample sizes, it is worth noting that the results obtained in a NPC framework can be extended to the corresponding reference population.
It is proved that it is possible to extend the permutation conditional inferences to unconditional or population inferences since permutation tests are provided with similarity and conditional unbiasedness properties (Pesarin, 2002). We note that in conditional inference, the reference space is obtained by considering the restriction of $\mathcal{X}$ to the sub-space associated with the conditioning event of interest, which may be denoted by A. Of course A must be a member of a suitable algebra of events $\mathcal{A}$ associated with $\mathcal{X}$, so that $A \in \mathcal{A}$. The symbol $\mathcal{X}_{/A}$ is used to denote the restriction of $\mathcal{X}$ induced by A; it contains the points of $\mathcal{X}$ which possess a given property associated with A, i.e. the coset of A. Of course, we assume that associated with $(X, \mathcal{X}_{/A})$ is a conditional distribution P_A.
If we are mainly interested in this subset, i.e we are not interested in extending conditional inference to the unconditional one, then no general problem arises. On the contrary, if we are interested in extending our inference to the entire target population, then we may be faced with quite difficult or even impossible problems. 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. But in practice this situation does not always occur and parametric inferential extensions might be wrong or misleading.
Otherwise 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. We refer the reader to Pesarin (2002) for details. Until now we focussed on the case of homogeneous, independent, spherical variation at each landmark. Hence, we have evalutated the power of the nonparametric permutation methods under the conditions in which parametric tests based on T^2 or F statistics and those based on interlandmark distances have a recognized better performance. Through the simulation study and the application, we have highlighted not only the power of these methods, but also that they enable the researcher to give local assessment using a combination with domains.
The case of heterogeneous and dependent variation at each landmark (nonzero covariance) is faced in Chapter 4.

3.A Appendix: some technical details on tests known in shape analysis literature

3.A.1 Hotelling's T^2 and Goodall's F tests

Let define two independent random samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ 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 $v_1, \ldots, v_{n_1}$ and $w_1, \ldots, w_{n_2}$ be the partial Procrustes tangent coordinates, where

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

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.

3.A.2 Euclidean Distance Matrix Analysis (EDMA) methods

The form of an object X is all the geometrical information about X that is invariant under traslation 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 $X_1, X_2, \ldots, X_n$ 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 $s^2 = n^{-1} \sum_{i=1}^{n} (e_{lm,i} - \bar{e}_{lm})^2$,
- $\hat{\delta}_{lm} = (\bar{e}_{lm}^2 - s^2(e_{lm}))^{0.5}$.
- $\mathbf{FM}(X) = (\hat{\delta}_{lm}^{0.5})_{lm=1,2,\ldots,k}$

In the same way the average form matrix $\mathbf{FM}(Y)$ of the sample $Y_1, Y_2, \ldots, Y_n$ 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

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

where $\mathbf{FM}(X)$ and $\mathbf{FM}(Y)$ are the 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.

Chapter 4

Evaluating NPC test power behaviour with GPA superimposition and correlation

4.1 GPA superimposition and NPC tests power

In Chapter 3, 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 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 behaviour is associated with transformations induced by GPA. Actually, including GPA, NPC tests are approximate, since GPA superimposition provides permutationally non-equivalent transformations. 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.

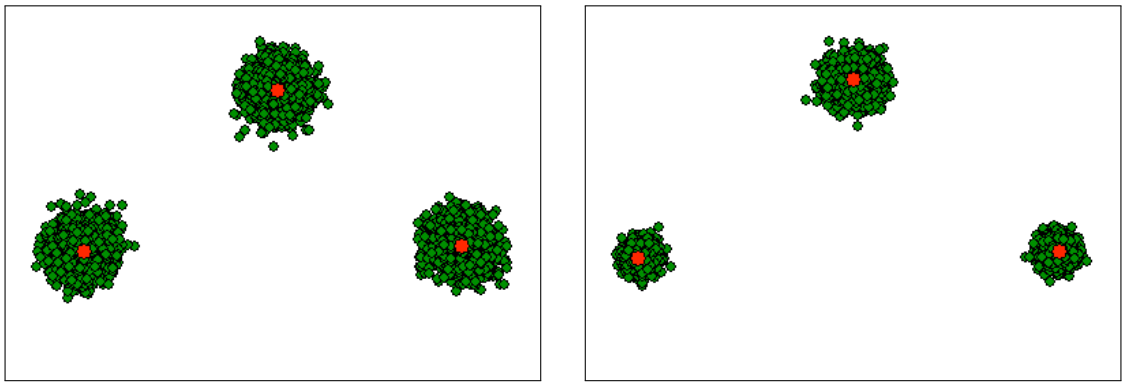
We could regard GPA superimposition as a method for standardizing shapes. It is well known that different results may be obtained using standardized or original data. For example, in multivariate statistics, this situation occurs in principal component analysis (PCA). Actually, the components obtained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

We recall this issue in Section 4.2, examining simulation results in presence of correlation between landmarks. In Figure 4.1 we show the effects of GPA superimposition. The scatterplot has been realized using the tutorial program TPStri (Rohlf,

2008b), allowing either to show some of the relationships between shape coordinates and to perform sampling experiments for triangles. In particular we have generated and plotted 2000 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 Figure 4.14.1(a) we have displayed the raw scatter, while in Figure 4.14.1(b) 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 behaviour.
Here we propose the same simulation study presented in Chapter 3, considering only NPC tests. We compare their power behaviour, 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), 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 induced by GPA (see the rows highlighted in darker blue in Tables 4.1-4.4).
The type I error is under control (see the column of nominal level $\alpha = 0.05$ highlighted in glow yellow in Tables 4.5-4.8). 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.



(a) (b)

Figure 4.1: Raw data (a) and superimposed data (b)

Table 4.1: 1st simulation ($n_1 = n_2 = 10$, B=CMC=1000, $\sigma^2 = 0.25$)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=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 4.2: 2nd simulation ($n_1 = n_2 = 10$, B=CMC=1000, $\sigma^2 = 0.50$)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=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 4.3: 3rd simulation: $n_1 = 50, n_2 = 20$, $\sigma^2 = 0.25$

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=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 4.4: 4th simulation ($n_1 = n_2 = 50$, B=MC=1000, $\sigma^2 = 0.25$)

	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=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 4.5: Simulations under H_0 ($n_1 = n_2 = 10$, B=CMC=1000, $\sigma^2 = 0.25$) using GPA

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=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 4.6: Simulations under H_0 ($n_1 = n_2 = 10$, B=CMC=1000, $\sigma^2 = 0.50$) using GPA

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=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 4.7: Simulations under H_0 ($n_1 = 50$, $n_2 = 20$, B=CMC=1000, $\sigma^2 = 0.25$) using GPA

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=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

Table 4.8: Simulations under H_0 ($n_1 = n_2 = 50$, B=CMC=1000, $\sigma^2 = 0.25$) using GPA

	α=0.01	**α=0.05**	α=0.10	α=0.20	α=0.30	α=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

4.2 Introducing correlation between landmarks

In this nonparametric framework we also analyze the case of heterogeneous and dependent 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. In order to obtain non singular 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 $\lambda^{1/10}$, rescaled by their trace (see the effect of transforming eigenvalues on the scatterplot in Figure 4.2). 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 4.9 we display hypothetical mean configurations, representing 3D male and female *Macaca fascicularis* monkey skulls (for details, see Frost et al. (2003)).

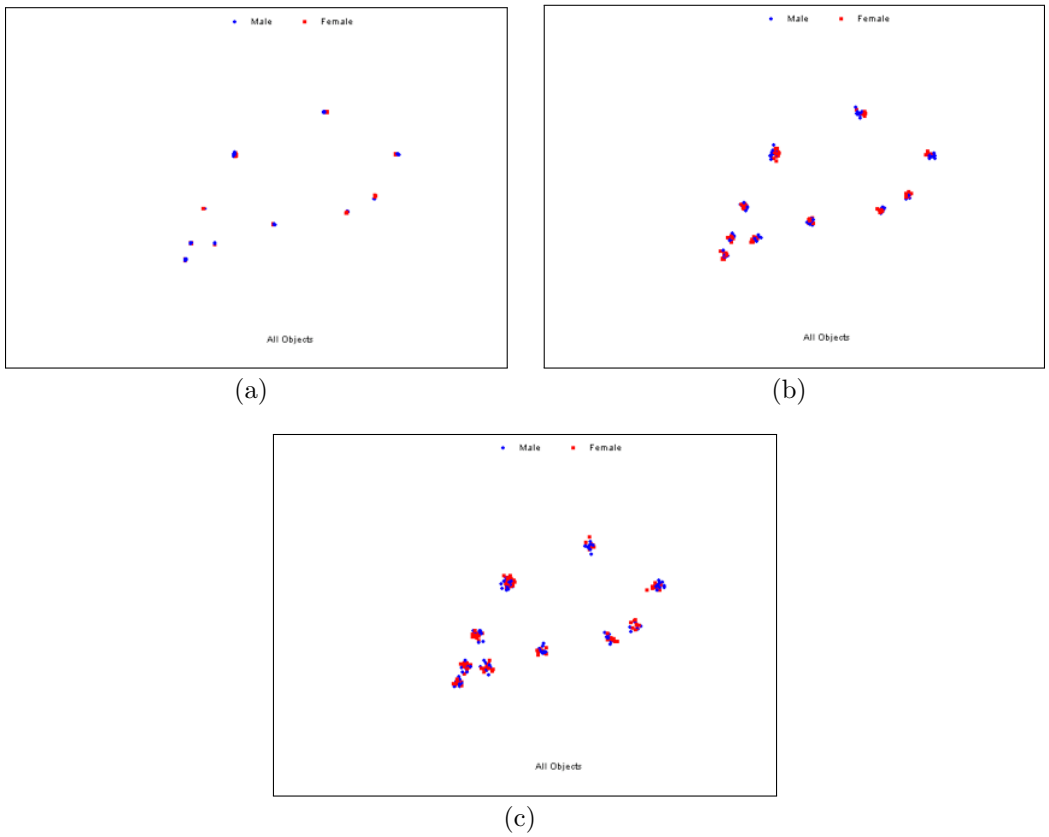


Figure 4.2: Original eigenvalues λ (a), $\lambda^{1/3}$ (b) and $\lambda^{1/10}$ (c)

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

Table 4.9: 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

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.
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 behaviour 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 ith specimen is given by

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

where α_i is a scale factor (size of the ith specimen), μ is the mean shape, $\mathbf{E}_i$ is a matrix of random errors (normally distributed with zero mean), Ω_i is a $m \times m$ matrix describing the orientation of the ith 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, Ω_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 $\Sigma = \Sigma_k \otimes \Sigma_m$, where Σ_k is the covariance matrix for the landmark points and Σ_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. $\Sigma_k = \sigma^2 \mathbf{I}_k$ and $\Sigma_m = \mathbf{I}_m$. 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 $\Sigma_k = \sigma^2 \mathbf{I}$ 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 4.10-4.13).
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 orange in Tables 4.10-4.11. For example, in Table 4.10, 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 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.
The test performing better are highlighted in a darker red (see Tables 4.12 and 4.13). 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 4.10). Hence the evaluation of power behaviour for these tests should be made with caution. 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).

Table 4.10: Simulations under H_0: $n_1 = n_2 = 10$, $\lambda^{1/3}$, $k = 6$, $m = 3$, $B = CMC = 1000$

	Achieved α-level					
Test	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

Table 4.11: Simulations under H_0: $n_1 = n_2 = 10$, $\lambda^{1/10}$, $k = 6$, $m = 3$, $B = CMC = 1000$

	Achieved α-level					
Test	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
$G_{(\mu,\mathbb{M}e),F}$	0.009	**0.053**	0.098	0.199	0.304	0.509
$G_{d,(\mu,\mathbb{M}e),F}$	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 4.12: Simulations under H_1: $n_1 = n_2 = 10$, $\lambda^{1/3}$, $k = 6$, $m = 3$, $B = CMC = 1000$

	Power					
Test	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
$\mathbf{G_{d,MA,T}}$	**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**
$\mathbf{T^2}$	**0.253**	**0.483**	**0.609**	**0.739**	**0.808**	**0.872**
$G_{(\mu,\mathbb{M}e),F}$	0.284	0.571	0.711	0.853	0.918	0.969
$G_{d,(\mu,\mathbb{M}e),F}$	0.164	0.431	0.596	0.797	0.880	0.952
$\mathbf{G_{(\mu,\mu^2),F}}$	**0.224**	**0.467**	**0.612**	**0.740**	**0.833**	**0.922**
$\mathbf{G_{d,(\mu,\mu^2),F}}$	**0.137**	**0.369**	**0.495**	**0.663**	**0.782**	**0.903**

Table 4.13: Simulations under H_1: $n_1 = n_2 = 10$, $\lambda^{1/10}$, $k = 6$, $m = 3$, $B = CMC = 1000$

	Power					
Test	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
$\mathbf{G_{d,MA,T}}$	**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**
$\mathbf{T^2}$	**0.054**	**0.155**	**0.244**	**0.362**	**0.448**	**0.585**
$G_{(\mu,\mathbb{M}e),F}$	0.097	0.257	0.369	0.524	0.631	0.782
$G_{d,(\mu,\mathbb{M}e),F}$	0.066	0.219	0.355	0.508	0.626	0.798
$\mathbf{G_{(\mu,\mu^2),F}}$	**0.066**	**0.193**	**0.298**	**0.435**	**0.557**	**0.741**
$\mathbf{G_{d,(\mu,\mu^2),F}}$	**0.047**	**0.152**	**0.265**	**0.424**	**0.543**	**0.727**

4.3 Another type of dependence among landmarks: paired data

When talking about dependent landmark data, we may allude not only to the case of nonzero correlation between landmarks, but also to the case of paired landmarks. Usually, dependent samples are either two groups matched on some variable (e.g. when examining ages of men and women in a relationship, the two groups are matched on relationship status) or are the same people being tested twice (e.g. blood pressure values recorded before and after the drug administration on the same subject). In shape analysis, paired data problem is 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 organisation 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 relabelling 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, 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 relabelled. 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 $(\tau_1, \ldots, \tau_k)$, 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

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

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

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

$t = 1, \ldots, k,\ i = 1, \ldots, n,\ j = 1, 2,\ h = 1, \ldots, q$, where μ_h represents a population constant for the hth variable, μ_{hit} represents a time effect on the hth variable at time t and specific to the ith individual; δ_{hjt} represents treatment time effect at level j on the hth variable which, without loss of generality, are assumed to be $\delta_{h1t} = 0$, $\delta_{h2t} \leq (\text{or} \geq) 0$, $\forall (h, t)$; $\sigma_{ht}(\delta_{hjt}) > 0$ 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_{hjt}, provided that, when $\delta_{h2t} \neq 0$, stochastic dominance relationships $X_{h1} \overset{d}{<} (\text{or } \overset{d}{>}) X_{h2}$, $h = 1, \ldots, q$, 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

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

against the alternatives of the form

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

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 $\mathbf{X}_{2t}$, $t = 1, \ldots, k$, are exchangeable within units. Note that in general case some of the sub-alternatives may be one-sides, 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

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

$t = 1, \ldots, k$, $i = 1, \ldots, n$, $h = 1, \ldots, q$, where it is shown that differences Y_{hit} depend only on treatment effects, exchangeable errors, and $\delta_{h\cdot t} = \delta_{h2t} - \delta_{h1t} = \delta_{h2t}$. 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 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

$$T_{ht} = \phi_{ht}\left(\textstyle\sum_i Y_{hit}\right), t = 1, \ldots, k,\ h = 1, \ldots, q,$$

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 wheter the htth sub-alternative H_{1ht} of interest is two-sided, $\neq$, or one-sided $>$ or $<$ respectively. It is worth noting that all these partial tests $\mathbf{T} = \{T_{ht},\ t = 1, \ldots, k,\ h = 1, \ldots, q\}$ are marginally unbiased. Actually each sub-hypothesis H_{0ht} against H_{1ht}, being separately related to the hth 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 $\mathbf{Y} = \{Y_{hit},\ t = 1, \ldots, k,\ i = 1, \ldots, n,\ h = 1, \ldots, q\}$, 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 $\{Y_{hit},\ t = 1, \ldots, k,\ i = 1, \ldots, n,\ h = 1, \ldots, q\}$. 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 firstly 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.

Chapter 5

Power behaviour of permutation tests with high dimensional data

5.1 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), 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 that 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 Chapter 3, we have introduced an Hotelling's T^2 permutation counterpart (denoted by T'') within the nonparametric combination approach (Pesarin, 2001). Simulation results displayed in Tables 3.1-3.3 show that the power for the suggested test statistic T'' 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 $\mathcal{X}^n_{/\mathbf{X}}$. On the basis of these results, we have performed a simulation study to evaluate the power of multivariate permutation tests combination based (Pesarin, 2001). 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.
These results allow us to introduce and then extend to shape analysis the notion of *finite-sample consistency* for NPC tests, illustrated in Chapter 6. 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.

5.2 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:

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

$i = 1, \ldots, n_j$, $j = 1, 2$, $h = 1, \ldots, k$, 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 δ_{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 behaviour 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 3 specimens are available in each group. Once more we are interested in investi-

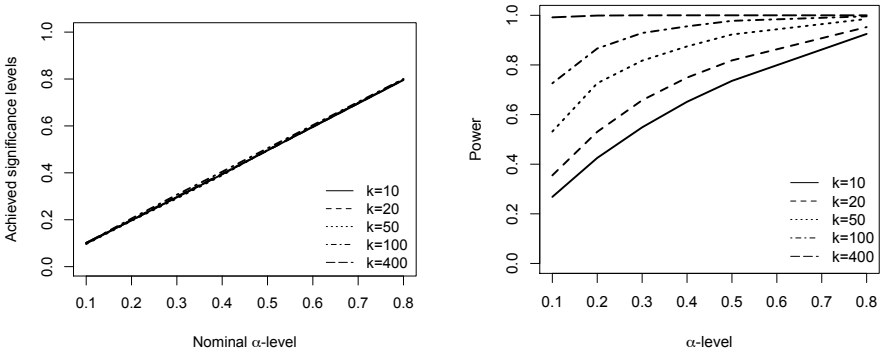
gating what happens to the power when the number of variables is larger than the cardinality of the permutation sample space. We recall that with $n_1 = n_2 = 3$ the cardinality of the permutation sample space is given by $\binom{6}{3} = 20$, hence we wish to analyze the power behaviour using all the possible permutations and recalling that in this case the minimum attainable α-level is 1/20 for two-sided tests. Figure 5.1 shows simulation results in the nondirectional case, and Figure 5.2 those for 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 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 Figure 5.3 are shown simulation results when the underlying distribution is the multivariate Cauchy $\mathcal{C}y(0, 1)$,
- in Figure 5.4 are displayed simulation results when data are distributed according to a Student t with 2 d.f., denoted by $\mathcal{S}t(2)$,
- in Figure 5.5 are shown simulation results when data follows a Pareto distribution $\mathcal{P}a(1, 1)$, with a fixed parameter of shape equal to 1.

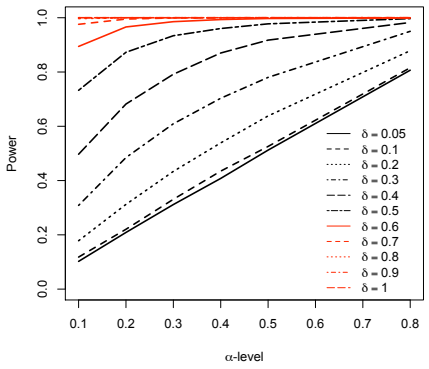
Both in the nondirectional and directional cases, under the multivariate normal distribution, the power increases when increasing the number of the processed variables or the value of the noncentrality parameter δ. We also have found that, under the Cauchy distribution, the power holds steady increasing 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 $\mathcal{S}t(2)$ 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 $\mathcal{P}a(1, 1)$, 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 $\mathcal{C}y(0, 1)$ has no mean and infinite variance and Pareto $\mathcal{P}a(1, 1)$ 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 $\mathcal{C}y(0, \sigma)$ or Pareto $\mathcal{P}a(\theta, \sigma)$, 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.

5.3 Including landmarks and semilandmarks

As application, getting back to our case-study, we have considered 12 Monachus monachus skulls (Mo, 2005), 4 of which are classified as young (i.e. the age class category is less or equal to 4) and 8 as adult (i.e. the age class category is greater than 4). We wish to assess whether or not there are cranial differences between

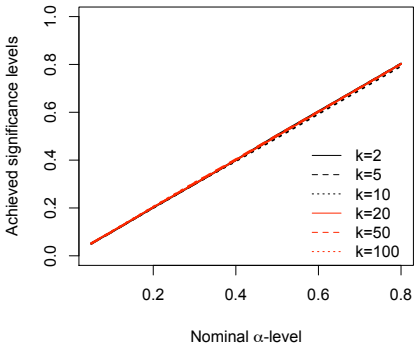


(a) Simulations under H_0, k variables (b) Simulations under H_1,k variables, δ=0.5)

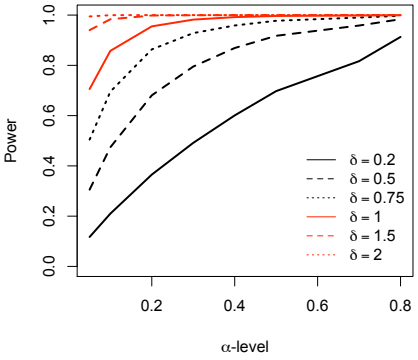


(c) Simulations under H_1, $k = 100$

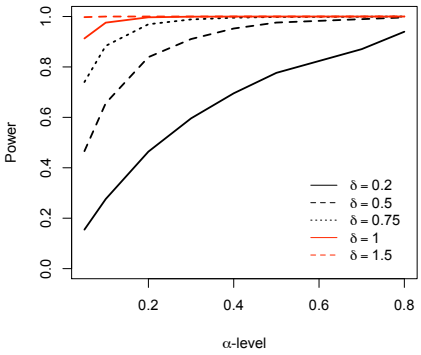
Figure 5.1: Simulation results under Multivariate Normal Distribution $\mathcal{N}(0, 1)$ distribution, nondirectional alternatives (a)-(c)



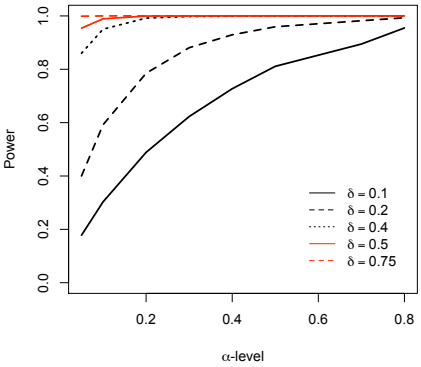
(a) Simulations under H_0, k variables



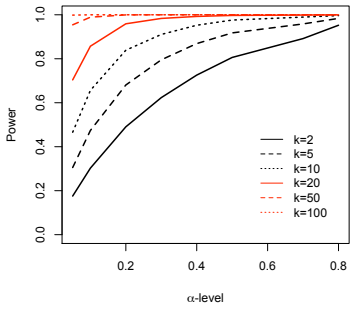
(b) Simulations under H_1, $k = 5$



(c) Simulations under H_1, $k = 10$

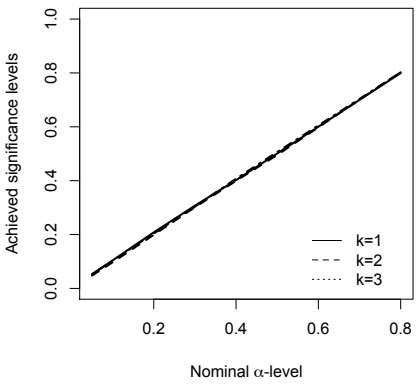


(d) Simulations under H_1, $k = 50$

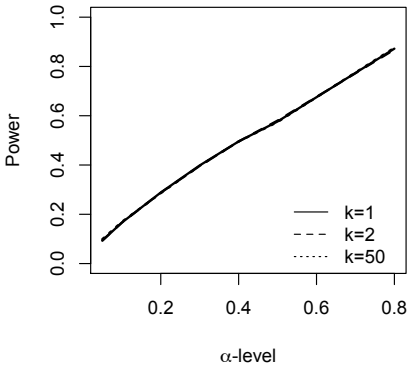


(e) Simulations under H_1, $\delta = 0.5$, k variables

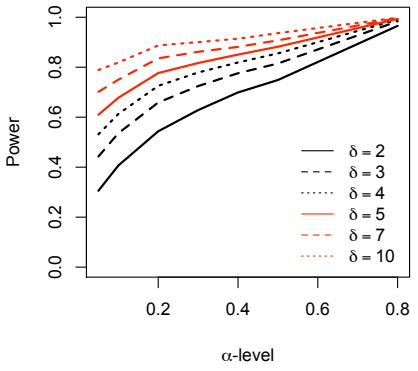
Figure 5.2: Simulation results under Multivariate Normal Distribution $\mathcal{N}(0, 1)$, directional alternatives (a)-(e)



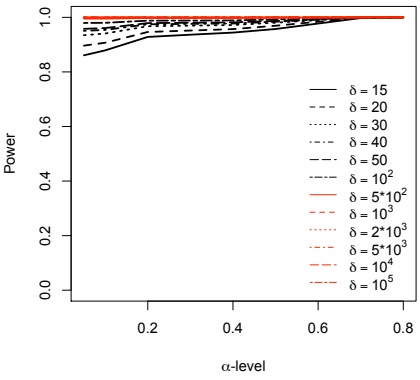
(a) Simulations under H_0, k variables



(b) Simulations under H_1, $\delta = 0.5$, k variables

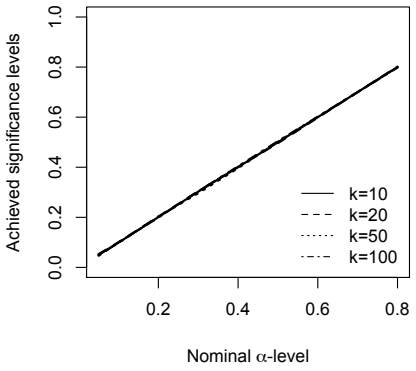


(c) Simulations under H_1, $k = 2$

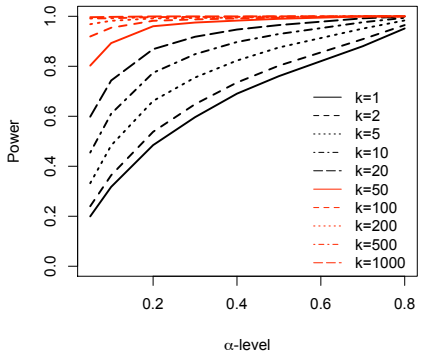


(d) Simulations under H_1, $k = 2$

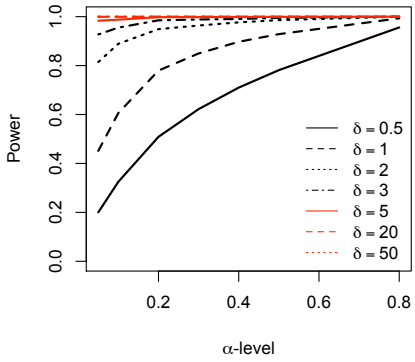
Figure 5.3: Simulation results under Cauchy $\mathcal{C}y(0, 1)$ distribution, directional alternatives (a)-(d)



(a) Simulations under H_0, k variables

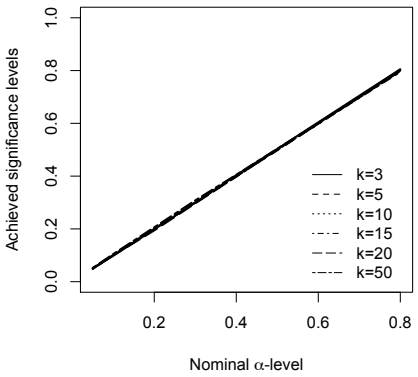


(b) Simulations under H_1, $\delta = 1$, k variables

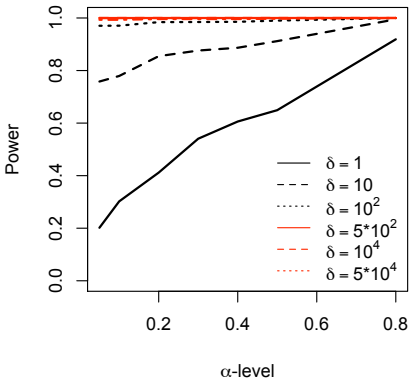


(c) Simulations under H_1, $k = 10$

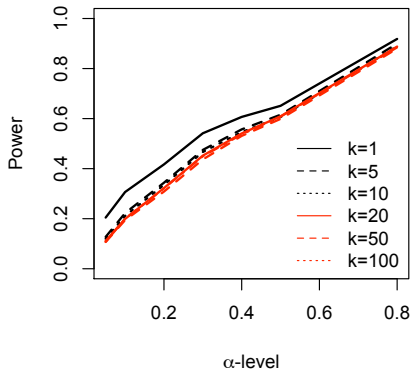
Figure 5.4: Simulation results under Student's $\mathcal{S}t(2)$ distribution, directional alternatives (a)-(c)



(a) Simulations under H_0, k variables



(b) Simulations under H_1, $k = 1$



(c) Simulations under H_1, $\delta = 1$, k variables

Figure 5.5: Simulation results under Pareto $\mathcal{P}a(1,1)$ distribution, directional alternatives (a)- (c)

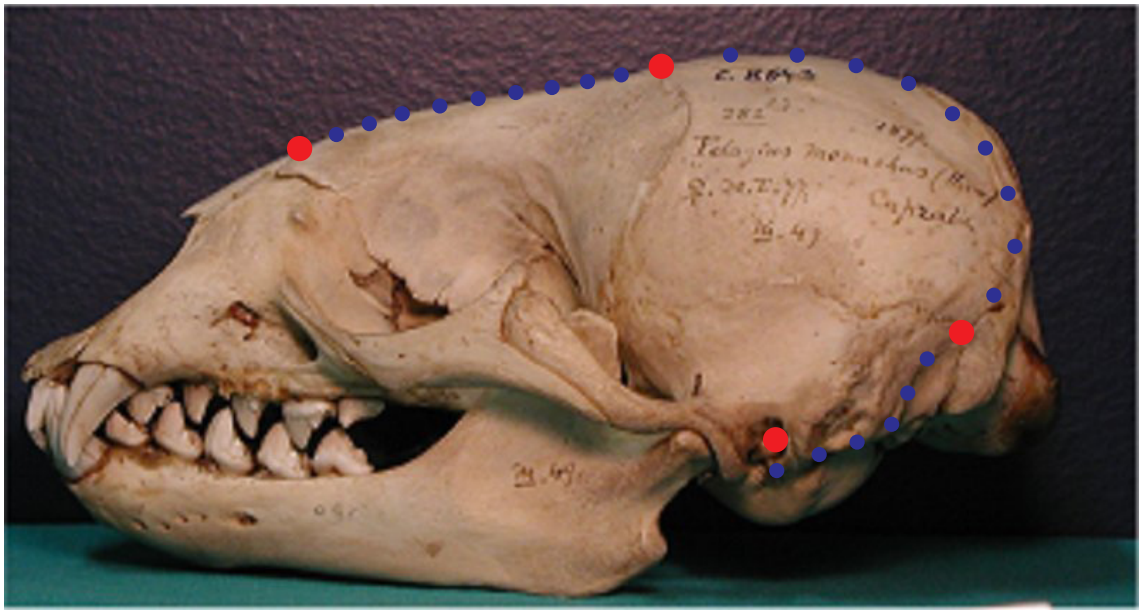


Figure 5.6: Landmarks and semilandmarks.

different age class categories.
Given the small sample sizes we have applied the NPC methodology in the two independent sample case. We have chosen 4 anatomical landmarks (auriculare, rhinion, and two suture points between them) and 24 semilandmarks, 6 in the first curve from auriculare to the second point, 9 in the other two curves from point 2 to point 3 and from point 3 to rhinion (see Figure 5.6). As a part of the GPA 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.

Let us recall that the NPC methodology is based on a decomposition of the hypotheses into r, $r > 1$, sub-hypotheses, where for each sub-hypothesis, a suitable partial permutation test statistic is available. In this case 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 (thus obtaining a second-order statistic) we can acquire informations (i.e. a p-value) for each 2D landmark and semilandmark. In particular, combining x and y coordinates, we can get a p-value of $p = 0.0026$ for auriculare, a p-value of $p = 0.0045$ for the second point, a p-value of $p = 0.0050$ for the third point and a p-value of $p = 0.0026$ for rhinion. We can also compute a p-value for each curve since a curve could be seen again as a combination of semilandmarks. We get a p-value of $p = 0.0975$ for the first curve, a p-value of $p = 0.0045$ for the second one and a p-value of $p = 0.0026$ for the third curve. We can finally consider a global combination of all the partial tests obtaining a global p-value $p = 0.0026$.
As it will be shown in Chapter 7, applying geometric morphometric techniques and performing a principal components analysis (PCA) in the tangent shape spaces, we are also able to assign without classification errors young and adult specimens to

the right group.
A noteworthy comment is that 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 $k = 28$, $m = 2$ and $n_1 + n_2 = 12$ hence we should calculate $F_{52,-41}$, which is impossible).
Through this application to shape analysis, we have highlithed the power and the flexibility of the nonparametric permutation solution, since it makes possible to deal with data sets including many informative landmarks and few specimens.

Chapter 6

Finite-sample consistency of combination-based tests in shape analysis

6.1 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 3 landmarks in 2D) corresponding to these aligned specimens lie on the surface of a unit hemisphere of the same dimensionality as Kendalls 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 Figure 6.1, taken from the paper by 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 aver-

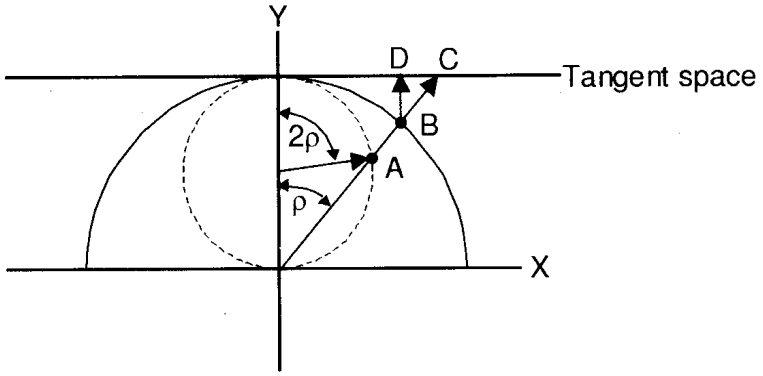


Figure 6.1: Kendall's shape space for triangles

age 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 preshapes 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 Figure 6.1.

6.2 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 decreases 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 $h = 1, \ldots, km$ 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. 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 dimostrate 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 $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ from independent populations with mean shapes $[\mu_1]$ and $[\mu_2]$.
Let $v_1, \ldots, v_{n_1}$ and $w_1, \ldots, w_{n_2}$ 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 Figure 6.1 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 Chapter 3 (Section 3.2), 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 $U = \{U_{hji},\ i = 1, \ldots, n,\ j = 1, 2,\ h = 1, \ldots, km\}$, 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 $(a_1^*, \ldots, a_n^*)$ a permutation of the labels $(1, \ldots, n)$, $U^* =$

$\{U^*_{hji} = U_{hj}(a^*_i),\ i = 1,\dots,n,\ j = 1,2,\ h = 1,\dots,km\}$ 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 δ_{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 non-negative 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

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

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 $\mathbf{T} = \{T_h, h = 1, \ldots, km\}$ necessary for nonparametric combination are obviously the same of those presented in Chapter 3 (Section 3.2). Hence all permutation partial tests are marginally unbiased, consistent and significant for large values.
Let $\lambda_h, h = 1, \ldots, km$ 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 Chapter 3 (Section 3.2) 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) it is proved that if at least one partial permutation test $T_h, h = 1, \ldots, km$ is weakly consistent for H_{0h} against H_{1h} respectively, then $T'' = \psi\{\lambda_1, \ldots, \lambda_{km}\}$, $\forall\ \psi$ 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 $P_{\mathbf{Z}}$. The extension from fixed to random effects will be presented too.
We limited out 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) 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 behaviour 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 6.2.1. *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 non-negative 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) to (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. $\mathcal{T}_{U(\delta)} : \{T^*(\delta) = T(U^*(\delta) : U^*(\delta) \in \mathcal{U}^n_{/U(\delta)}\}$. 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.

$$T^o(U(\delta_Z)) = \max_{\mathcal{T}_{U(\delta_Z)}} \{T^*(\delta_Z) : U^*(\delta_Z) \in \mathcal{U}^n_{/U(\delta_Z)}\}.$$

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 $\mathbf{U}^*(\delta_Z) \neq \mathbf{U}(\delta_Z)$, $T^o(\mathbf{X}^*(\delta_Z)) < T^o(X(\delta_Z))$.

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 $U^*(\delta_Z) \in \mathcal{U}^n_{/U(\delta_Z)}$ 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

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

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

Theorem 6.2.2. *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 non-negative fixed effects;

(iv) fixed effects δ diverge to the infinity according to the monotonic sequence $\{\delta_v, v \geq 1\}$ as in Lemma 6.2.1.

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 $\mathrm{P}_Z(z) = \Pr\{Z \leq z\}$ 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 6.2.1, 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

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

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 $\min_{n_1}(Z_{1i}+\delta_{\mathbf{Z}}) > \max_{n_2}(Z_{2i})$ converges to zero monotonically as δ diverges. This, by taking also account of Lemma 6.2.1, implies that the unconditional rejection rate

$$W_\alpha(\delta) = \int_{\mathcal{U}} \Pr\{\lambda(U(\delta)) \leq \alpha | \mathcal{U}^n_{/U(\delta)}\} \, \mathrm{d}P_Z(z),$$

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

Theorem 6.2.3. *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 6.2.2.

If conditions i) to 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 towards $[\tilde{\delta}, 0]$. 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 6.2.2, we have that

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

The limit of this probability, as δ tends to infinity, is equal to 1, since the associated sequence of probabilities $\{\mathrm{P}_Z[t - \delta_v], v \geq 1\}$ monotonically converges to zero and (n_1, n_2) are fixed and finite.
According to Theorem 6.2.2, the related rejection rate converges to 1 for all $\alpha \geq \alpha_a$. T is weak unconditional finite-sample consistent. ∎

Results obtained in Lemma 6.2.1, concerning the conditional finite-sample consistency of T, and in Theorem 6.2.2, concerning the weak unconditional finite-sample consistency of T, even in presence of a function $\rho(\delta) > 0$ of effect δ (Theorem 6.2.3), can be extended to the case of random effects Δ.

Theorem 6.2.4. *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 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. $\Delta_v \overset{d}{\leq} \Delta_{v+1}, \forall v \geq 1$;

(v) $\lim_{v\uparrow\infty} \Pr\{\Delta_v > u\} \to 1$ for every finite u.

If conditions (i) to (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

$$\mathrm{P}_Z(t - \Delta'' \leq u) \overset{d}{\leq} \mathrm{P}_Z(t - \Delta' \leq u), \forall u,$$

whenever $\Delta' \overset{d}{\leq} \Delta''$. So that 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 behaviour 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 6.2.5. *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 6.2.4 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 ($r = 1, \ldots, Q$) 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 6.2.2 and/or of Theorem 6.2.3 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 towards 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 6.2.3, 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 to 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 $\{(vn'_1, vn'_2), v \geq 1\}$.

6.3 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 Figure 6.2. 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 Figure 6.2) 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.
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 non-centrality 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 5 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 3 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 it is due to the Procrustes superimposition process, that probably has a more considerable stretching or shortening strength in 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 corresponds 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 Figure 6.2 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 6.2, 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 6.1-6.2, in order to draw the attention of the reader, we have highlighted in a darker blue and in bold, the case $k = 10$.
In Table 6.1 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 6.2, 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. 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 Figure 6.2 (i.e. the baseline). Hence Procrustes superimposition process involve just a shortening step in order to obtain completely matching configurations.

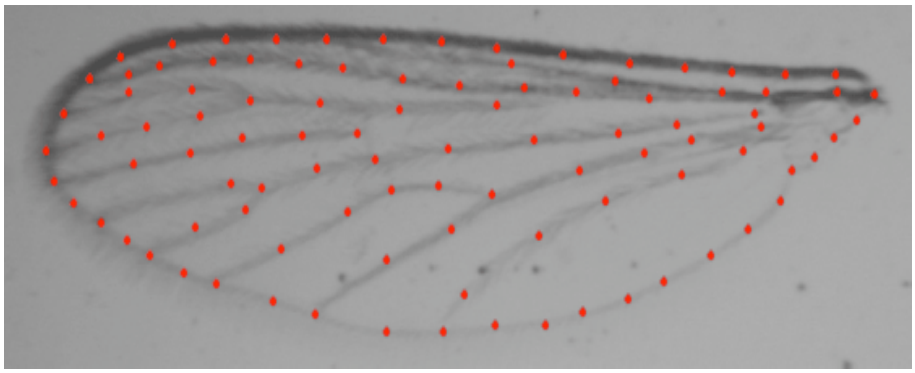


Figure 6.2: $k = 98$ points registered in the left wing of female *Aedes canadensis*

Table 6.1: Controlling achieved α level and evaluating power: $n_1 = n_2 = 5$, $m = 2$, $B = MC = 1000$

Δ=0						
Nominal α	$k = 2$	$k = 5$	$\mathbf{k = 10}$	$k = 20$	$k = 50$	$k = 98$
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

Table 6.2: Evaluating power: $n_1 = n_2 = 5$, $m = 2$, $B = MC = 1000$

α-level	$k = 2$	$k = 5$	**k = 10**	$k = 20$	$k = 50$	$k = 98$
Δ effect up 1%						
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

Chapter 7

Applications to real case studies

7.1 Further results on Mediterranean monk seals data

Let us now come back to the case study introduced in Chapter 2. The dwindling population of Mediterranean monk seals (*Monachus monachus*), split into two subpopulations living in Mauritanian and Aegean waters respectively, adds up to a total of 350-400 individuals. Its is thus possible that this species is indeed on the eve of extinction, following the fate of the closely related West Indian monk seal Monachus tropicalis (Jefferson et al., 2008). Collection of biological data on endangered species is restricted by limited number of free-ranging individuals and by obvious caution necessary to approach them. Under these difficult conditions, field studies are scarce and consist mainly of sightings of animals in the wild. Modern evaluation and careful study of museum specimen could surely further improve the general knowledge on the Mediterranean monk seal systemic anatomy, general biology, diet and even comparative pathology. To this effect, in a previous study we evaluated the possible relationships between bone density of selected areas of the skull of this species and age and gender categories. Using mineral bone deposition as an investigation tool, we were able to identify 5 age classes of individuals in a series of 17 specimens presently housed in several Museums of Natural History or Zoology. Based on the age and sex categories defined in our former study, we try here to further improve our research possibilities by evaluating whether a statistical approach based on osteometric measurements and surface analysis of skull photographs may lead to a new method of age and sex classification in this critical marine mammal.
Hence, our sample consists of 17 Mediterranean monk seal skulls, belonged to individuals originating from Italian or nearby coasts. Information about specimen catalogue number, sex, estimated age, date of collection, origin of location, morphometric and osteometric measurements have already been reported elsewhere (Mo, 2005). A photographic documentation including left-lateral, frontal, posterior, dorsal and ventral views of the skull is also available for each subject. We refer to Table 7.1 for details where 'nk' stands for those animals without any classication for sex and/or age.
By the way, in the previous work by Mo (2005) data consist in 31 morphometric

measurements recorded with a plastic manual caliper, mostly taken on the ventral and dorsal planes of the seal skull.
While in Mo (2005) these data have been processed using standard multivariate morphometrics, we propose to apply geometric morphometric techniques, carrying out inference in a nonparametric permutation framework.

Table 7.1: Our sample

Age category	M	F	nk
3	-	-	2
4	1	1	-
7	-	2	2
8	-	2	-
9	2	-	-
nk	1	-	4

Before getting to the core, we wish to briefly summarize some of the techniques and the notions already presented throughout the thesis which will "help" the statistical analysis.
In many biological and biomedical investigations, the most effective way to analyze shapes is by recording and locating a finite number of landmarks and semilandmarks (also referred to as landmark based approach). The main strategies to analyze databases of landmark locations are multivariate morphometrics and deformation analysis (Bookstein, 1986). One can measure configurations of landmark points by variables that express aspects of size or shape of single specimens (distances or ratios of distances) or can directly measure the relation between one form and another as a deformation (varying rearrangement of the configuration of landmarks considered as a whole). Either approach may be turned to the investigation of group differences in size and shape or between size change and shape change (Dryden and Mardia, 1998). Only recently, in the late 1980s, various authors, among which we mention Fred Bookstein and James Rohlf, proposed a synthesis of these two experiences called geometric morphometrics, that is a collection of approaches for the multivariate statistical analysis of Cartesian coordinate data, usually (but not always) limited to landmark point locations.
A key benefit to use the geometric morphometric methods, instead of traditional morphometric methods, is that since all the geometric information is retained throughout a study, results of high-dimensional multivariate analyses can be mapped back into physical space to achieve appealing and informative visualizations that are frequently not possible with alternative methods (Slice, 2005).
Most inferential methods in shape analysis field analyses can be interpreted in terms of differences between congurations of landmarks optimally superimposed using a least-squares procedure. For instance, shape variables are usually constructed using a Procrustes superimposition. Through this method, the raw coordinates are superimposed by translating the configurations to a common centroid, scaling to unit centroid size, and rotating until the sum of the squared distances between cor-

responding landmarks is minimized. When including semilandmarks, the method of sliding semilandmarks proposed by Bookstein (1997), representing an extension of the standard Procrustes superimposition procedure, could be apply. Once the optimally adjusted positions of the landmarks and semilandmarks are determined, they can all be treated in the same way in subsequent statistical analyses. Among traditional morphometric methods, we have mentioned deformation analysis.
In particular deformation analysis includes relative warps, smoothing thin plate splines and tangent space methods. Being available a random sample of shapes, it could be intriguing to explore the structure of the within group variability in the tangent space to shape space. This problem may be addressed using either PCA with respect to Euclidean metric or the method of relative warps (Dryden and Mardia, 1998). Briefly, the method of relative warps consists of fitting an interpolating function, such as the thin-plate spline, to the coordinates of the landmark for each specimen in a sample. Variation among the specimens within a sample is described in terms of variance in the parameters of the fitted functions. The relative warps are simply principal components of a distribution of shapes in shape tangent space. They are used to describe the major trends in shape variation among specimens within a sample as deformations in shape. As principal components do, they summarize the variation among the specimens as principal directions of the variation in shape. Indeed, relative warps and relative warps scores are useful tools for describing the non-linear shape variation in a dataset. Moreover when variation is not well localized, the method of relative warps displays the pattern of covariation between the displacements at different landmarks very effectively (see for details, Rohlf, 1993; Bookstein, 1991).
Rohlf (2000) reviews the main statistical tests proposed in shape analysis comparing their statistical power. But all these classical tests are parametric in nature and need strong assumptions to be applied. Alternative inferential procedures are represented by permutation methods. We remark that these tests are distribution-free and allow us for efficient solutions even when the number of cases is less than the number of variables (Blair et al., 1994). In the wake of these considerations, we proposed an extension of the NonParametric Combination (NPC) methodology (Pesarin, 2001).
Coming back to our case study, we recall that we have carried out two analyses using available specimen information about sex (first analysis) and estimated age (second analysis). In Table 7.2 are shown the details for subjects involved in the two analyses. At a glance, it is clear that sample size changes when considering different statification variable (age or sex). The experimental design (i.e. the choice of the landmarks) as well as some inferential results have been presented in Chapter 5. As a reminder, we have chosen 4 anatomical landmarks (auriculare, rhinion, and two suture points between them) and 24 semilandmarks, 6 in the first curve from auriculare to the second point, 9 in the other two curves from point 2 to point 3 and from point 3 to rhinion. The three curves represents the domains. In Figure 7.1 are shown unregistered (observe carefully, not registered or superimposed by GPA) outlines of young and adult female, along with outlines. Coordinate locations of landmarks ad semilandmarks have been digitized using tpsDig2 software (Rohlf, 2007). It is of interest to assess whether there is a difference between the sexes

Table 7.2: Subjects involved in sex based analysis (a) and in age based analysis (b)

(a)

#	Specimen	Age class
1	FI8644_M	9
2	FI8646_M	nk
3	GE225_M	4
4	GE713_M	9
5	TS759_F	7
6	FI8643_F	4
7	FI8645_F	8
8	GE714_F	7
9	GE17760_F	8

(b)

#	Specimen	Age class
1	FI8643_F	4
2	FI8644_M	9
3	FI8645_F	8
4	GE225_M	4
5	GE713_M	9
6	GE714_F	7
7	GE17760_F	8
8	PV3402_nk	7
9	PV3799_nk	7
10	TS394_nk	3
11	TS395_nk	3
12	TS759_F	7

(male vs. female) and the ages (adult vs. young), hence whether it is possible to discriminate, to classify subjects into predefined categories according to the available information on sex and age class category.
To this purpose, we have performed principal components analysis (PCA) in the tangent shape spaces, displaying the position of each specimen with respect to the first and second PC. Then we have carried out inferential analysis, using NPC methodology. TpsRelw software (Rohlf, 2008a) allows for sliding semilandmarks along outlines or curves to be combined with landmark points. The procedure involves first sliding the semilandmarks to the left or right along a curve to as to minimize the amount of shape change between each specimen and the Procrustes average of all the specimens. The computations are iterative. TpsRelw is used to explore the overall diversity of within-sample shape variation. It provides a low dimensional approximation (via a principal components analysis) to the tangent space approximation of shape space. A plot of the relative warp scores matrix, showing the position of each specimen with respect to the first and second partial warps, is displayed in Figure 7.2.
After sliding semilandmarks and superimposing subjects minimizing Procrustes distance we would expect to find in Figure 7.2 (a) two separate groups: males (# 1,2,3,4) vs. females (# 5,6,7,8,9). But given the low sample size and the fact that using information on sex we have only 2 young specimens over 9 specimens in total, it is very hard to detect a sexual dimorphism and if there is any it is masked by age effects. For these reasons, focussing on the age based analysis, we have decided to apply the nonparametric methodology, in order to detect differences between different age class categories. In such a way, we can obtain a sample of 12 specimens, 4 of those are considered young (i.e. the age class category is less or equal to 4) and 8 are adult (i.e. the age class category is greater than 4). The data are the coordinates of

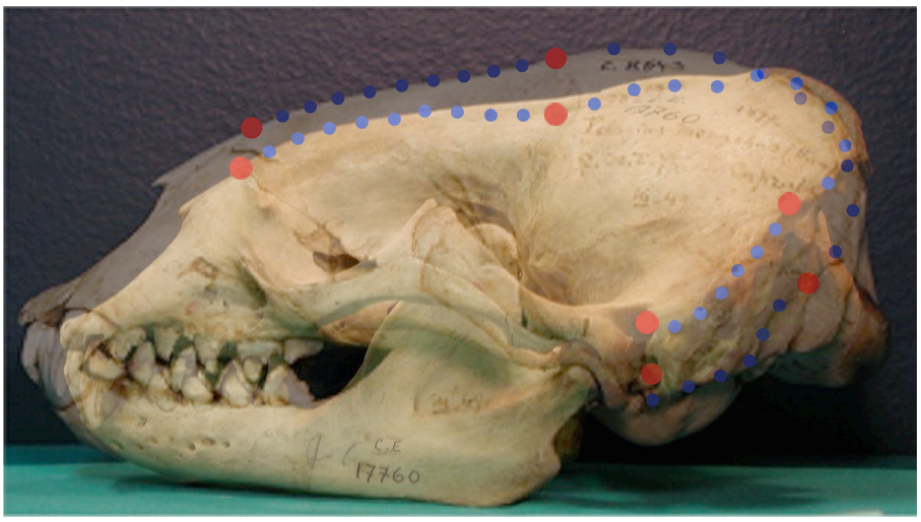
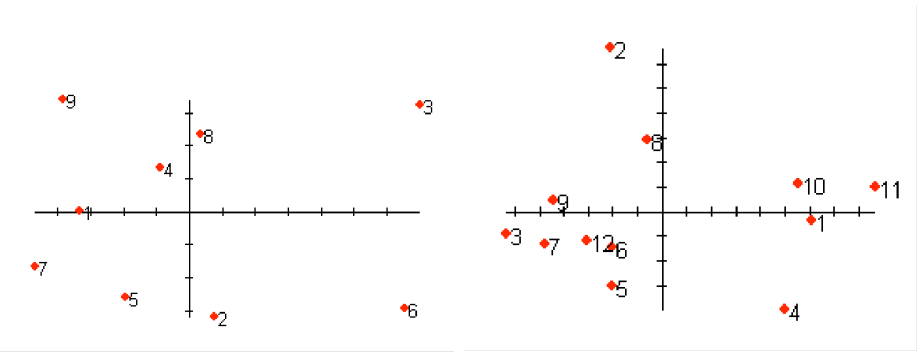


Figure 7.1: Chosen landmarks (4 red points) and semilandmarks (24 blue points).



(a) (b)

Figure 7.2: Plot of the relative warp scores matrix in (a) sex based analysis and (b) age analysis.

landmarks and semilandmarks after sliding and after superimposition, minimizing the Procrustes distance (see superimposed data in Figure 7.3).

(a) (b)

Figure 7.3: Consensus configuration and points (a) or vectors indicating the variability at each point (b) in the age based analysis (12 specimens). The particular disposition of the points (they lie along a straight line) is due to the choice of minimizing the Procrustes distance. A different layout would be obtained minimizing bending energy (for details, see Rohlf, 2008a).

Results obtained by the means of NPC methodology are displayed in Table 7.4.

Table 7.3: Results

Points and curves	p-value
auriculare	0.0026
2nd point	0.0045
3rd point	0.0050
rhinion	0.0026
1st curve	0.0975
2nd curve	0.0045
3rd curve	0.0026
global	0.0026

Our results confirm what stated in Mo (2005) and Marchessaux (1989) according to which there is no apparent sexual dimorphism although it is likely that the number of adult males reaching the maximum size is higher than that of adult females. Also, whether it is found, dimorphism is more related to the length of body or to the coloration of the pelage, than to cranial differences (Brunner et al., 2003). On the other hand, in Figure 7.2 (b) is it possible to see that using information on age class category we able to define two separate groups: young (age class category ≤ 4, # 1,4,10,11) and adult (age class category > 4, # 2,3,5,6,7,8,9,12) specimens. Hence, while it seems difficult to discriminate between male and female specimens also using different views of the skull (frontal, posterior, dorsal and ventral views), it is possible to assign without classification errors young and adult specimens. Even

if we cannot find sexual dimorphism in our sample, probably because of the sample itself, we may state that there is a statistically significant difference between young and adult specimens analyzing landmark and semilandmark point locations.
Studying shape changes in relation with to sex and age seem to have only biological significance and implications. At this point we wish to recall the introductory concerns about the viability of the animal. As already said, the Mediterranean monk seal is believed to be the world rarest pinniped and one of the most endangered mammals of the world, with fewer than 600 individuals currently surviving (Johnson et al., 2006). Once widespread throughout the Mediterranean, the Black Sea and the northwestern coast of Africa, this monk seal has suffered a devastating decline. Today the largest population of Mediterranean Monk Seals is found near Greece. In 2007 it was listed as critically endangered (CR) in the IUCN (International Union for Conservation of Nature) Red List of Threatened Species. This term is use to indicate that a taxon is facing an extremely high risk of extinction in the wild in the immediate future.
Using shape analysis tecniques in a nonparametric permutation framework, enable us to solve effectively high-dimensional small sample size problems. We recall that one of the main advantage of the proposed approach is that, using the MA procedure and the information about domains, it is possible to obtain not only a global p-value, as for traditional tests, but also a p-value for each of the defined aspects or domains. Hence following this procedure it is possible to construct a hierarchical tree, allowing for judgements at different levels of the tree (coordinates level, on the landmarks level after re-combining aspects and coordinates, on the domain level as well as on the global p-value).
Moreover the proposed extension of the NPC methodology to shape analysis enables us to study also the symmetry in an object, that could be useful to answer and solve biological matters, e.g. related to fluctuating and directional asymmetry (Klingenberg et al., 2002; Mardia et al., 2000).
Appealing applications are related to museological finds. One of the problem recently detected in the museums, is that there are many museological pieces that have been collected and stored but they cannot be exhibited since there are difficul to classify on the basis of sex and age class category informations using standard tecniques. Osteological teriofauna collection are very impressive and always attract italian and foreign researchers. Developing geometric morphometrics techniques in a nonparametric permutation framework could be useful in solving these problems. By means of shape analysis, we could at first record the landmarks of interest in collaboration with expert biologists. Standard shape analysis techiniques could be applied to define mean shapes of the objects, to evaluate variability and classify the species. Then inferential analysis could be carried out in a nonparametric framework, thus relaxing the demanding assumptions required by parametric tests. Informative and systematic lists of the museum pieces (e.g. databases available online) or "user-friendly" softwares may be realized. Of course this is convenient especially when dealing with specimens facing an extremely high risk of extinction. Actually, we feel confident that, combining geometric morphometrics techniques and nonparametric permutation methodology, it is possible to give rise to a long run perspectives work.

7.2 On morphology of aortic valve

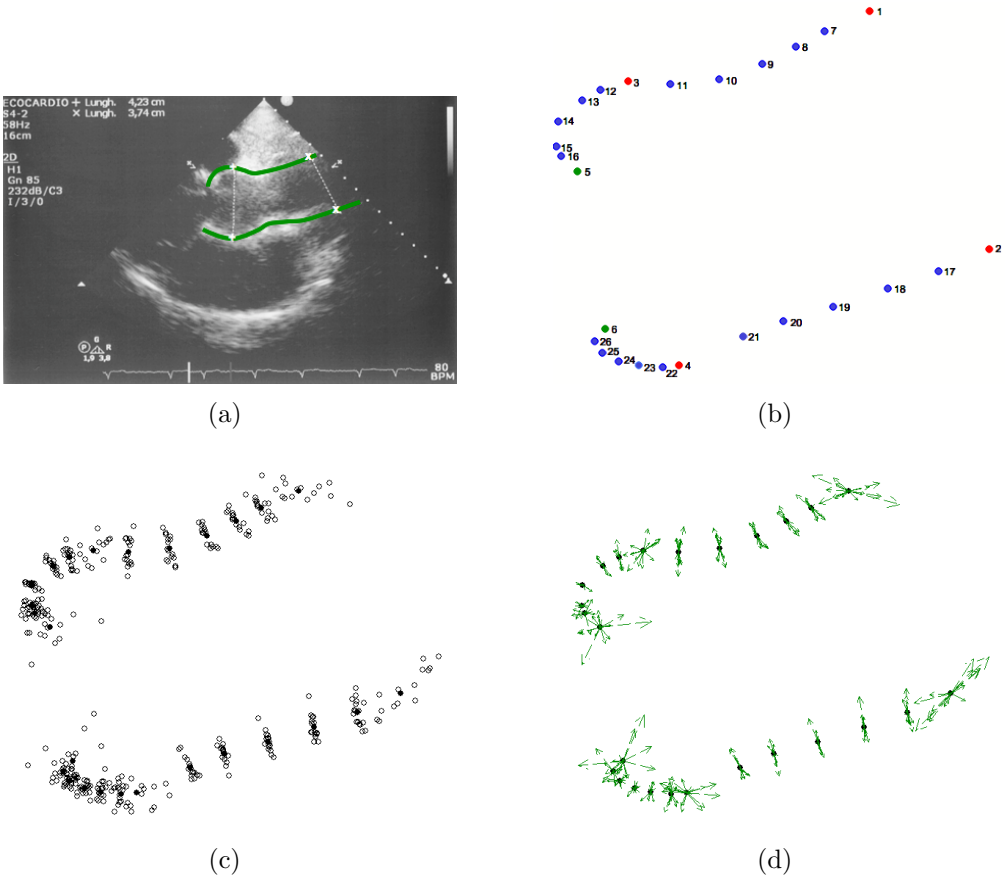
7.2.1 Introduction

In another application we have examined aortic valve shape. Preliminary results are given. Data consists 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 non-invasive 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). 7 patients are overweighted or belong to the obese Class I (that means that 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. 2D 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 Figure 7.4 (b)), related to the measurements taken at the sinuses of Valsalva and ascending aorta (see dashed lines in Fig. 7.4 (a)).Moreover we have digitized 20 semilandmarks (blue points), 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 Figure 7.4 are shown consensus and all subject 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. 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 matrix also showing the position of each specimen with respect to the first and second partial warps (see for example Figure 7.5 (a)). It also allows us to explore the deformations associated with different position in this ordination. For example, in Figure 7.5 (b) we show the deformation corresponding to the position of the consensus. In Figure 7.5 (c), we show the deformation produced when moving from the centre 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 Figure 7.5 (d), 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 Figure 7.5 (e) 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 Figure 7.5 (f) 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.

7.2.2 Inferential results

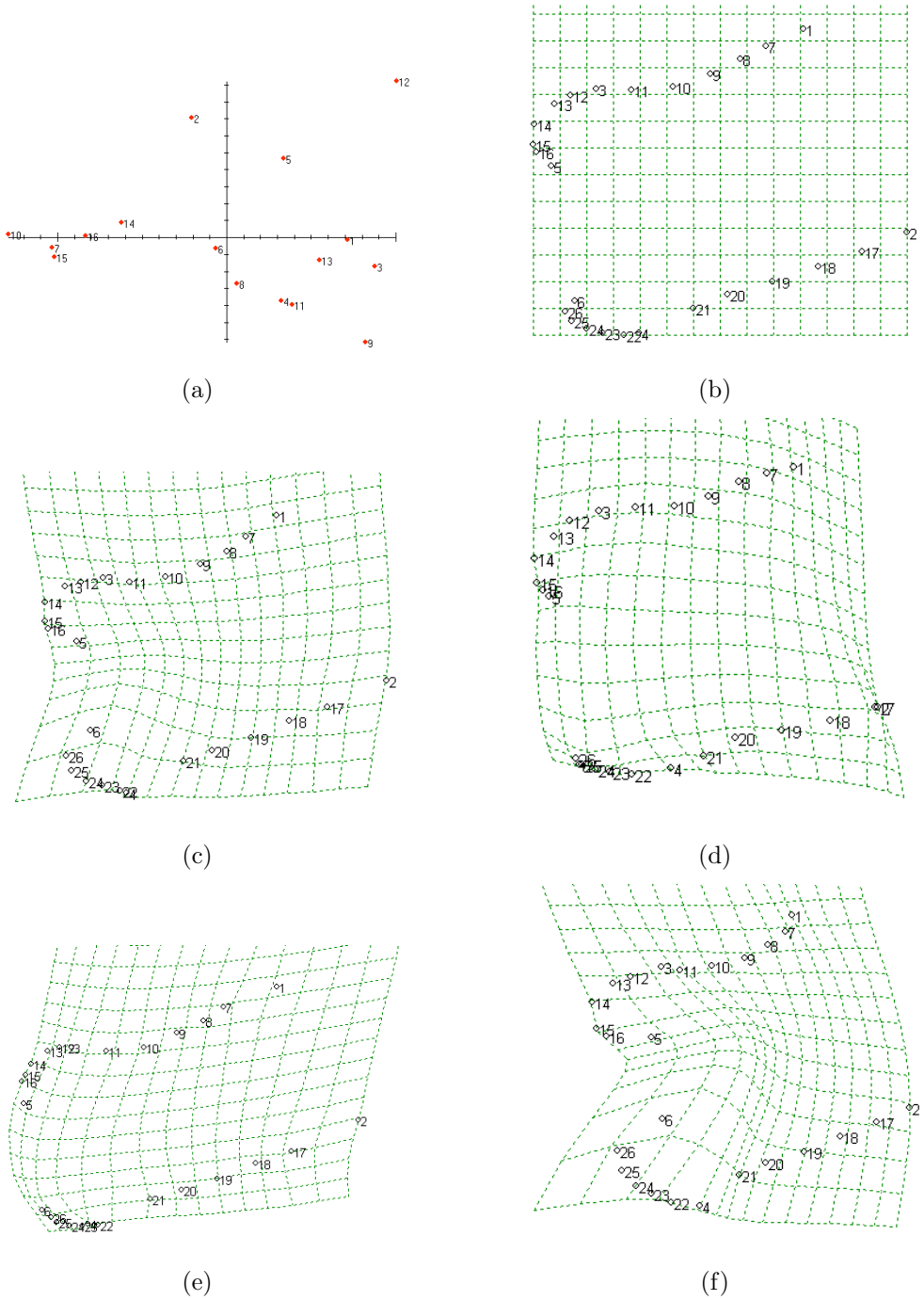
We have used the plot of the relative warp scores matrix to define two groups. The first group includes patients 2, 6, 7, 10, 14, 15 and 16 ($n_1 = 7$). The second includes patients 1, 3, 4, 5, 8, 9, 11, 12, 13 ($n_2 = 9$). 5 of the 7 overweighted or obese patients in our sample are allocated in this 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



(a) (b)

(c) (d)

Figure 7.4: 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).



(a) (b) (c) (d) (e) (f)

Figure 7.5: 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 centre to patient 2 (c), or to patient 9 (d), 10 (e) and 12 (f).

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 7.4). We have found that the two groups are significantly different in all the 6 landmarks, in curves 1, 2 and 3 and globally. We recall that p-values associated to 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 to curve 1 is obtained as a combination of the x and y coordinates of points 7, 8, 9, 10 and 11). In Table 7.4, 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 7.4: 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.

7.3 Some remarks

Case studies have shown that 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-dened 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.

7.4 Future research

A particular case we wish to investigate in the future is that concerning what happens placing a myriad of discrete points (to better define the contour) and then processing them according to their "importance", using closed testing or stepwise procedures. Some work has been already done with reference to the adjustment of stepwise p-values in generalized linear models (Finos et al., 2009). We briefly summarize the results obtained till now. Actually stepwise variable selection methods are an important feature in the analysis of biological data, especially in presence of several covariates. The goal of this method is to identify the better set of predictors of any general linear model (GLM). The interest of applied data analysis for stepwise methods widely addresses the statistical research in the last decades; as a consequence several multiple variable selection methods have been developed. Despite its widespread use inside the scientic community, stepwise procedures are often criticized. The biasedness of standard p-values for stepwise regression is not a novelty (Freedman et al., 1992; Hjorth, 1994; Austin and Tu, 2004; Harshman and Lundy, 2006). So one must be cautious in glm-stepwise findings evaluation, mainly when regressors have been data-steered (Grechanovsky and Pinsker, 1995). We have proposed a nonparametric permutation solution to this problem. The proposed algorithm (Finos and Salmaso, 2006) controls the α-level under the global null hypothesis that all covariates are unrelated to the outcome variable. Moreover, the procedure ensures the unbiasedness and the consistency of the p-values of the

selected model. The proposed method controls the FWE in a weak sense but is valid for any GLM and any stepwise selection method. Our proposal is exact, flexible and potentially adaptable to most different applications of model selection. The correction becomes more severe when many variables are processed by the stepwise machinery. We think that these results may represent a starting point for subsequent extensions and applications to shape analysis.

Bibliography

Adams, D. C. (1999). Methods for shape analysis of landmark data from articulated structures. *Evolutionary Ecology Research* **1**, 959–970.

Adams, D. C., Rohlf, F. J. and Slice, D. E. (2004). Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology* **71**, 5–16.

Amaral, G., Dryden, I. and 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.

Austin, P. C. and Tu, J. V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology* **57**, 1138–1146.

Benjamini, Y. and 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, R. C., Higgins, J. J., Karniski, W. and Kromrey, J. D. (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, F. L. (1986). Size and shape spaces for landmark data in two dimensions. *Statistical Science* **1**, 181–242.

Bookstein, F. L. (1991). *Morphometric Tools For Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.

Bookstein, F. L. (1996). Combining the tools of geometric morphometrics. In *Advances in morphometrics*, vol. 284. Plenum Press, New York, pp. 131–151.

Bookstein, F. L. (1997). Shape and the information in medical images: A decade of the morphometric synthesis. *Computer Vision and Image Understanding* **66**, 97–118.

Brunner, S., Bryden, M. M. and Shaughnessy, P. D. (2003). Cranial ontogeny of otariid seals. *Systematics and Biodiversity* **2**, 83–110.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford University Press, New York.

Dryden, I. L. and Mardia, K. V. (1992). Size and shape analysis of landmark data. *Biometrika* **79**, 57–68.

Dryden, I. L. and Mardia, K. V. (1993). Multivariate Shape Analysis. *Sankhyā Series A* **55**, 460–480.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. John Wiley & Sons, London.

Ferraty, F. and Vieu, P. (2006). *NonParametric Functional Data Analysis: Theory and Practice*. Springer, New York.

Finos, L., Brombin, C. and Salmaso, L. (2009). Adjusting stepwise p-values in generalized linear models. Submitted to *Communications in Statistics: Theory and Methods*.

Finos, L. and Salmaso, L. (2005). A new nonparametric approach for multiplicity control: *Optimal Subset* procedures. *Computational Statistics* **20**, 643–654.

Finos, L. and Salmaso, L. (2006). Weighted methods controlling the multiplicity when the number of variables is much higher than the number of observations. *Journal of Nonparametric Statistics* **18**, 245–261.

Finos, L. and Salmaso, L. (2007). FDR- and FWE-controlling methods using data-driven weights. *Journal of Statistical Planning and Inference* **137**, 3859–3870.

Freedman, L. S., Pee, D. and Midthune, D. N. (1992). The problem of underestimating the residual error variance in forward stepwise regression. *The Statistician* **41**, 405–412.

Friman, O. and Westin, C. F. (2005). Resampling fMRI time series. *NeuroImage* **25**, 859–867.

Frost, S. R., Marcus, L. F., Bookstein, F. L., Reddy, D. P. and 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.

Galilei, G. (1638). *Discorsi e Dimostrazioni Matematiche, intorno a due nuoue Scienze Attenenti alla Mecanica & i Movimenti Locali*. Elzeviri, Leida.

Good, P. (2000). *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer, New York.

Goodall, C. R. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, Series B* **53**, 285–339.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika* **40**, 33–50.

Grechanovsky, E. and Pinsker, I. (1995). Conditional p-values for the f-statistic in a forward selection procedure. *Computational Statistics & Data Analysis* **20**, 239–263.

Harshman, R. A. and Lundy, M. E. (2006). A randomization method of obtaining valid p-values for model changes selected "post hoc". Poster presented at *Seventy-first Annual Meeting of the Psychometric Society*, Montreal, Canada. Available at http://publish.uwo.ca/∼harshman/imps2006.pdf.

Hjorth, J. S. U. (1994). *Computer intensive statistical methods: Validation model selection and bootstrap*. Chapman & Hall, New York.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *Annals of Mathematical Statistics* **23**, 169–192.

Huxley, J. S. (1932). *Problems of Relative Growth*. The Dial Press, New York.

Jefferson, T. A., Webber, M. A. and Pitman, R. L. (2008). *Marine mammals of the world*. Elsevier - Academic Press, Amsterdam.

Johnson, W. M., Karamanlidis, A. A., Dendrinos, P., Fernández de Larrinoa, P., Gazo, G., González, L. M., Güçlüsoy, H., Pores, R. and Schnellmann, M. (2006). Monk seal fact files. Biology, behaviour, status and conservation of the foca monje del mediterráneo, Monachus monachus. *The Monachus Guardian, www.monachus-guardian.org* .

Katina, S., Bookstein, F. L., Gunz, P. and 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, D. G. (1977). The diffusion of shape. *Advances in Applied Probability* **9**, 428–430.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* **16**, 81–121.

Kent, J. T. (1994). The complex bingham distribution and shape analysis. *Journal of the Royal Statistical Society: Series B* **56**, 285–299.

Klingenberg, C. P., Barluenga, M. and Meyer, A. (2002). Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution* **56**, 1909–1920.

Klingenberg, C. P. and McIntyre, G. S. (1998). Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with procrustes methods. *Evolution* **52**, 1363–1375.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. Wiley, New York.

Lele, S. (1993). Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Mathematical Geology* **25**, 573–602.

Lele, S. and Cole, T. M. (1995). Euclidean distance matrix analysis: a statistical review. *Current issues in statistical shape analysis* **3**, 49–53.

Lele, S. and Cole, T. M. (1996). A new test for shape differences when variance-covariance matrices are unequal. *The Journal of Human Evolution* **31**, 193–212.

Lele, S. and Richtsmeier, J. T. (1991). Euclidean distance matrix analysis: a coordinate free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology* **86**, 415–427.

Ludbrook, J. and Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician* **52**, 127–132.

Marchessaux, D. (1989). *Recherches sur la Biologie, l'Ecologie et le Statut du Phoque Moine, Monachus monachus*. GIS Posidonie Publ., Marseille, France.

Marcus, R., Peritz, E. and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**, 655–660.

Mardia, K. V., Bookstein, F. L. and Moreton, I. J. (2000). Statistical assessment of bilateral symmetry of shapes. *Biometrika* **87**, 285–300.

Mardia, K. V., Edwards, R. and Puri, M. L. (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*. Ph.D. thesis, University of Padova.

Palmer, A. R. (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**, 14279–14286.

Palmer, A. R. and 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**, 85–100.

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.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rao, C. and 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**, 12132–12136.

Rao, C. and 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.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester.

Rohlf, F. J. (1993). Relative warp analysis and an example of its application to mosquito wings. In *Contributions to morphometrics*, L. F. Marcus, E. Bello and A. Garcia-Valdecasas, eds., vol. 8. Monografias del Museo Nacional de Ciencias Naturales: Madrid, pp. 131–159.

Rohlf, F. J. (1999). Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification* **16**, 197–225.

Rohlf, F. J. (2000). Statistical power comparisons among alternative morphometric methods. *American Journal of Physical Anthropology* **111**, 463–478.

Rohlf, F. J. (2007). *tpsDig2, digitize landmarks and outlines, version 2.12*. Department of Ecology and Evolution, State University of New York at Stony Brook.

Rohlf, F. J. (2008a). *tpsRelw, Relative warps analysis, version 1.46*. Department of Ecology and Evolution, State University of New York at Stony Brook.

Rohlf, F. J. (2008b). *TPStri, Explore shapes of triangles, version 1.25*. Department of Ecology and Evolution, State University of New York at Stony Brook.

Rohlf, F. J. and Slice, D. E. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* **39**, 40–59.

Romano, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association* **85**, 686–692.

Salmaso, L. and Solari, A. (2005). Multiple Aspect Testing for case-control designs. *Metrika* **12**, 1–10.

Salmaso, L. and Solari, A. (2006). Nonparametric iterated combined tests for genetic differentiation. *Computational Statistics & Data Analysis* **50**, 1105–1112.

Savriama, Y. and Klingenberg, C. P. (2006). Geometric morphometrics of complex symmetric structures: Shape analysis of symmetry and asymmetry with procrustes methods. In *Interdisciplinary Statistics and Bioinformatics*, S. Barber, P. D. Baxter, K. V. Mardia and R. E. Walls, eds. Leeds University Press. ISBN 0-85316-252-2.

Seber, G. (1984). *Multivariate Observations*. John Wiley, New York.

Siegel, A. F. and Benson, R. H. (1982). A robust comparison of biological shapes. *Biometrics* **38**, 341–350.

Slice, D. E. (2005). *Modern Morphometrics In Physical Anthropology*. Springer-Verlag New York, LLC.

Slice, D. E., Bookstein, F. L., Marcus, L. F. and Rohlf, F. J. (1996). A glossary for geometric morphometrics. *Advances in Morphometrics* **284**, 531–551.

Small, C. G. (1996). *The Statistical Theory of Shape*. Springer, New York.

Srivastava, A., Joshi, S., Mio, W. and Liu, X. (2005). Statistical shape analysis: Clustering, learning and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, 590–602.

Terriberry, T. B., Joshi, S. and Garig, G. (2005). Hypothesis testing with nonlinear shape models. In *Information processing in medical imaging*, vol. 3565. Springer Berlin-Heidelberg, pp. 15–26.

Thompson, D. W. (1961). *On Growth and Form*. University Press, Cambridge.

Westfall, P. H. and Wolfinger, R. D. (2000). *Closed Multiple Testing Procedures and PROC MULTTEST*. Observations, SAS Institute Inc.

Westfall, P. H. and Young, S. S. (1993). *Resampling-Based Multiple Testing*. Wiley, New york.

Chiara Brombin - Curriculum Vitae

Personal information

Address: Department of Statistics, University of Padova,
Via Cesare Battisti 241-243 - 35121 Padova (Italy)

Phone: +39 049 827 4174

E-Mail: chiara.brombin@unipd.it

Education

University

From January 2006 until present (expected completion by December 2008):
enrolled at the Doctoral School in Statistical Sciences, at the Department of Statistical Sciences, University of Padua, Italy. Actually I'm working on nonparametric methods for statistical shape analysis, under the supervision of Prof. Luigi Salmaso.

June 2005:
Laurea Degree (four years) in Statistics, at the Faculty of Statistical Sciences, University of Padova, Italy. Thesis on "Multivariate analysis with paired observations and missing values", advisors: Prof. F. Pesarin and Prof. G. A. Fava.

Secondary school

July 2000:
Secondary School Diploma in Classical Studies, at Liceo G.B. Ferrari, Este (Padova) Italy.

Study periods abroad

From March 2008 to June 2008:
Visiting Ph.D. student at the Department of Ecology and Evolution, State University of New York, Stony Brook, NY. Advisor: Prof. F. J. Rohlf.

From April 2007 to June 2007:
Visiting Ph.D. student at the Department of Anthropology, EVAN (European Virtual Anthropology Network) Centre, Vienna, Austria. Advisor: Prof. Stanislav Katina.

Conferences and Presentations

24-29 Agosto 2008
Porto-Portogallo
COMPSTAT2008, International Conference on Computational Statistics, talk titled "Searching for powerful tests in shape analysis".

23-24 May 2008
Penn State's University Park campus, State College, Pennsylvania
Nonparametric Statistics and Mixture Models: Past, Present, and Future, poster presentation "The Hotelling's T^2 permutation counterpart in shape analysis".

19-22 September 2007
Palermo
IV Congresso Nazionale SISMEC, poster presentation"Nonparametric methods in observational studies for testing for association in bladder cancer".

8-11 July 2007
Vienna
MCP 2007 Conference, talk "Adjusting stepwise p-values in generalized linear models".

8-11 May 2007
Rolduc (Kerkrade, The Netherlands)
First IBS Channel Network Conference (International Biometric Society British and Irish Region), talk titled "How confident can you be in stepwise-GLM findings?".

Collaborations

Prof. Ferronato G., MD Peretta R. and MD Galzignato P. F., Dept. of Medical and Surgical Specialties, University of Padova, Italy.

Prof. Floreani A., Dept. of Surgical and Gastroenterological Science, University of Padova, Italy.

Prof. B. Mozzanega, Dept. of Gynaecology and Human Reproduction Sciences, University of Padova, Italy.

Prof. B. Cozzi, Dept. of Experimental Veterinary Science, University of Padova, Italy.

Prof. P. F. Bassi, Dept. of Urology, Sacro Cuore Catholic University Hospital "A. Gemelli", Roma, Italy.

Teaching experiences

From November 2007 to January 2008:
Descriptive Statistics at the Department of Statistical Sciences, University of Padova (Dott. B. Scarpa)

From November 2006 to January 2007:
Descriptive Statistics at the Department of Statistical Sciences, University of Padova (Prof. F. Pesarin)

Publications

Brombin C. and Salmaso L. (2008) Multi-aspect permutation tests in shape analysis with small sample size. Submitted to Computational Statistics and Data Analysis.

Brombin C., Pesarin F. and Salmaso L. (2008) Dealing with multivariate data and small sample sizes: an application to shape analysis as a particular case. Submitted to the Annals of Applied Statistics (a Festschrift in the Lecture Notes-Monograph Series of the Institute of Mathematical

Statistics (LNMS) from the Nonparametric Statistics and Mixture Models Conference).

Brombin C. and Salmaso L. (2008) Searching for Powerful Tests in Shape Analysis. Proceedings of COMPSTAT2008, International Conference on Computational Statistics, Vol. II, Contributed Papers, 3-10, P. Brito (Editor) Physica-Verlag: A Springer Company.

Caroli D., Variola A., Ferrara F., Rizzotto E. R., Baldo V., Salmaso L., Chiaramonte M., Brombin C. and Floreani A. (2008) The early diagnosis improves survival in primary biliary cirrhosis: A 35-year follow-up in a single centre. Digestive and Liver Disease, 40, A119.

Arboretti R., Brombin C., Pellizzari P., Salmaso L., Mozzanega B. (2008) Nonparametric methods applied to nuchal translucency and foetal macrosomia. JP Journal of Biostatistics, 2, 19-36. This paper is available online at "http://www.pphmj.com", JP Journal of Biostatistics.

Finos L., Brombin C., Salmaso L. (2008) Adjusting stepwise p-values in generalized linear models. Submitted to Communications in Statistics: Theory and Methods.

Guarda-Nardini L., Stifano M., Brombin C., Salmaso L., Manfredini D. (2007) A one-year case series of arthrocentesis with hyaluronic acid injections for temporomandibular joint osteoarthritis. Oral Surgery, Oral Medicine and Oral Pathology, 103 (6), e14.

Ruini C., Belaise C., Brombin C., Caffo E., Fava G. A.(2006) Well-Being Therapy in School Settings: A Pilot Study. Psychotherapy and Psychosomatics, 75, 331-336.

Brombin C., Pesarin F., Fava G. A. (2005) Multivariate analysis with paired observations and missing values: a case study (in italian). Working Paper no. 2005.13, Dept. Of Statistical Sciences, University of Padova.

Brombin C., Pesarin F., Fava G. A. (2005) Application of multivariate observation testing to a randomized controlled trial of CBT and WBT. Abstract in 2nd National Congress of Clinimetry, Bologna, 23 September 2005.