UNIVERSIDADE FEDERAL DE LAVRAS

SOME TOPICS IN PROCRUSTES ANALYSIS APPLIED TO SENSOMETRICS

ERIC BATISTA FERREIRA

2007

Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da Biblioteca Central da UFLA

Ferreira, Eric Batista.
Some topics in Procrustes analysis applied to Sensometrics / Eric Batista Ferreira. -- Lavras : UFLA, 2007.
131 p. : il.

Tese (Doutorado) – Universidade Federal de Lavras, 2007.
Orientador: Marcelo Silva de Oliveira.
Bibliografia.

1. Procrustes. 2. Sensorial analysis. 3. Missing values. 4. *Bootstrapping*. I. Universidade Federal de Lavras. II. Título.

CDD - 519.535

ERIC BATISTA FERREIRA

Some topics in Procrustes analysis applied to Sensometrics

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de "Doutor".

Orientador: Prof. Dr. Marcelo Silva de Oliveira

LAVRAS

MINAS GERAIS - BRASIL

2007

ERIC BATISTA FERREIRA

Some topics in Procrustes analysis applied to Sensometrics

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de "Doutor".

APROVADA em 13 de novembro de 2007.

Prof. Dr. Daniel Furtado Ferreira	UFLA
Profª. Dra. Rosemary Gualberto F. A. Pereira	UFLA
Prof. Dr. Fernando Antônio Resplande Magalhães	EPAMIG
Prof. Dr. Renato Ribeiro de Lima	UFLA

Prof. Dr. Marcelo Silva de Oliveira
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

PARA LUCIVANE

NÃO PODERIA TER ENCONTRADO TAMANHA DEDICAÇÃO E AMOR EM OUTRA PESSOA.

AGRADECIMENTOS (ACKNOWLEDGEMENTS)

Embora não consiga listar o nome de todos aqueles que merecem ser agradecidos e embora o faça misturando duas línguas, seguem meus principais agradecimentos.

Agradeço a Deus. Agradeço a minha família. Minha esposa, Lucivane, pela compreensão, paciência e orações nos momentos mais difíceis de ausência, principalmente durante o período que passei na Inglaterra. Meus pais, Rui e Marilene, pelo apoio incondicional e por nunca duvidarem de mim. Minha irmã, Ester, pela admiração e apoio. Meus tios, avós, primos, sogros, sobrinhos e cunhados que, mesmo distantes, torceram e torcem por mim.

Agradeço a meus amigos. Para não correr o risco de esquecer de citar pessoas importantes, me reservo ao direito de agradecer meus colegas de Pós-graduação do DEX/UFLA e da Open University (Milton Keynes, UK), meus professores, meus amigos do Karate Kyokushin, meus amigos da banda Doravante e Ensaio Aberto, meus alunos e ex-alunos da UFLA, do UNIFOR e da UNIFAL-MG, meus amigos de Juiz de Fora, Lavras, Piúma, Alfenas, Formiga, Milton Keynes (UK) e tantas outras cidades pelas quais passei durante meu doutorado; e, nominalmente, Fabyano, Denismar, Washington e Beijo, que são como irmãos para mim; e a todos os funcionários do DEX/UFLA.

Agradeço particularmente, a cinco pessoas que, para mim, são exemplos de vida e profissionalismo [1]:

- Professor e amigo Daniel Furtado Ferreira, pela enorme lição de como deve ser um cientista ético, dedicado e produtivo;
- Professor e amigo Eduardo Bearzoti, pelo exemplo de disciplina, ética e brilhantismo profissional, e, acima de tudo, por ensinar que nunca é tarde para realizarmos nossos sonhos.
- Professor e amigo Fernando Antônio Resplande Magalhães, por ter me apresentado a UFLA e Lavras, pela amizade, pelos conselhos, pelo exemplo de profissional da área de Análise Sensorial, pelo carinho e consideração;

[1] Listados em ordem alfabética.

- Professor and friend John Clifford Gower, for the friendship, for treating me so well in a foreign land, for being an exemple to be followed of brilliant Mathematician and Statistician and for the great help in exploring missing value problems in Procrustes Analysis.
- Professor e amigo Marcelo Silva de Oliveira, pela lição de humanismo e sensibilidade no fazer científico e na enorme confiança que depositou em mim durante ambos mestrado e doutorado;

Agradeço a Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), por fomentar o meu doutorado e proporcionar minha temporada na Inglaterra para aprimorar meus conhecimentos e ter experiências pessoais e profissionais inesquecíveis.

Agradeço ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) por me dar a oportunidade de prosseguir minhas pesquisas em Sensometria no Pós-doutorado.

I thank the Open University, for kindly receiving me and contributing to my human and technical formation.

E, finalmente, agradeço a minha querida Universidade Federal de Lavras (UFLA) que, além de me proporcionar a graduação, o mestrado e o doutorado, me acolheu como filho em seu leito de conhecimento e tradição. Me sinto honrado em defender minha tese em meio as comemorações do centenário dessa Instituição de prestígio nacional e internacional.

Por meu respeito, amor e admiração à UFLA, publico aqui o hino que compus para essa Instituição, que ficou colocado em quarto lugar no Concurso do Hino da Universidade Federal de Lavras, realizado em 2002.

Hino à Universidade Federal de Lavras

De um sonho estrangeiro gerada e criada com orgulho nacional
Abençoa a cidade de Lavras com a graça e glória da ESAL.

Com empenho de tantos amores, a benção de Deus e o rufar dos corações
Surge a UFLA repleta de cores contemplando as novas gerações.

Tu consegues respeito dos filhos e dos mestres tens admiração
Tu ajudas o homem do campo a desempenhar melhor sua função.

A ciência que tanto praticas eleva teu nome e de quem te seguiu
És a dama que mora em Minas, flor do coração do meu Brasil.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

RESUMO

FERREIRA, Eric Batista. **Alguns tópicos em Análise de Procrustes aplicada à Sensometria**. Lavras: UFLA, 2007. 131 p. (Tese - Doutorado em Estatística e Experimentação Agropecuária)*

Os problemas de Procrustes são estudados desde a segunda metade do século XX, quando o primeiro problema foi proposto. Desde então a Análise de Procrustes foi muito estudada e desenvolvida. Contudo ainda persistem lacunas como o estudo da estimação de parcelas perdidas e a falta de ferramentas de Inferência Estatística. O presente trabalho estuda a influência dos chutes iniciais na estimação de parcelas em problemas ordinários de Procrustes, relata e propõe novos aspectos em algoritmos de estimação em problemas generalizados de Procrustes e descreve uma proposta de método decisório para permitir a Inferência em tal análise. Esse método é ilustrado com três experimentos reais e dois estudos de simulação de dados. Os chutes iniciais mostram interferência no resultado dos ajustes e o método decisório se mostrou coerente, estável e eficiente na detecção de produtos semelhantes, associando uma fase liberal com uma fase conservativa.

*Comitê orientador: Marcelo Silva de Oliveira - UFLA (orientador); Daniel Furtado Ferreira (UFLA) e John Clifford Gower (Open University).

ABSTRACT

FERREIRA, Eric Batista. **Some topics in Procrustes analysis applied to Sensometrics.** Lavras: UFLA, 2007. 131 p. (Thesis - Postgraduation Program in Statistics and Agricultural Experimentation)*

Procrustes problems have been studied since the second half of the 20th century, when the first problem was stated. Since then Procrustes analysis has been developed. However, some gaps still hold mainly in estimating missing values and the lack of tools for statistical inference. This work analises the influence of putative values in the estimation of missing cells in ordinary Procrustes problems, reports and suggest new aspects for estimation algorithms in generalised Procrustes analysis and describes a decision method to allow inference features. Such method is illustrated with three practical experiments and two simulation studies. Inadequate putative values have shown to ability to lead to local minima and the decision method performed coherently, stably and efficiently in detecting similar products, associating a liberal and a conservative stage.

*Guidance committee: Marcelo Silva de Oliveira - UFLA (supervisor); Daniel Furtado Ferreira (UFLA) and John Clifford Gower (Open University).

1 INTRODUCTION

Hurley & Cattell (1962) did a homage to Procrustes, personage of Greek mythology, when they used his name, for the first time, to refer to a minimization problem that intended to transform a given matrix into another. Depending on how that minimization problem is set it receives a particular name, setting up a particular Procrustes problem.

Besides the importance on itself, i.e., on solving how two (set of) matrices best fit themselves only using some allowed transformations, the Procrustes problems have found a crucial role in sensorial sciences. Williams & Langron (1984), for instance, stated a sensorial methodology called Free Choice Profiling (FCP), only possible after the progress of the Procrustes analysis. On the other hand, the development of FCP certainly motivated the Procrustes problems to keep in progress (Figure 1). Free Choice Profiling allow each assessor to score his/her own set of attribute, in his/her own scale. That is what the Procrustes problems are all about, comparing configurations apart from bias of scaling, positioning or orientation.

So many others areas of science have used and have been influenced by Procrustes apart from sensorial sciences. Particularly, Statistical Shape Analysis uses Procrustes rotation from so long ago (Ian Dryden) and is the another important application of Procrustes analysis. However, it can be found a lot of others areas where Procrustes have been used, like Paleontology (Rodrigues & Santos, 2004), Medicine (Rangarajan, 2003), Ecology

(Moraes, 2004), Molecular genetics (Mironov et al., 2004), automotive industry (Dairou et al., 2004) and Photogrammetry (Akca, 2003).

Meaning of the term *Procrustes* is shown in Table 1.

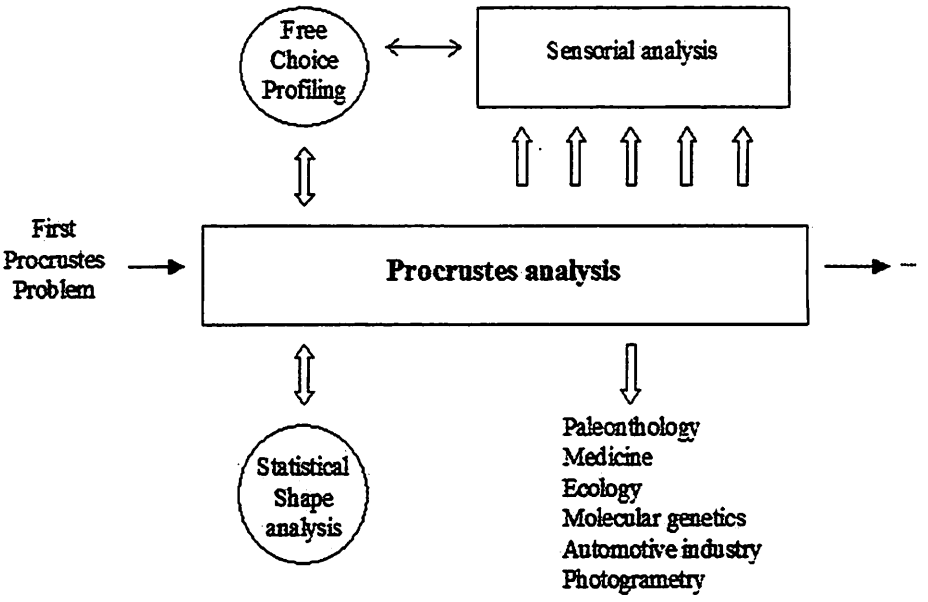


FIGURE 1: Origin and progress of Procrustes analysis and its main related areas.

The joint development of Procrustes and sensorial analysis is justified by the need of correctly modeling and comparing the subjectiveness of the human perceptions and likes. In manufactures, for instance, it is more than ever worth to measure that subjectiveness and take it into account when producing for a specific share of the market.

Due to their close relationship, some topics on Procrustes analysis are discussed here merged by sensorial issues and sensorial and shape analysis examples through all the text.

TABLE 1: Meaning of the term *Procrustes* along some areas of knowledge.

Area	**Relation with Procrustes**
Greek Mythology	Presents the myth of Procrustes, evil personage that adjusted his guests to his "magical" bed. He was murdered in his own bed.
Mathematics	Beginning of Procrustes Problems. Use of the term to refer to the problem of fitting a matrix to another.
Sociology/Economics	Procrustes is used referring to autocratic/dictatorial/unflexible systems, rules, policies, governments or regimes.
Statistics	As a multivariate descriptive tool is frequently used, mainly after Generalised Procrustes Analysis (Gower, 1975). This work proposes some stochastic point of view.
Shape analysis	Procrustes Analysis is largely used in Statistical Shape Analysis (Dryden & Mardia, 1998), fitting landmarks.
Sensometrics	Impelled by the developments of Procrustes problems, arises the Free Choice Profiling (Williams & Langron, 1984). Sensometrics arises as a multidisciplinar area intending to quantify and understand stimuli perceived by humans.

1.1 History of Procrustes

Procrustes is a Greek Mythology personage. Son of Poseidon, he had a special bed in his chamber that would have the power to adjust itself to anyone laid on it. Procrustes use to offer a great banket to the peregrines that passed by his chamber, full of food and wine. After the dinner, he offered a restful night on a magical bed. When they had no way out he announced that they should exactly fit his, what never happened. Just in case, he had two beds to ensure anyone would never fit it. In order to fit the guest to the bed, Procrustes stretched or racked his legs and arms. However, Procrustes had a tragic fait. Theseus, in his way to claim Athens reign, killed him on his own bed, as one of his six works. Hurley & Cattell (1962) did homage to that Greek Mythology personage stating the *Procrustes problem*, a mathematical fitting problem, for the first time.

Procrustes myth seems to denounce one (or several) human blemish(es): the intolerance (besides rigidity and cruelty). It denounces a dark side of human beings that tends to impose its will at any cost, in spite of chopping and stretching.

1.2 Definitions and notation

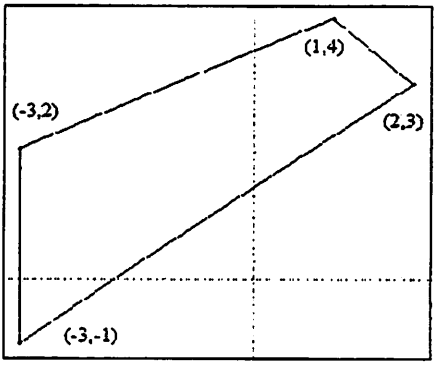
In this section some basic definitions are presented and the notation adopted along the text is stated.

Reading the following definitions, consider that, mainly in generalised Procrustes problems, some definitions use index i to refer to several matrix (e.g. X_i, $i = 1, \ldots, k$).

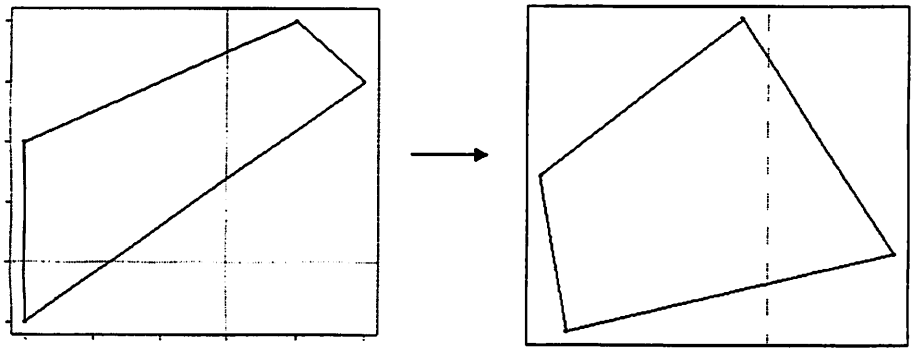
Moreover, along the text, the sign $'$ denotes transposed matrices (e.g. X' is X transposed); tr stands for the *trace* of a matrix (e.g. $tr(A) = trace(A)$); and the sign $*$ denotes the element-wise product of two matrices.

X: a given matrix $(n \times p)$ of raw scores, distances or transformed data. It can be seen as a configuration, i.e., let each row of X represent the coordinates of a point. So, the n points can be joint by an imaginary line forming a configuration. Because of that correspondence between a matrix and a polygon, there is meaning saying "rotation of a matrix", "translation of a matrix", and so on. The next example shows a configuration on a plan.

$$\begin{pmatrix} 1 & 4 \\ 2 & 3 \\ -3 & -1 \\ -3 & 2 \end{pmatrix} \longrightarrow$$

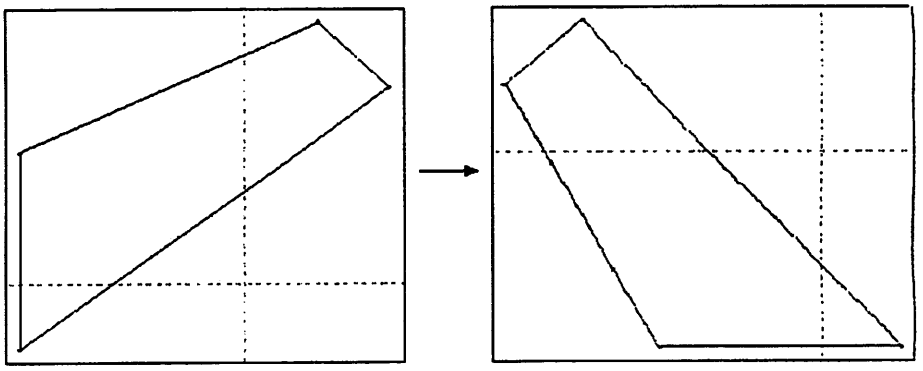


T: a transformation matrix $(p \times p)$.[1] It is unknown (i.e., has to be estimated) and promotes a general transformation in X when post multiplying it. The illustration shows a given configuration X (left hand side) and a transformed one XT (right hand side). Note that the general transformation does not necessarily respect the relative distances between vertices, i.e., it *must not* be called a rigid body transformation.



[1] T is not necessarily square. Since the number of columns p may vary along the matrices X, it will be denoted by p_i, so T turns out a $p_i \times p_{i'}$ matrix.

Q: an orthogonal matrix ($p \times p$). One matrix Q is considered orthogonal when it is square and both column orthonormal ($QQ' = I$) and row orthonormal ($Q'Q = I$). In this case $Q^{-1} = Q'$. It promotes a rigid body rotation in X when post multiplying it. The illustration shows a given configuration X (left hand side) and a rotated one XQ (right hand side).

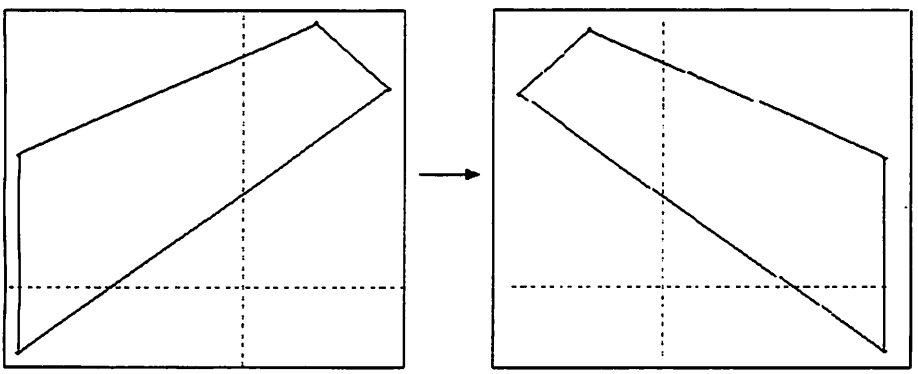


It is worth noting that terminology usually is a source of confusion in the literature concerning transformations. Here, I transcribe such a reflection about that found in Gower & Dijksterhuis (2004):

"There is a considerable confusion in the literature between the terms orthogonal and orthonormal. The prefix **ortho** *refers to the inner product of two vectors being zero (geometrically they represent perpendicular directions) and* **normal** *refers to vectors of length one. Thus an orthonormal matrix P ($p_1 \times p_2$) satisfies $P'P = I$. More precisely P, so defined, is column-orthonormal; we may also have row orthonormality, defined by $PP' = I$. By an orthogonal matrix Q, we mean a square matrix, in so that $QQ' = Q'Q = I$. Actually for square matrices, row orthonormality implies column orthonormality and vice versa; thus a $p \times pn$ orthonormal matrix is very common in the literature but it is a little misleading because from the etymological point of view we would understand only the diagonality of QQ' and $Q'Q$ and no the unit-length. However, orthogonal matrices are of such importance that some special terminology is needed to distinguish*

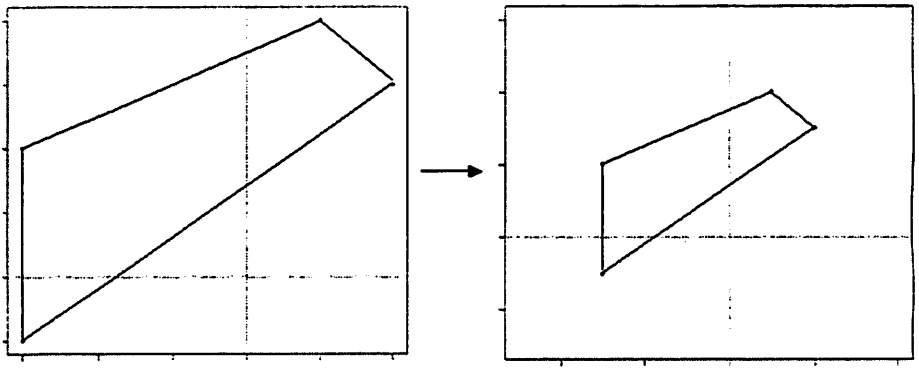
them from non-square orthonormal matrices. In the literature, what we call orthonormal matrices are sometimes described as orthogonal, which is etymologically correct but a source of confusion. Another source of confusion in the literature is that quite general matrices T may be referred to as representing rotations whereas, strictly speaking, rotation only refers to orthogonal matrices and then not all orthogonal matrices."

P: a column orthonormal matrix $(p \times p)$[2], i.e., $PP' = I$. It promotes a reflection in X when post multiplying it. The illustration shows a given configuration X (left hand side) and a reflected one XP (right hand side).



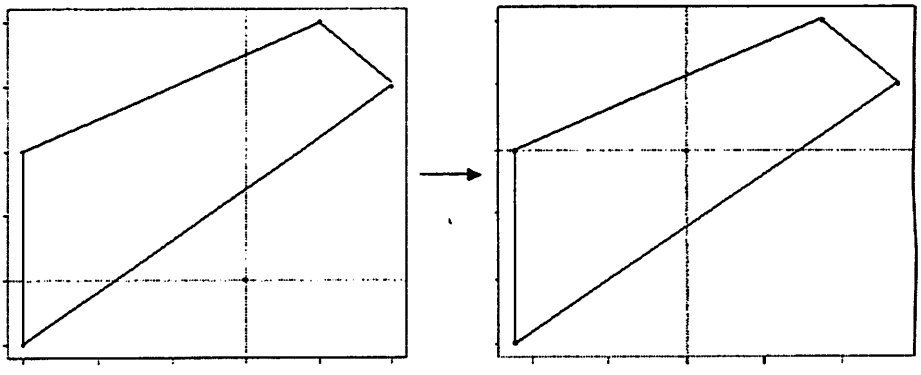
s: a scaling factor or central dilation (scalar). Usually associate with a matrix X, it stretches or shrinks the configuration X represents. The illustration shows a given configuration X (left hand side) and a shrunk one sX (right hand side).

[2] P is not necessarily square either. When X_i is $n \times p_i$, then P is $p_i \times p_{i'}$ but constrained to $PP' = I$



u: a vector containing the average of the X columns, i.e., the centroid coordinates of the X configuration. In the translation step, $1u'$ represents the matrix to be subtracted from X to center it.

1: a vector of ones and of the suitable length. The illustration shows a given configuration X (left hand side) and a centered one $X - 1u'$ (right hand side).



1.3 Review of Procrustes Problems

According to Gower & Dijksterhuis (2004), there are three main elements in Procrustes history, the *unfortunate traveler* (matrix X_1), the *bed* (matrix X_2) and the *stretch* or the *rack* to turn him of the same size of the bed (matrix T). The simplest statement of the Procrustes problem

seeks for a matrix T to minimize

$$\|X_1T - X_2\|,$$

where T is a matrix with dimensionality $p_1 \times p_2$; X_1 and X_2 are given matrices with dimensionality $n \times p_1$ and $(n \times p_2)$, respectively; and $\|\cdot\|$ denotes the Euclidean/Frobenius norm, that is $||A|| = trace(A'A)$. In sensorial analysis, those quantities can be understood as: n, number of evaluated objets; p_1, number of variables used by assessor 1 to evaluate such objects; p_2, number of variables used by assessor 2. Note that, since T is a general matrix, if X_1 is invertible, there is a matrix $T = X_1^{-1}X_2$ which is a ready solution for the problem, i.e., there is T such that X_1 fits exactly X_2. But, in order to T equals $X_1^{-1}X_2$, it is necessary X_1 to be invertible, that is, X must be a square matrix *and* must have non-zero determinant. In general, X_1 neither is square nor is guaranteed to have non-zero determinant. So, the minimization problem proceeds.

$$\left\| \begin{bmatrix} x_{11}^{(1)} & \dots & x_{1p_1}^{(1)} \\ \vdots & \ddots & \vdots \\ x_{n1}^{(1)} & \dots & x_{np_1}^{(1)} \end{bmatrix} \begin{bmatrix} t_{11} & \dots & t_{1p_2} \\ \vdots & \ddots & \vdots \\ t_{p_11} & \dots & t_{p_1p_2} \end{bmatrix} - \begin{bmatrix} x_{11}^{(2)} & \dots & x_{1p_2}^{(2)} \\ \vdots & \ddots & \vdots \\ x_{n1}^{(2)} & \dots & x_{np_2}^{(2)} \end{bmatrix} \right\|.$$

The minimization problem is called a Procrustes problem because X_1 is transformed by T to best fit X_2. Under that ordinary point of view, T is unrestricted real matrix. According to Gower & Dijksterhuis (2004), that is a multivariate multiple regression problem (1.1). Therefore, one can derive a least squares estimator[3] of T.

[3]Demonstration in (A.1).

$$\hat{T} = \left(X_1' X_1\right)^{-1} X_1' X_2. \tag{1.1}$$

The groundwork for Generalised Procrustes Analysis (GPA) was in Factor Analysis laid by Green (1952), Cliff (1966) and Schönemann (1966). They solved the problem of orthogonally transforming one matrix X_1 to another X_2 only allowing an orthogonal transformation, minimizing the least squares criterium. It is called the *orthogonal Procrustes problem* (1.2). The two configurations had an equal number of columns. The solutions of Green and Cliff only work for matrices of full column rank, while Schönemann generalised the solution to matrices of deficient column rank.

$$\|X_1 Q - X_2\|, \tag{1.2}$$

where Q is an orthogonal matrix.

Schönemann & Carrol (1970) proposed the first extended problem. Their problem allowed a rotation, a translation and the multiplication by a scaling factor for adjusting the two given matrices. That was called an *extended orthogonal Procrustes problem* or *Ordinary Procrustes Analysis* (OPA), equation 1.3.

$$\|s(X_1 - 1u_1')Q - (X_2 - 1u_2')\|, \tag{1.3}$$

where 1 is a vector of ones with n positions and u_i is a vector of length p_i, that is, $1u_i'$ is a translation matrix $n \times p_i$; and s is a scalar factor associated with X_1 to stretch or shrink it, in order to turn it as similar as possible to X_2. Note that, if the elements of u_i are the average of the columns of X_i, the matrices as translated to have their centroids superimposed at the origin.

Gower (1971) had taken the first steps towards the *generalised orthogonal Procrustes analysis* (GPA or GOPA). In that context, X_1 and X_2 are replaced by k sets $X_1, \ldots, X_k$. This may have been the first instance where the matrices $X_1, \ldots, X_k$ are regarded as configuration matrices rather than as coordinates matrices of factor loadings (Gower & Dijksterhuis, 2004).

Studying factor loadings, Kristof & Wingersky (1971) considered minimizing the multi-set criterion $\sum_{i<i'}^{k} ||X_i Q_i - X_{i'} Q_{i'}||$ to determine the orthogonal matrices and to give an average "factor matrix" referred to by $G = \frac{1}{k} \sum_{i=1}^{k} X_i Q_i$.

Gower (1975) introduced GPA minimizing the same criterion as Kristof & Wingersky (1971) but introducing scaling factors for each set and viewing the whole problem as one of matching configurations (1.4). He was also the first to coin the term *Generalised Procrustes Analysis* and to introduce scaling factors s_i $(i = 1, \ldots, k)$ for each matrix X_i. The problem follows:

$$\sum_{i<i'}^{k} \left\| s_i (X_i - 1u_i') Q_i - s_{i'} (X_{i'} - 1u_{i'}') Q_{i'} \right\|, \tag{1.4}$$

what is equivalent to minimize

$$k \sum_{i=1}^{k} ||s_i (X_i - 1u_i') Q_i - G||, \tag{1.5}$$

where G is called *group average configuration* (1.6),

$$G = \frac{1}{k} \sum_{i=1}^{k} s_i (X_i - 1u_i') Q_i. \tag{1.6}$$

Gower (1975), in the generalisation of the Procrustes analysis, presents an iterative algorithm that seeks for the parameters of interest (Q_i, s_i, $1u_i'$ e G) while the euclidian distances between similar vertices along the configurations are minimized. In 1977, Ten Berge corrected Gower's method for the determination of the central dilations, and improved Gower's procedure for the calculation of the orthogonal transformations. Ten Berge & Knol (1984) extended the problem to the case where two or more configurations have a different number of columns (although only for orthogonal transformations). More recently, Peay (1988) has solved the problem of fitting more than two configurations with differing dimensionalities including translations and central dilations.

According to Lingoes & Borg (1978) Generalised Procrustes Analysis can be seen as a model that treat each configuration of a panel as a mean configuration plus an error (1.7). That statement is possible since those transformations do not affect the relative distances between the stimulus[4] points (rows of X_i). If a generalised Procrustes analysis does not result in a satisfactory match, that is, if the minimum residual sum of squares is still a big number, this may be due to the fact that the relations between the configurations are of a more complex nature (Commandeur, 1991).

$$s_i (X_i - 1u_i') Q_i = G + E_i \tag{1.7}$$

[4] *Stimulus* is another term used to refer to the rows of X_i. *Stimulus* can be considered as a synonym of *objects* and *products*. By the way, *subjects* is another term for *assessors*, used by psychologists. It can be found along the text.

where E_i is a $(n \times p)$ residual matrix and $i = 1, \ldots, k$. That is the first of the PINDIS (Procrustean INdividual DIfferences Scaling) models brought by Lingoes & Borg (1978) and was called *the similarity transformation model* or *Generalised Procrustes Analysis* (1.7). In the second and in the third PINDIS models, (1.8) and (1.9), a transformation called *dimension weighting* is introduced, that is, dimensions are allowed to be stretched or shrunk differently when matching n configurations. Follow the second and the third PINDIS models, respectively,

$$(X_i - 1u_i')Q_i = (G - 1g_i')SW_i + E_i \tag{1.8}$$

and

$$(X_i - 1u_i')Q_i = (G - 1g_i')S_iW_i + E_i, \tag{1.9}$$

where g_i is a translation vector for the group average configuration specific for assessor i; S and S_i $(p \times p)$ are orthogonal matrices and W_i is a diagonal matrix of weights $(p \times p)$.

The phycological interpretation is that in (1.8) the assessors agree on the underlying dimensions that structure the stimuli under investigation (that is, the axes of GS), but differ in the importance they attach to the dimension on which the stimuli are ordered; on the other hand, in (1.9) they not only differ in the importance they attach to the dimension, but also have their own views on *which* dimensions of G are to be considered as important. That is represented by the index i in S_i (Commandeur, 1991). These PINDIS models are less parsimonious than the GPA model since they require the estimation of more parameters than in GPA model. This means that the dimension weighting models always fit the data equally as

well as or better than the GPA model.

The forth (1.10) and the fifth (1.11) PINDIS models proposed by Lingoes & Borg are called by Comnandeur (1991) as *stimulus weighting models*, as follows:

$$(X_i - 1u_i')Q_i = V_i(G - 1g') + E_i \tag{1.10}$$

$$(X_i - 1u_i')Q_i = V_i(G - 1g_i') + E_i \tag{1.11}$$

where g and g_i are translations vectors; V_i is an unknown diagonal matrix of stimuli weights.

The essential difference between the two stimulus weighting models is that G is only translated once in (1.10) while it is translated differently for each configuration i in (1.11).

Lingoes & Borg (1978) even discuss a sixth model where the combined effects of dimension weights and stimulus weights are investigated. Commandeur (1991) call it a *double weighting model.*

However the algorithms for estimating the parameters of PINDIS models brought by Lingoes & Borg (1978) present a list of problems. Commandeur (1991) brings improved algorithms for those estimations, generalisations to the p-dimensional case for all estimation procedures, algorithms to deal with incomplete configurations (missing rows) and analysis of variation allowing one to assess the relative contributions of configurations, stimuli and dimensions to the total fit of each model.

Most of the extensions of GPA typically are concerned with data sets consisting of matrices having unequal number of columns. However,

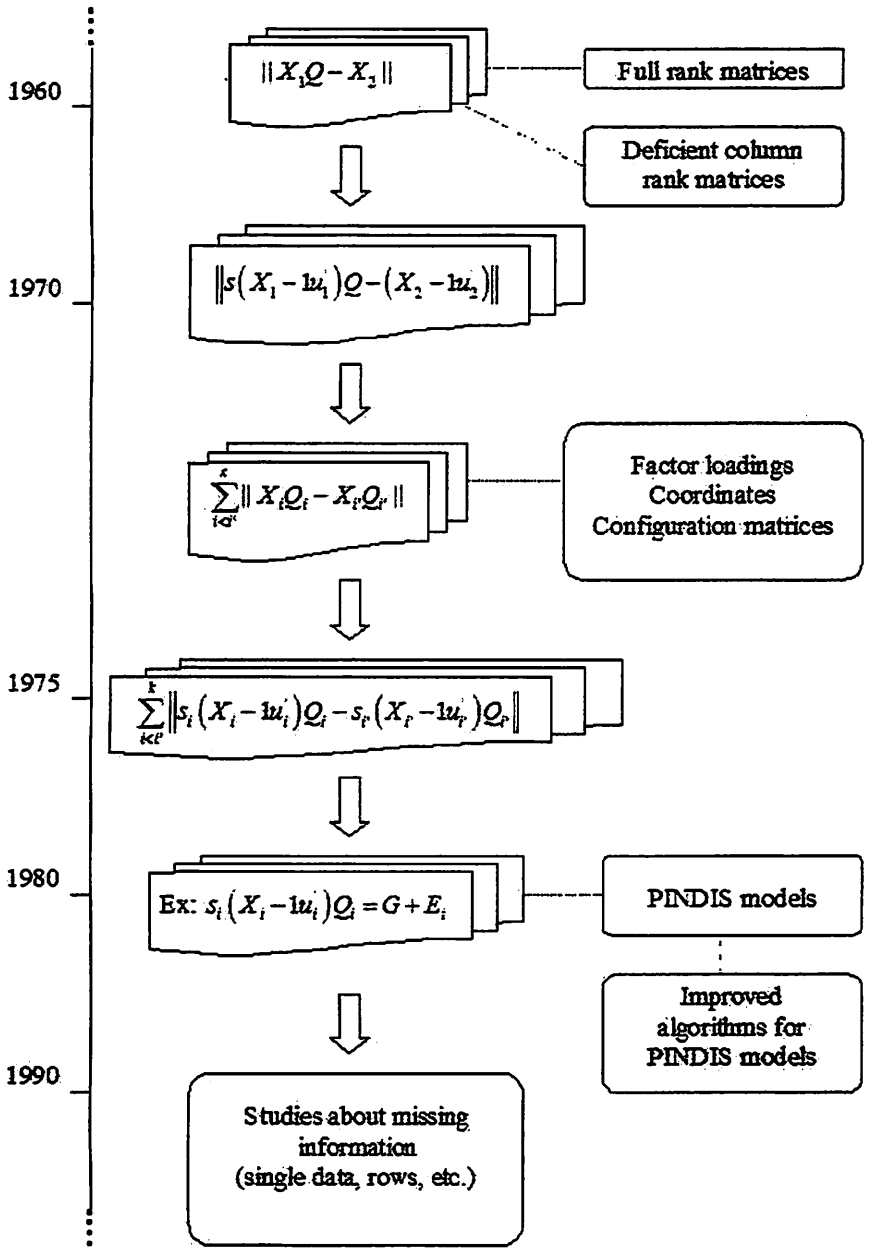


FIGURE 2: Scheme summarizing Procrustes problem since its origin until recent days.

the case of different number of rows (that is, missing information about one or more stimuli) are mentioned by: (i) Everitt & Gower (1981) apud Commandeur (1991), in the context of what they called a *weighted generalised Procrustes method*, where they did not incorporate central dilations; and De Leeuw & Meulman (1986) apud Commandeur (1991); (ii) in Jackknife context, where they allowed only one missing stimulus per matrix. Commandeur (1991) bring a matching procedure called MATCHALS that allows any pattern of missing rows.

A scheme is presented in Figure 2 roughly summarizing the path of Procrustes problems.

1.4 Sensorial Analysis

Sensorial analysis or Sensory analysis is a very wide field of the science concerned to understand, describe, measure and/or even reproduce the mechanisms of perception of external stimuli by human basic senses. The word *sensorial* comes from Latin *sensus*, which means *sense* (Anzaldua-Morales, 1994).

Mueller (1966) says that the senses are our *highways to the outside world.* Though that phrase is not very related to senses, it remind us that the only way to perceive the world is through our basic senses. Everything one knows besides his/her instincts came to him/her by the basic senses (Table 2).

Depending on how one looks to that system, it receives a different name. For instance, when one wants to understand how a person perceives an object (focus on the person), we are in the field of Psychology. It can receive the name *Sensory Psychology* and/or *Psychophysics* when regards the physiology of perception and physical phenomena related to perceiving something (Mueller, 1966); and *Experimental Psychology* or *Psychometrics*,

TABLE 2: Comparison of human basic senses according to the number of discernible stimuli, sharpness of discrimination, average speed of reaction and duration of sensation (adapted from Amerine et al., 1965).

	Stimuli (#)	**Discrimination** (%)	**Speed** (sec)	**Duration** (sec)
Sight	40,000	1	0.189	0.03
Hearing	15,000	33	0.146	0.002
Smell	?	25	?	?
Touch	$3-4$ (?)	33	0.149	10^{-3}
Taste	4 (+,?)	5	$0.3-1$	?

when effort is made to measure the senses (O'Mahony, 1985). On the other hand, when the main interest of the researcher is to understand how the object is perceived by someone (focus on the object), we are in the field of *Sensory Evaluation* (Noronha, 2003). Therefore, Sensory Evaluation is a group of techniques used to investigate, for instance, whether a tiny modification in a manufacture process can be detected by a consumer.

A huge subset of Sensorial analysis is the Sensometrics. It regards to every effort made to measure stimuli been perceived by human basic senses. That measurement processes is consider Sensometrics in spite of whether the focus is on person or object, though most works in Sensometrics are focussed on objects.

Sensometrics arose into 50's and 60's and is a kind of modern term to refer to Statistics applied to Sensory Sciences (Heymann, 2006). One of Sensometrics' strongest characteristics is that, in the last decades, multivariate Statistics have been the preferred set of tools for exploring sensations (Sinesio et al., 1992; Daems & Delvaux, 1997; Hummer, 1998; Thybo & Martens, 2000; Allcroft et al., 2007). Due to its multidisciplinary profile, Sensometrics is able to congregate applied statisticians, experimental psy-

chologists, food engineers, chemical engineers, mathematicians, physicists, doctors, marketing people and so many other practitioners interested in human senses.

Though Sensory Evaluation can be used to investigate how any kind of object is perceived by human senses, its use is by far greater in food science (Ferreira, 2004; Jorge, 1989; Aquino, 1992; Aguiar, 1995; Cardoso, 1994; Malvino Madrid, 1994; Chagas, 1994; Magalhães, 2002).

However, other products can be studied regarding the sensations they produce. For instance, automotive industry is an increasing field of studies. Sensation of car breaks and interaction between consumer, car and road have been studied by Dairou et al. (2004) and Astruc et al. (2006), respectively. The cosmetic industry is also very developed due to constant investigations of mainly odor and color perceptions (Wortel et al., 2000; MacDougall, 2000; Chrea et al., 2004).

In a sensory context, in a large sense, a set of people evaluate a set of objects. When those people are trained assessors or experts, they are called a *panel*. When they are not trained at all and are searched at marketplace, they are called *consumers*. In that context, one even can imagine *degrees of training*, i.e., people who are rapidly trained or even are used to the sensory world, were trained before but not for the object at issue. The prior can be understood, therefore, as a *semi trained* assessor. At last, there are untrained people who are not asked for scoring objects in their usual marketplaces (but anywhere else). Lets call them *untrained assessors*.

How was said above, in Experimental Psychology the main focus of the work is the person. So, attention has to be paid to the fact that one is seeking for *random samples* of people, i.e., they should be random elements of a *population* of interest for which conclusions and statements are to be maid (O'Mahony, 1985). On the other hand, when a panel of

trained assessor or experts is used to evaluate something, it has to be clear that it is not a random sample of a population of consumers. They were screened from a population of consumers for been skillful, trained to be able to precisely detect several sensorial attributes and distinguish between products supposedly similar.

In the sensory evaluation of foods (or other products), there can be more than one goal. In some situations, people may be used for a preliminary analysis to measure an attribute of the food and provide clues for later chemical or physical analysis. In others, they may be used to measure, say, flavor changes due to a change in the processing of the food. They can describe a product or compare a set of possible recipes to the market leader. Consequently, assessors can be evaluated with respect to the precision (variance), exactness, accuracy, i.e., success of the training process. In turn, consumers can be used to evaluate preference for one product rather than others, indicate the level of that preference or even say whether they would buy it.

Depending on how the group of assessors is chosen, the factor *assessor* can be assumed to be a *fixed* or a *random* effect. For instance, a panel comprising of a few people is selected and trained to become judges in the specific task required, whether it be detecting off-flavors in beverages or rating the texture of cakes. That would be a fixed effect. Generally, potential judges are screened to see whether they are suitable for the task at issue. This careful selection of judges is in no way a random sampling and is the first reason why the judges could not be regarded as representative of the population in general. The training then given to the selected judges will usually make them more skilled in their task than the general population would be; this is the second reason why they can no longer be regarded as presentative of the population from which they were drawn. Those judges are not a sample to be examined so that conclusions can be

drawn about a population; they *are* the population (O'Mahony, 1985).

Statistically thinking, a good training is expected to scale their mean scores (vector of means, in a multivariate sense) closer to the parameter, i.e., remove possible natural bias specific for that person; and decreases their variance (covariance matrix). Note that as a good training can remove the bias more and more, a bad one can insert a bias not observed since then!

Considering the whole process of selection, training and scoring to be an experiment, the panel can be considered a sample from a theoretical population of trained assessors. That enables statistical inference. Usually, this is not treated this way because the theoretical population of trained assessors is kind of immaterial and judged as of low importance by some. The panel is frequently consider the population itself and each assessor is like a measurement instrument. O'Mahony (1985) says that *when the focus of the study is very much on the food, the judges are merely instruments used for measuring samples of food. Logically, one good judge would be adequate for this purpose, as is one good gas chromatograph or atomic absorption spectrometer. More than one judge is used as a precaution because unlike a gas chromatograph, a judge can become distracted; the cautious use of second opinion can provide a useful fail-safe mechanism.* Apart from that, one can consider that theoretical population particularly when dealing with semi trained assessors.

On the other hand, if one wants to investigate how regular people perceive a particular product, it can be done drawing untrained assessors from the potential buying public (at the marketplace or not). Then from this sample, inferences can be made about the population of untrained people in general. The experiment now becomes similar to one in Sensory Psychophysics.

In both cases, it is worth considering each assessor to be a block. He/she has exactly the attributes of a block, like homogeneity within and

heterogeneity between scores. But the differences between considering random or fixed effects becomes important for analysis of variance, where different denominators are used for calculating F values. Where conclusions apply only to the judges tested, the judges are said to be a fixed effect. Where conclusions apply to the whole population from which the judges were drawn, the judges are said to be a random effect.

According to O'Mahony (1985), in Psychology or Psychophysics, people are the subject of the investigation and thus tend to be called *subjects*. In sensory evaluation, or sensory analysis as it is also called, people are often specially selected and trained and tend to be referred to as *judges*. In sensory analysis, judges are tested while isolated in booths; experimenters and judges communicate by codes, signs or writing. In Sensory Psychophysics, the experimenter and subject often interact and communicate verbally; this requires special training for experimenters so that they do not influence or bias on the subjects. The methods of sensory analysis are often derived from those of Psychophysics, but care must be taken when adapting psychophysical tests to sensory evaluation. This is because the goals of the two disciplines are different and they can affect the appropriateness of various behavioral methods of measurement used, as well as the statistical analysis employed. In Psychophysics, the aim is to measure the "natural" functioning of the senses and cognitive processes. Extensive training will not be given if it alters a subject's mode of responding or recalibrates the person, so obscuring his or her "natural" functioning.

2 MISSING VALUES IN PROCRUSTES PROBLEMS

2.1 Orthogonal Procrustes Rotation

The minimization problem of transforming one given matrix X_1 by a matrix, say T, such that best fits a given target matrix X_2 is called a Procrustes Problem. The term Procrustes Problem is due to Hurley & Cattel (1962) whose suggested the problem of transforming one matrix into another by minimizing the Residual Sum of Squares (RSS)

$$||X_1T - X_2||$$

for the first time. Here, T refers to a general transformation matrix but Schönemann (1966) solved that problem for a restricted case where Q, instead of T, is orthogonal. Of course, when there are no restrictions, that problem can be solved as a multivariate multiple regression, i.e.,

$$T = (X_1'X_1)^{-1}X_1'X_2.$$

In the orthogonal case, i.e., when the general transformation T is replaced by an orthogonal matrix Q, that problem is called Orthogonal Procrustes Rotation (OPR). In fact, if one consider that the post multiplication by an orthogonal matrix is itself a rotation (Gower & Dijksterhuis, 2004), that term could be consider a pleonasm. The terms *Procrustes Rotation* (PR) or *Ordinary Procrustes Rotation* (OPR), where *Ordinary* refers to

the match of only two configurations, would be more suitable. In this case, Q provides a rigid body rotation in X_1 in order to best fit X_2, leaving the relative distances among its rows untouched (Figure 3). Solution for that case can be achieved by $Q = VU'$, where V and U come from the singular value decomposition of $X_2'X_1 = U\Sigma V'$ (Schönemann, 1966).

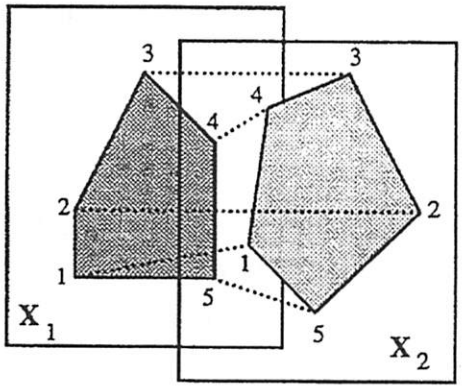


FIGURE 3: Illustration of rotating a matrix X_1 to fit X_2, in order to show the corresponding vertices. Source: Commandeur (1991).

For a better match, translations to the origin can be done and a scaling factor can be allowed in the orthogonal case, modifying the problem to

$$||s(X_1 - 1u_1')Q - (X_2 - 1u_2')||.$$

where u_i is a vector of column means of the *ith* configuration. Gower & Dijksterhuis (2004) consider both configurations already centered, turning out in a problem with simpler notation (2.1),

$$||sX_1Q - X_2||. \tag{2.1}$$

Furthermore, they show that such scaling factor may now be estimated by minimizing (2.1) subject to the constraint that Q must be orthogonal, given

$$s = \frac{trace(X_2'X_1Q)}{||X_1||} = \frac{trace(\Sigma)}{||X_1||}.$$

Both s and Q are usually estimated in an iterative algorithm that seeks for the minimum of (2.1), until convergence.

2.1.1 Missing values

Missing values are typical in sensory experiments due to several limitations: (i) physical (fatigue); (ii) financial (samples and consumers); (iii) operational (samples); (iv) sections duration; (v) eventual losses (several reasons).

In practical situations, X_1, X_2 or both may present missing values due to either the impossibility of collecting all data or loss of information. In a general statistical problem, one has basically two ways of dealing with missing data: (i) modeling the problem without the missing information or; (ii) estimating it under some constraint. Here, one is concerned in estimating them in some way. Then, the algorithm can be modified to iteratively alternate steps of missing values estimation and Residual Sum of Squares minimization (Gower & Dijksterhuis, 2004).

According to Gower & Dijksterhuis (2004), the estimates of the missing values must be such that the RSS between configurations is minimum. It is clear that one wants the estimates to be as similar as possible to the missing values, but it is not possible to formally impose that constraint since the only information one has about the missing values are, of course, the non-missing values. Nevertheless, in typical Procrustes situations, i.e., when the columns of the matrices are not variables, but dimensions, the remaining information tell us even less about the missing values.

In iterative algorithms, as the one suggested by Gower & Dijksterhuis (2004) for two matrices, it is necessary to start the process of estimation/minimization by setting putative values in the missing cells. Then, the following question is natural: Does any putative value lead to the same

estimate? Of course, when one has to minimize functions that have some local minima, the answer can be negative. In this case, one has to seek for the best way of determining them.

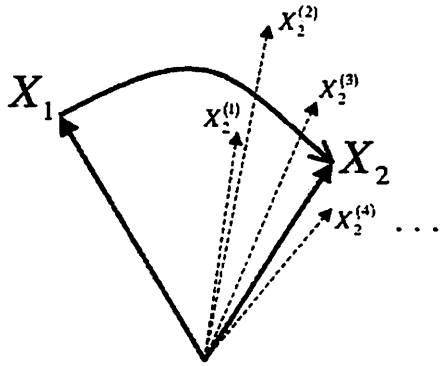


FIGURE 4: Scheme of the distances (RSS) between X_1 (before rotation) and the unknown full data X_2 and the possible arbitrary configurations $X_2^{(i)}$, $(i = 1, 2, 3, 4, \ldots)$ according to the putative values chosen.

For instance, let X_2 present some missing values. There is an infinity of ways of filling its missing cells with putative values, therefore, an infinity of different initial configurations $X_2^{(i)}$ $(i = 1, 2, \ldots)$, most of them likely different from the unknown (full) configuration X_2 (Figure 4).

Summarizing, three information sources are used to estimate the missing cells: (i) the information contained in the non-missing data; (ii) the imposed constraint and; (iii) the information brought in by the putative starting values. The first one tends to be the weakest; the second tends to be the strongest; and the level of interference of the the third is not well known and is going to be investigated in this paper.

With real and simulated data, we observe the behaviour of the estimation algorithm for two matrices described by Gower & Dijksterhuis (2004) imposing the constraint that minimizes the RSS. Simulating different levels of loss for non-error and real sized error data sets, we seek for the

best way of establishing putative values to start the iterative process.

2.1.2 Methodology

Let W_i $(i = 1, 2)$ be indicator matrices which elements are unit where the correspond position in X_i is missing and zero otherwise, for instance

$$W_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

If X_2 has missing data, one can estimate them constrained to

$$||X_1 Q - (X_2 * W_2)|| \tag{2.2}$$

to be minimum. The whole process begins by inserting putative values in the missing cells of X_2 and finding the Q that minimizes (2.2). In the estimation step one must replace the missing cells' values by the estimates given by the estimator that minimizes (2.2), i.e.,

$$W_2 * (X_1 Q).$$

With configuration X_2 updated, one now updates Q in minimization step and iterate those steps until the RSS converges to a minimum.

If, instead, X_1 has missing values, one can just reverse the roles of X_1 and X_2 in the procedure described above, noting that $||(W_1 * X_1)Q - X_2|| = ||(W_1 * X_2)Q' - X_1||$. Therefore, the estimator for the missing values of X_1 is $W_1 * (X_2 Q')$. Finally, when they both have missing values, both procedures must be alternated until convergence.

Simulation study

For testing the behaviour of such procedures according to the number of missing values, it was considered matrices X_1 and X_2 (6×3) in situations of perfect match (error-free) and plausible sized error (using empirical data). It was simulated the loss of 5%, 10%, 25% and 50% of the data in each X_i $(i = 1, 2)$ and 24 combinations of them (Table 3). Those combinations correspond to lose 3%, 6%, 8%, 11%, 14%, 17%, 19%, 25%, 28%, 31%, 39%, and 50% of the whole information. Each single combination was repeated 500 times. Of course, the combination of no loss in both matrices was the full data situation, used as a reference.

TABLE 3: Simulated loss in X_1, X_2 and in both matrices, expressed as a percentage of the cells.

Loss in X_1	Loss in X_2	Loss in the whole system
5%	0%	3%
10%	0%	6%
25%	0%	14%
50%	0%	25%
5%	5%	6%
10%	5%	8%
⋮	⋮	⋮
50%	50%	50%

For the error-free situation, one fixed matrix X_1 was rotated by a random orthogonal matrix Q for generating the target X_2. For the plausible error situation, one empirical data set was used (Lele & Richtsmeier, 2001), who measured six landmarks on nine Macqaque skulls.

In order to check the influence of different putative starting values in the missing cells, three sources were considered: Gaussian and Uniform distributions and a fixed scalar.

Gaussian distributions were defined by parameters μ_c and σ_c^2, where μ_c represents the average of the non-missing cells in one column. Then, one distribution was used to draw values for each column that presented missing cells. When a whole column was missing, μ_c was set to zero. Parameter σ_c^2 represents the variance in column c. When the whole column was missing, σ_c^2 was set to one. Note that this procedure is repeated along all columns that contain missing cells both in X_1 and X_2.

Uniform distributions were defined by parameters a_c and b_c. The lower and upper limits $a_c = -\sigma_c\sqrt{3}$ and $b_c = \sigma_c\sqrt{3}$, were σ_c is the standard deviation of column c, of course, when c contains missing cells.

The scalar value was set to be zero because all configurations were centered before the algorithm to begin. Then, zero was always the mean of every column.

It is clear that when one inserts a putative value coming from a distribution instead of a fixed mean value, he/she is inserting variability in the process. However, if the algorithm leads to a global minimum, all answers should be the same. Of course, where the putative values are more far away from the mean, it is expected the convergence to happen after a higher number of iterations.

The behaviour of the estimates was evaluated along all situations, namely, the combinations of two *types of error* size: non-error (perfect match) and real sized error (empirical data set); three *sources of putative values*: Gaussian distribution, Uniform distribution and scalar zero; and all levels of *induced loss*. To check the behaviour of the estimates five parameters were examined, ω, δ, ϕ, η and R^2. Their mean and the standard deviation were recorded along 500 replications.

(a) Residual Sum of Squares discrepancy (ω):

$$\omega = RSS - RSS_e,$$

where RSS is the Residual Sum of Squares after adjustment of the full data sets and RSS_e is the Residual Sum of Squares after match the data sets under loss and posterior estimation.

That parameter ω indicates how similar the estimated data sets are from the full data sets in terms of match. It is one of the possible indices of quality of estimation. The closer to zero ω is, the similar are the RSS's. However, it is worth noting that is expected the RSS_e to be smaller than RSS since the estimation constraint imposes that the estimates are such that the RRS is minimum.

(b) Mean squared loss (δ):

$$\delta = \frac{\sum_{i=1}^{2} \sum_{j=1}^{m_i} \left(x_{ij}^M - \hat{x}_{ij}^M \right)^2}{m_1 + m_2},$$

where m_i is the number of missing values in X_i $(i = 1, 2)$; x_{ij}^M is the real value of the jth missing cell in X_i $(j = 1, \ldots, m_i)$; and $\hat{x}_{ij}^M$ is its estimate.

That is a really quality of estimation index. It achieves the mean quadratic distance between the real value and the estimate of a missing cell, along all of them. Clearly, the smaller is δ, the better are the estimates.

(c) Scaling factor discrepancy (ϕ):

$$\phi = s - s_e,$$

where s is the scaling factor obtained when one matches the full data sets and s_e is the scaling factor when the sets have estimates in their missing positions.

Of course, one wants s_e to be as similar as possible to s, but the behaviour of ϕ along an increasing number of missing values is not trivial.

(d) Number of iterations until convergence (η).

That parameter indicates efficiency. Besides achieve the right answer (the global minimum of the function), one wants to do that as fast as possible. Though, one wants η to be small.

(e) Explained variation (R^2):

$$R^2 = \frac{Var(fitted)}{Var(Total)} = \frac{2strace(\Sigma)}{s^2||X_1|| + ||X_2||}.$$

In this case, R^2 suggests how similar X_1 transformed can be to X_2. Again, this parameter indicates quality of match. The closer to one, the better the match.

All computations were performed by specifically made R functions (Appendix B).

2.1.3 Results and discussion

Results were divided into two sections: (a) Exact match and (b) Real sized error. Here, is explored the behaviour of each parameter along the increasing value of missing cells in the whole system. It could be explored the behaviour along missing cells in just one matrix, but it would be redundant, since they were quite similar.

2.1.3.1 Exact match

In this situations there is no error, since X_2 is a perfect transformation of X_1. According to Figure 5a one notes that the discrepancy between RSS of full data sets and estimated data sets is small with few or too many

missing values. That is easy to understand since when one has few missing values, the situations are very similar; and when one has too many ones, their estimates are such that the RSS is minimum. Since the original RSS is already zero (perfect match), then the discrepancy tends to decrease.

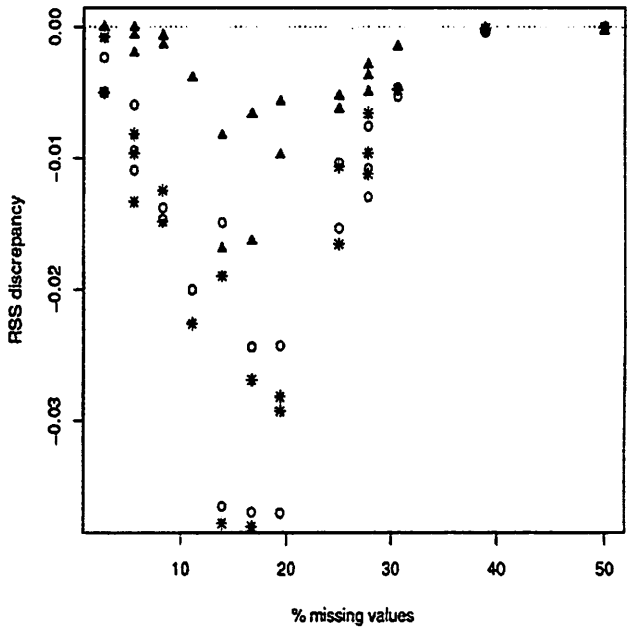
A similar behaviour is observed in R^2 (Figure 5b). With the same explanation given above one can understand why R^2 is maximum in cases of few and too many missing values.

For both ω and R^2 the best behaviour were given by the putative values set to zero. In fact, if one constructs confidence intervals based on Normal distribution for those parameters one can note that the estimates do not differ statistically from source to source, but, since setting a fixed scalar is very much easier than sampling from a distribution, there is no reason to consider the most complicated options.

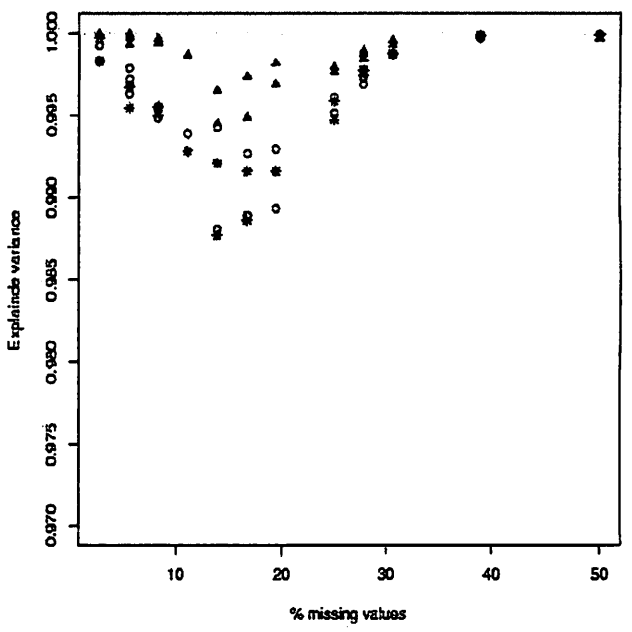
In Figure 6a, one can observe that δ is much smaller when the scalar zero is used. Again, the worst situation happens in a median rate of loss. The estimates of η are more similar along putative sources, but the worst behaviour happens in the same percentage of missing values.

It is worth remembering that the apparent straightforward estimation when there are too many missing values is *not* a good thing, since the estimates are necessarily similar to the missing data. They are just convenient to turn the matrices similar and the match better.

Scaling factor discrepancy ϕ brings us a more fare result in Figure 7. The more missing values, the worst the scaling factor that results from the matching. That is a strong argument against the apparent good situation of having several missing values.

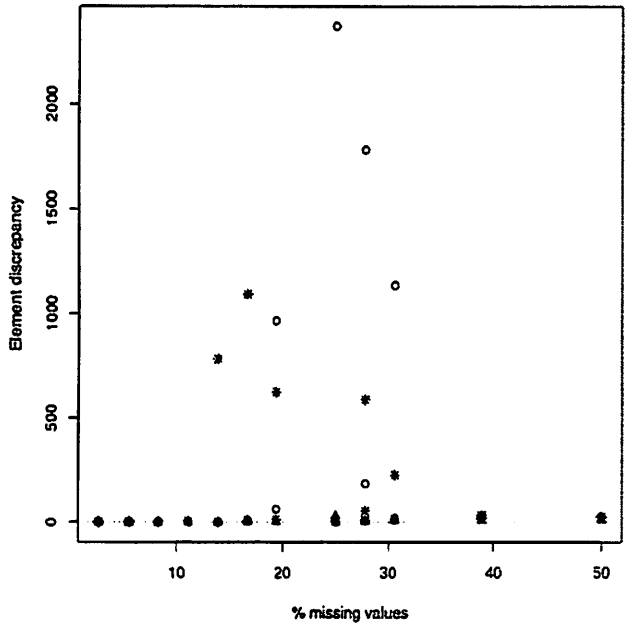


(a)

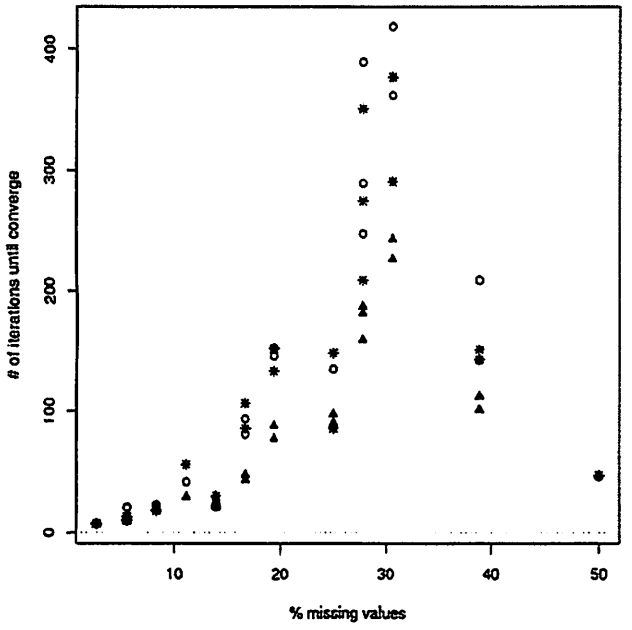


(b)

FIGURE 5: Behaviour of parameter (a) ω and (b) R^2 along an increasing rate of missing values, when their putative values come from Gaussian ($*$) and Uniform (o) distributions and a fixed scalar ($\blacktriangle$).



(a)



(b)

FIGURE 6: Behaviour of parameter (a) δ and (b) η along an increasing rate of missing values, when their putative values come from Gaussian ($*$) and Uniform ($\circ$) distributions and a fixed scalar ($\blacktriangle$).

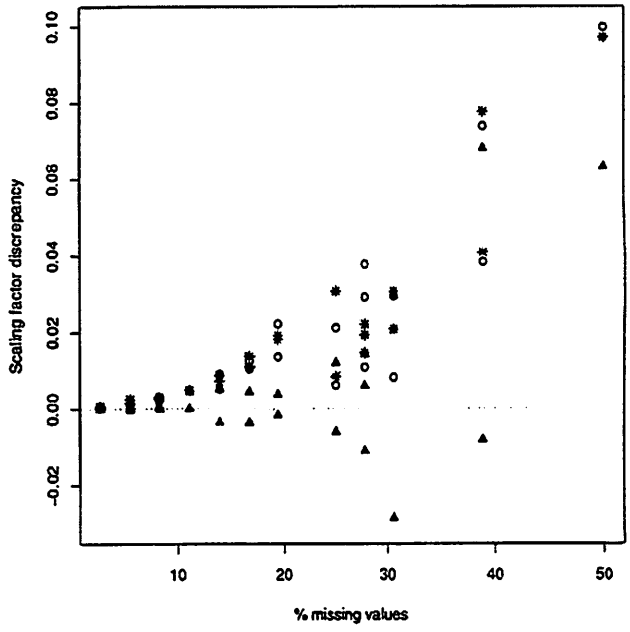


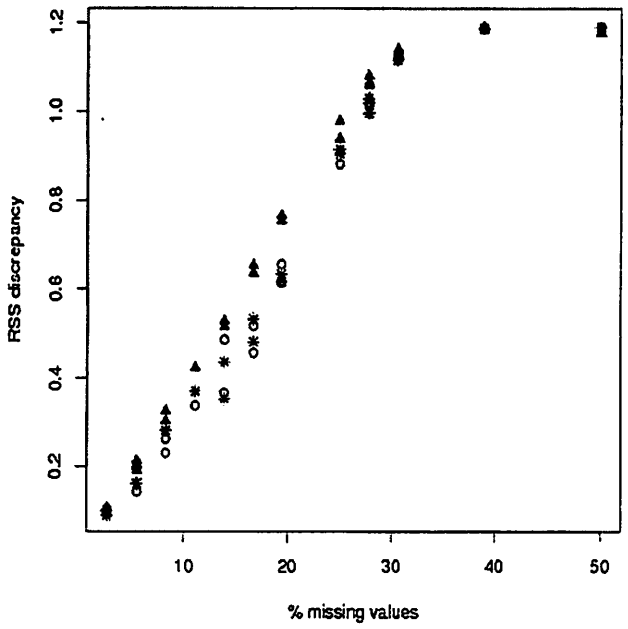
FIGURE 7: Behaviour of parameter ϕ along an increasing rate of missing values, when their putative values come from Gaussian (*) and Uniform (o) distributions and a fixed scalar (▲).

2.1.3.2 Real sized error

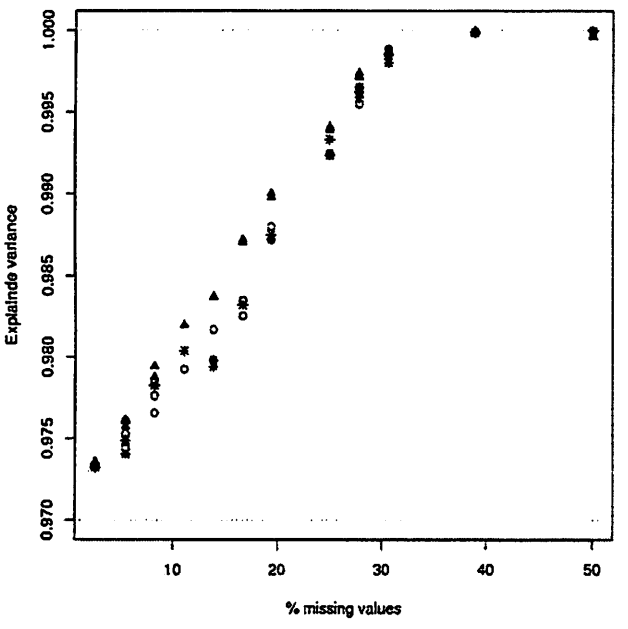
In this situations we look to the estimation in a real sized error context. However, we noticed that this particular data set already gives a good match, i.e., the minimum RSS is already small.

Figure 8a shows an asymptotic behaviour of ω (difference between RSS's) going to 1.2 (original RSS), indicating that RSS_e goes to zero as the number of missing values increases. Here, all three sources of putative values give statistically the same answer as well. Variance explained R^2 goes to 1, in a similar behaviour (Figure 8b).

δ and η show almost the same behaviour of exact match situations because, as explained above, this particular empirical data set already had



(a)



(b)

FIGURE 8: Behaviour of parameter (a) ω and (b) R^2 along an increasing rate of missing values, when their putative values come from Gaussian (*) and Uniform (o) distributions and a fixed scalar (▲).

a great match (Figure 9a and 9b).

In real sized error situations ϕ shows an even worst behaviour when the number of missing cells increase. Figure 10 displays the values spreading along x axis and tends to suggest that the scalar zero seems to provide better estimates. However, the sources of putative values can be considered equal comparing confidence intervals based on Normal distribution.

2.1.4 Conclusions (A)

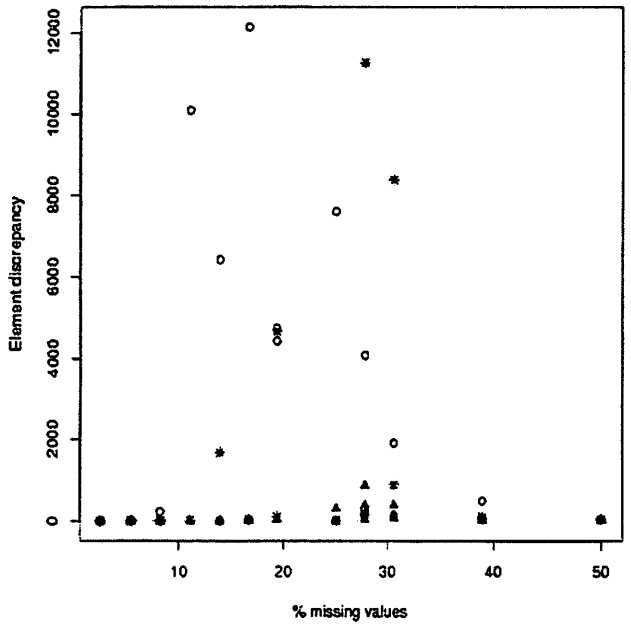
The algorithm described by Gower & Dijksterhuis (2004) to estimate missing data in Orthogonal Procrustes Rotation seems to play a good role estimating them subject to minimize the RSS of the match.

The more missing values the system has, the better is the fit, but it does not guarantee that the estimates will be similar to the lost values.

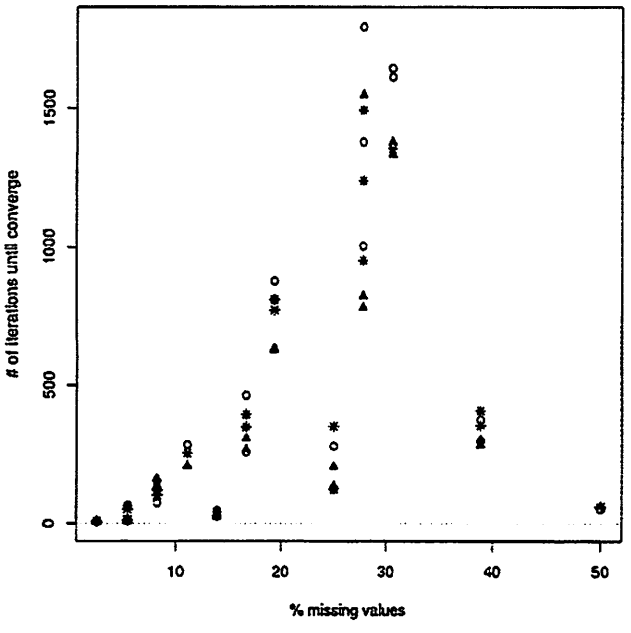
In situations of typically low RSS, median rates of loss (between 14% and 42% of the total information) tend to provide worst estimates.

Scaling factor discrepancy ϕ seem to be a good quality parameter to denude an excess of missing values.

In most of the cases, it does not matter the putative values source. However, there is no logic reason to recommend a procedure that inserts variability in the process. For that reason, centering and starting the estimation process with the scalar zero seems to be the more reasonable attitude.



(a)



(b)

FIGURE 9: Behaviour of parameter (a) δ and (b) η along an increasing rate of missing values, when their putative values come from Gaussian ($*$) and Uniform ($\circ$) distributions and a fixed scalar ($\blacktriangle$).

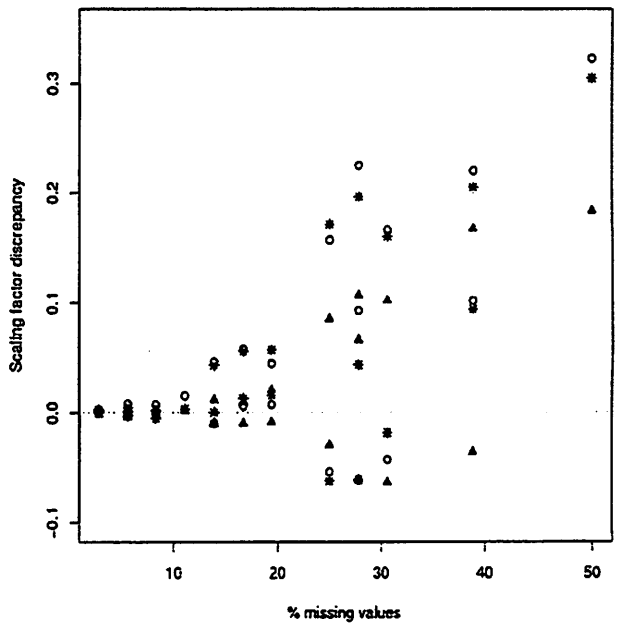


FIGURE 10: Behaviour of parameter ϕ along an increasing rate of missing values, when their putative values come from Gaussian ($*$) and Uniform ($\circ$) distributions and a fixed scalar ($\blacktriangle$).

2.2 Generalised Procrustes Analysis (GPA)

Until 1975, Procrustes problems used to concern only with two data sets, that is, when one matrix is transformed to match one target. In that year, Gower generalised those problems to deal with k sets. In nowadays, many practical applications demand matching more than two configurations. To state a notation, consider the following Procrustes problem[1]

$$f(s_1,\ldots,s_k,u_1,\ldots,u_k,T_1,\ldots,T_k) = \\ = \sum_{i<i'}^{k} ||s_i(X_i - 1u_i')T_i - s_{i'}(X_{i'} - 1u_{i'}')T_{i'}||, \quad (2.3)$$

[1]Although in general transformations the scaling factor (s_i) becomes meaningless, since it is absorbed by the matrix T_i ($i = 1,\ldots,k$), here the most general problem is stated and all estimators derived in order to promptly obtain any particular case latter.

where X_i is a known $n \times p$ data matrix; s_i is a scaling factor; u_i is a translation vector; and T_i is a general transformation matrix ($i < i' = 1, \ldots, k$).

For algebraic facility and to increase interpretability, it is useful to express the same general problem (2.3) as a summation of distances between each transformed configuration and the group average. The latter is the way of expressing the Procrustes problem called by Commandeur (1991) as the *centroid approach*. It is easy to demonstrate that the following basic identity holds (similar to A.2)

$$\sum_{i<i'}^{k} ||s_i(X_i - 1u_i')T_i - s_{i'}(X_{i'} - 1u_{i'}')T_{i'}|| = k \sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G||, \tag{2.4}$$

where G is the group average,

$$G = \frac{1}{k} \sum_{i=1}^{k} s_i(X_i - 1u_i')T_i. \tag{2.5}$$

In turn, (2.4) can be rewritten, using another basic identity (A.3), to express the problem as the summation of distances between each configuration and the *ith*-excluded group average.

$$\begin{aligned} k \sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G|| &= k \left(\frac{k-1}{k}\right)^2 \sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G_i|| \\ &= \frac{(k-1)^2}{k} \sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G_i|| \end{aligned} \tag{2.6}$$

where G_i is the *ith*-excluded group average, that is

$$G_i = \frac{1}{k-1} \sum_{i' \neq i}^{k} s_{i'}(X_{i'} - 1u_{i'}')T_{i'}. \tag{2.7}$$

Expressing the Procrustes problem in terms of the G_i instead of the

commonly used G, one observes two advantages: (i) it avoids the pathological case where the group average is zero; and (ii) the *ith*-excluded group average does not depend on X_{ci}, what helps improving the efficiency of the estimation algorithm.

For notation facility, let X_{ci} denote the *ith* centered configuration, i.e. $X_{ci} = X_i - 1u_i'$. Moreover, from now on lets suppress the constant $\frac{(k-1)^2}{k}$ from (2.6) since it is imaterial for the minimization process. Therefore, (2.6) becomes

$$\sum_{i=1}^{k} ||s_i X_{ci} T_i - G_i||. \tag{2.8}$$

2.2.1 Considerations about non-orthogonal GPA

Considering non-orthogonal transformation matrices T_i for post-multiplying data matrices X_i in Procrustes problems is a quite old and not so developed issue. It might be so because of the good properties of orthogonal matrices Q_i. They are able to simplify a great deal of the algebra and allow prompt practical interpretations. Moreover, they preserve the relative distances between vertices of the configurations, i.e. they promote rigid body rotations and preserve the original *classification* done by the assessor.

However, some effort must be done towards finding and interpreting non-orthogonal transformations able to minimize Procrustes problems (by the way, to lower minima than orthogonal ones), under a suitable constraint. Moreover, orthogonal transformations (Q_i) are a particular case of general transformations (T_i). Thus, since the algebra is well known for T_i it is immediately known for Q_i, in a simplified form. Algorithms as well, can be proposed for allowing the more general case and then computer packages can allow the user to choose the desired transformation according to his/her practical needs.

Particularly, Gower & Dijksterhuis (2004) are concerned in developing a lot of algebra involving general transformations T_i. A tiny part of them is rewritten here.

Following, some algebra is developed and some is reported for orthogonal and non-orthogonal cases, along the estimation of missing values, transformation matrices, scaling factors and translation vectors.

2.2.2 General transformation

(a) Estimating missing values

First of all, lets describe an estimation process of missing values suggested by Gower & Dijksterhuis (2004). It is a modified EM (estimation/minimization) procedure that seeks for those estimates of missing values that minimize a least squares metric. That procedure is described using the notation adopted here rather than the notation used by the authors. The following describes the estimation procedure for the ith X matrix and, of course, the same process must be repeated for every matrix X that have missing values.

Suppose X_i has M values missing in cells $(i_1, j_1), (i_2, j_2), \ldots, (i_M, j_M)$. The missing values may be estimated by a variant of the iterative EM algorithm (Dempster et al., 1977) where "M", rather than representing a step for maximum likelihood estimation, now stands for the least-squares Procrustes minimization, while "E" is a step, described below, that gives expected values for the missing cells.

For fixed $T_{i'}$ $(i' = 1, \ldots, k)$ and $X_{i'}$ $(i' \neq i)$ the terms of $\sum_{i<i'}^{k} ||X_i T_i - X_{i'} T_{i'}||$ that involve X_i requires the minimization of the criterion $||X_i T_i - G_i||$ over the unknown values in given cells $(i_1, j_1), (i_2, j_2), \ldots, (i_M, j_M)$. Recall that G_i represents the ith-excluded group average, which is inde-

pendent of X_i. One assumes that the cells with unknown values contain putative values that one seeks to update by minimizing the criterion. Suppose even that the updates are given in a matrix X_{ui}, which is zero except for the missing cells which contain values denoted by $x_1, x_2, \ldots, x_M$. Thus, one wishes values of X_{ui} that minimize[2]:

$$||(X_i - X_{ui})T_i - G_i|| = ||X_{ui}T_i - (X_iT_i - G_i)||. \tag{2.9}$$

This is a Procrustes problem itself, where now it is X_{ui}, rather than T_i, that is to be estimated. Transposition would put (2.9) into the basic form. The constraint on X_{ui} may be written:

$$X_{ui} = x_1 e_{i_1} e'_{j_1} + x_2 e_{i_2} e'_{j_2} + \ldots + x_M e_{i_M} e'_{j_M} = \sum_{m=1}^{M} x_m e_{i_m} e'_{j_m} \tag{2.10}$$

where, as usual, e_i represents a unit vector, zero except for its ith position. This function is linear in the parameters x_m so, in principle, the minimization of (2.9) is a simple linear least-squares problem; unfortunately, the detailed formulae are somewhat inelegant (A.5). The terms in (2.9) that involve X_{ui} are:

$$tr[(X'_{ui}X_{ui})(T_iT'_i) - 2X'_{ui}(X_iT_i - G_i)T'_i] \tag{2.11}$$

which one writes:

$$tr[(X'_{ui}X_{ui})T - 2X'_{ui}Y] \tag{2.12}$$

where $T = T_iT'_i$ is symmetric and $Y = (X_iT_i - G_i)T'_i$, is the current residual matrix.

The minimization of (2.12) over one missing cell x_r, $r = 1, 2, \ldots, M$,

[2](2.9) is demonstrated in (A.4).

subject to the constraint (2.10) yields[3]:

$$y_{i_r,j_r} = x_1(e'_{i_r}e_{i_1})(t_{j_r,j_1}) + x_2(e'_{i_r}e_{i_2})(t_{j_r,j_2}) + \\ + \ldots + x_M(e'_{i_r}e_{i_M})(t_{j_r,j_M}) \tag{2.13}$$

which may be written:

$$y_r = \sum_{m=1}^{M} x_m(e'_{i_r}e_{i_m})(t_{j_r,j_m}), \quad r = 1, \ldots, M, \tag{2.14}$$

or $y = A_i x$, yielding $x = A_i^{-1}y$. The matrix A_i is symmetric with elements $a_{rm} = (e'_{i_r}e_{i_m})(t_{j_r,j_m})$. This will be zero unless $i_r = i_m$, which would imply that missing values x_r and x_m are in the same row of X_{ui}. When no row contains more than one missing cell, A_i is diagonal and then $x_m = y_{i_m,j_m}/t_{j_m,j_m}$.

In general, A_i contains diagonal blocks, each block being symmetric, with as many rows/columns as there are missing cells in the pertinent row of X_{ui}.

For a better comprehension, consider the following example: suppose X_i (3×2) with cells x_{12}, x_{31} and x_{32} missing (denoted by $\circ$),

$$\begin{pmatrix} x_{11} & \circ \\ x_{21} & x_{22} \\ \circ & \circ \end{pmatrix}. \tag{2.15}$$

So, the indices of the missing cells can be disposed as in Table 4.

Then, matrix A_i is a block diagonal formed by two blocks: the first is 1×1 (because the first row of X_i has just one cell missing in its first row); and the second is 2×2 (because there are 2 missing cells in the second row

[3](2.13) and (2.14) are demonstrated in (A.8).

TABLE 4: The indices of the missing cells.

m	i_r	j_r
1	1	2
2	3	1
3	3	2

of X_i):

$$\boldsymbol{A}_i = \begin{pmatrix} t_{22} & 0 & 0 \\ 0 & t_{11} & t_{12} \\ 0 & t_{12} & t_{22} \end{pmatrix}. \tag{2.16}$$

As a general formule, one may consider filling the diagonal of $\boldsymbol{A}_i$ first: $diag(\boldsymbol{A}_i) = \{a_{mm}\} = t_{j_r j_r}$; and then filling up the non-diagonal elements of the blocks crossed indexes of t.

At this point it is important talking about translation. Although one could recentre after adjusting for the missing cells, it can be done better. One assumes that the current settings of X_i and hence G_i are centered. To preserve centring we replace X_{ui} by its centred form

$$(I - N)X_{ui},$$

where I is a $n \times n$ identity matrix and $N = \frac{1}{n}\mathbf{1}\mathbf{1}'$; where $\mathbf{1}$ is a vector of ones of length n.

The effect of this is to replace the previous definition of a_{rm} by $a_{rm} = (e'_{i_r}(I - N)e_{i_m})(t_{j_r,j_m})$. Thus, when $i_r = i_m$, $a_{rm} = (1 - 1/n)(t_{j_r,j_m})$ else $a_{rm} = -1/n(t_{j_r,j_m})$, so that $\boldsymbol{A}_i$ is now replaced by

$$\boldsymbol{A}_i^* = \boldsymbol{A}_i - 1/n\boldsymbol{T}^*,$$

where $\boldsymbol{T}^*$ is made up of those elements of $\boldsymbol{T}$ corresponding to the columns

of the missing cells (e.g. the third column of Table 4):

$$\boldsymbol{T}^* = \{t^*_{ij}\} = t_{j_r j_m},$$

and when $\boldsymbol{T}$ is symmetric, $t_{j_r j_m} = t_{j_m j_r}$.

In the example, assume $\boldsymbol{T}^*$ is a tabular form to turn explicit the indices r and m:

$$\begin{array}{c} \text{m/r} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} t_{22} & t_{12} & t_{22} \\ t_{21} & t_{11} & t_{12} \\ t_{22} & t_{12} & t_{22} \end{pmatrix} \end{array}.$$

This minor change gives $\boldsymbol{x} = \boldsymbol{A}_i^{*-1}\boldsymbol{y}$, so defining X_{ui}, and then $(I_N)X_{ui}$ gives the required correction matrix. This is all that is necessary for handling translations. Recentring X_{ui} derived from $\boldsymbol{x} = \boldsymbol{A}_i^{-1}\boldsymbol{y}$ and using $(I-N)X_{ui}$ derived from $\boldsymbol{x} = \boldsymbol{A}_i^{*-1}\boldsymbol{y}$ will give different updates to the missing values, though both must converge to the same final estimates. A near exception to this rule occurs for orthogonal transformations when the two approaches hardly differ (Gower & Dijksterhuis, 2004, Section 9.2.1).

Now, an alternative procedure is presented for finding estimates of the missing values. Such procedure intends to be an easier and computationally lighter alternative to the one described above. However it is based on the same principle, that is, finding those estimates that minimize a residual sum of squares. The main difference is the way the problem of Procrustes is stated.

Suppose the following Procrustes problem, in a least squares metric:

$$\sum_{i=1}^{k} ||s_i(X_{ci} + W_i * X_{ui})T_i - G_i||, \tag{2.17}$$

where X_{ci} has putative values in unknown cells (say, zeros); W_i is an indicator matrix that contains ones in unknown cells addresses and zeros

otherwise; X_{ui} is an update matrix for X_i; and $*$ stands for an element-wise product. Now, for notation matters, let $X_{ui}^r = W_i * X_{ui}$, where the upper index r stands for *restricted*.

At this point a remark is necessary: when considering general transformations T_i it is not necessary to consider scaling factors s_i. The effect of the scaling factors is absorbed by the T_i's. However, one intends to state the problem on its more general form, for which orthogonal rotation Q_i is a special case. In turn, when orthogonal rotation Q_i is allowed, the use of s_i does make sense.

If one consider all parameters known but the update matrix X_{ui}^r, for each $i = 1, \ldots, k$, one has to minimize

$$||s_i(X_{ci} + X_{ui}^r)T_i - G_i||,$$

what is the same as minimizing[4]:

$$tr[(X_{ui}^{r\prime}X_{ui}^r)T - 2X_{ui}^{r\prime}Y], \tag{2.18}$$

where $T = T_iT_i'$ and $Y = (s_i^{-1}G_i - X_{ci}T_i)T_i'$.

Minimizing (2.18) over X_{ui}^r leads to the least squares estimator[5]:

$$\hat{X}_{ui}^r = YT^{-1} \quad (i = 1, \ldots, k). \tag{2.19}$$

However X_{ui}^r is not necessarily restricted, that is, nothing guarantees it has only values in the missing addresses and zeros otherwise, but as the estimation process is already iterative, it possibly tends to converge to a restrict matrix as is desired. Therefore, for obtaining estimates specifically

[4]Proof is given in (A.10).
[5]Demonstration is given in (A.9).

for the missing values, one can restrict the answer:

$$\hat{X}^r_{ui} = YT^{-1}$$
$$\widehat{W_i * X_{ui}} = YT^{-1}$$
$$\widehat{W_i * X_{ui}} = W_i * YT^{-1}.$$

Accordingly to the prior procedure described here, the translation correction can be insert as well. Derivation of the estimator is omitted, but it is similar:

$$\hat{X}^r_{ui} = W_i * [(I - N)^{-1} YT^{-1}]. \tag{2.20}$$

where $(I-N)$ is as defined above. However is important noting that $(I-N)$ can require generalised inverse due to possible singularity.

Of course such procedure is not so elegant as Gower & Dijksterhuis' but it is a (at first sight) easier and computationally lighter alternative that might lead to the same answer when it converges.

(b) **Estimating transformation matrices**

Deriving the estimator of transformation matrices T_i depends on the constraint inserted into the minimization process.

With no restriction, i.e. deriving the estimator without any kind of constraint, estimating T_i is solving a multivariate multiple regression problem. In problem (2.8), considering s_i, X_{ci} and G_i as fixed[6], it is easy to solve the multivariate multiple regression problem[7]:

$$\hat{T}_i = \frac{1}{s_i}(X'_{ci}X_{ci})^{-1}X'_{ci}G_i. \tag{2.21}$$

However, some kind of constraint is suitable for avoiding trivial so-

[6]Remember that X_{ci} is not usually fixed because it is function of Xui; and G_i is function of $T_{i'}$, for $i' \neq i$, considered fixed by the algorithm when estimating T_i.

[7]Demonstration in (A.11).

lutions as $T_1 = \ldots = T_k = \emptyset$, where $\emptyset$ is a matrix $(p \times p)$ of zeros.

Gower & Dijksterhuis (2004) derive an estimator of T_i minimizing a Procrustes problem under a general quadratic restriction. That derivation follows.

Consider the Procrustes problem rewritten in a hypermatrix form:

$$S = \sum_{i<i'}^{k} ||X_{ci}T_i - X_{ci'}T_{i'}|| = tr(kV'\Delta V - V'ZV), \tag{2.22}$$

where $Z = \{X'_{ci}X_{ci'}\}$, $\Delta = diag(X'_{c1}X_{c1}, X'_{c2}X_{c2}, \ldots, X'_{ck}X_{ck})$ and $V' = T'_1, T'_2, \ldots, T'_k$. Now, consider the constraint in a general form:

$$diag(\alpha V'\Delta V + \beta V'ZV) = I, \tag{2.23}$$

which can turn into several constraints. For instance, setting $\alpha = 1/k$ and $\beta = 0$ generates the constraint $diag(V'\Delta V) = kI$; and setting $\alpha = 0$ and $\beta = 1/k^2$ gives $diag(V'ZV) = k^2I = k^2G'G$. Differentiation of (2.22) with Lagrange multipliers $\gamma_1, \gamma_2, \ldots, \gamma_R$, and writing $\Gamma = diag(\gamma_1, \gamma_2, \ldots, \gamma_R)$, gives:

$$k\Delta V - ZV = (\alpha\Delta V + \beta ZV)\Gamma. \tag{2.24}$$

Pre-multiplying the above by V' and using (2.22) and (2.23), immediately shows that $S = tr(\Gamma)$, so the criterion is minimized by selecting the R smallest values of γ_r that satisfy (2.24). One may rewrite (2.24):

$$ZV = \Delta V\Phi, \tag{2.25}$$

where $\Phi = (kI - \alpha\Gamma)(I + \beta\Gamma)^{-1}$. As usual, the matrices $V'ZV$ and $V'\Delta V$ are diagonal. Thus one has orthogonality even though (2.23) only constrains the diagonal elements of these matrices. Thus (2.23) becomes:

$$\alpha V'\Delta V + \beta V'ZV = I,$$

which in combination with (2.25) gives the scaling of the columns of T_i:

$$(V'\Delta V)(\alpha I + \beta\Phi) = (V'ZV)(\alpha\Phi^{-1} + \beta I) = I.$$

However a two-sided eigenvalue problem requires at least one of the given matrices in (2.25), (Z or Δ) to be positive definite, and since Z and Δ are formed by $X'_{ui}X_{ui}$ they are both typically positive semi definite. For that reason, the determination of T_i, that is V, is not straightforward. Therefore, the further works are needed for developing Procrustes problems with general transformations T_i.

(c) Estimating scaling factors

It is worth noting that, in the general transformation case, the scaling factors become meaningless since they are absorbed by the matrices T_i ($i = 1, \ldots, k$). However, if, for some reason, they need to be estimated in a general case, it would be possible. Moreover, its estimator will be derived here for didactic reasons, namely, the estimation process of scaling factors in the orthogonal case will be showed to be a particular case of this one.

First of all, lets rewrite (2.3) allowing for missing values as in (2.17) but with X_i already centered as in (2.8).

$$f(s_1, \ldots, s_k) = \sum_{i<i'}^{k} ||s_i(X_{ci} + W_i * X_{ui})T_i - s_{i'}(X_{ci'} + W_{i'} * X_{ui'})T_{i'}|| \quad (2.26)$$

To simplify the notation, one can let $X_{ci} + W_i * X_{ui} = X^*_{ci}$, i.e. configurations centered and enabled for missing data. That yields

$$f(s_1, \ldots, s_k) = \sum_{i<i'}^{k} ||s_i X^*_{ci} T_i - s'_i X^*_{ci'} T_{i'}||. \quad (2.27)$$

Moreover, at the *scaling factor* step, one must consider the other variables to be fixed, namely X^*_{ci} and T_i. Therefore, (2.26) and (2.27) turn

out into functions of s_i.

Rewriting equation (2.27) in a matrix form[8], gives

$$f(s_1, \ldots, s_k) = k(s' diag(\mathbf{S})s) - s'\mathbf{S}s \tag{2.28}$$

where s is a vector containing all s_i and $\mathbf{S}$ is a symmetric matrix with elements $\mathbf{S}_{ii'} = \{tr(T_i' X_{ci}'^{*} X_{ci'}^{*} T_i')\}$, $i = 1, \ldots, k$.

Note that minimizing the loss function (2.28) is maximizing

$$h(s) = s'\mathbf{S}s. \tag{2.29}$$

In order to avoid the trivial solution $s_1 = s_2 = \ldots = s_k = 0$, one must consider a suitable restriction. Suppose one imposes that the configurations must have the same total size before and after the scaling transformation, i.e.,

$$\sum_{i=1}^{k} ||X_{ci}^{*} T_i|| = \sum_{i=1}^{k} ||s_i X_{ci}^{*} T_i||. \tag{2.30}$$

Gower & Dijksterhuis (2004) emphasize that the scaled configuration is important to be set to a fixed value, what is respected assuming X_{ci}^{*} and T_i to be fixed at this estimation step.

Expressing (2.30) in matrix notation using the prior definition of s and $\mathbf{S}$ and assuming $diag(\mathbf{S})$ to be a diagonal matrix corresponding to the main diagonal of $\mathbf{S}$, gives

$$tr(diag(\mathbf{S})) = s' diag(\mathbf{S})s. \tag{2.31}$$

Since $diag(\mathbf{S})^{\frac{1}{2}}\, diag(\mathbf{S})^{-\frac{1}{2}} = I_k$, where I_k is a k dimensional identity, (2.29) can be rewritten as

$$h(s) = s'\, diag(\mathbf{S})^{\frac{1}{2}} diag(\mathbf{S})^{-\frac{1}{2}}\, \mathbf{S}\, diag(\mathbf{S})^{-\frac{1}{2}} diag(\mathbf{S})^{\frac{1}{2}}\, s.$$

[8]Proof is given in (A.12).

Now, let

$$diag(\mathbf{S})^{-\frac{1}{2}}\ \mathbf{S}\ diag(\mathbf{S})^{-\frac{1}{2}} = P\Phi P'$$

be an eigenvalue-eigenvector decomposition, where P is a $(k \times k)$ orthogonal matrix containing the eigenvectors and Φ is a $(k \times k)$ diagonal matrix containing the eigenvalues of $diag(\mathbf{S})^{-\frac{1}{2}}\ \mathbf{S}\ diag(\mathbf{S})^{-\frac{1}{2}}$. Then, the expression (2.29), to be maximized, can be expressed as

$$h(s) = s'\, diag(\mathbf{S})^{\frac{1}{2}}\ P\Phi P'\ diag(\mathbf{S})^{\frac{1}{2}}\ s.$$

Now, if one let $a = P' diag(\mathbf{S})^{\frac{1}{2}} s$, then

$$h(s) = a'\Phi\, a.$$

Since

$$h(s) = a'\Phi\, a \leq \phi_1 a'a,$$

where ϕ_1 is the largest eigenvalue and

$$\begin{aligned} a'a &= s'\, diag(\mathbf{S})^{-\frac{1}{2}}\ PP'\ diag(\mathbf{S})^{\frac{1}{2}}\ s \\ &= s'\, diag(\mathbf{S}) s \\ &= tr(diag\,(\mathbf{S})), \end{aligned}$$

then,

$$h(s) = a'\Phi\, a \leq \phi_1 tr(diag\,(\mathbf{S})). \tag{2.32}$$

Since one wants to maximize $a'\Phi\, a$, one seeks for s such that $h(s) = \phi_1 tr(diag\,(\mathbf{S}))$.

Lets denote p_1 the eigenvector associated with the largest eigenvalue

ϕ_1, then rewriting (2.32) in the form of (2.29), gives

$$\begin{aligned} h(s) &= \phi_1 tr(diag\,(\mathbf{S})) \\ &= tr(diag\,(\mathbf{S}))\; p_1' P \Phi P' p_1 \\ &= \left(tr(diag\,(\mathbf{S}))^{\frac{1}{2}} diag(\mathbf{S})^{-\frac{1}{2}} p_1\right)' \mathbf{S} \left(tr(diag\,(\mathbf{S}))^{\frac{1}{2}} diag(\mathbf{S})^{-\frac{1}{2}} p_1\right) \\ &= h\Big(tr(diag\,(\mathbf{S}))^{\frac{1}{2}} diag(\mathbf{S})^{-\frac{1}{2}} p_1\Big). \end{aligned}$$

So, the estimator of s should be

$$\hat{s} = tr(diag\,(\mathbf{S}))^{\frac{1}{2}}\; diag(\mathbf{S})^{-\frac{1}{2}} p_1, \tag{2.33}$$

since it minimizes the loss function (2.28) and satisfies the constraint (2.31),

$$\begin{aligned} s' diag(\mathbf{S}) s &= \left(tr(diag\mathbf{S})^{\frac{1}{2}} diag(\mathbf{S})^{-\frac{1}{2}} p_1\right)' diag(\mathbf{S}) \left(tr(diag\mathbf{S})^{\frac{1}{2}} diag(\mathbf{S})^{-\frac{1}{2}} p_1\right) \\ &= tr(diag\,(\mathbf{S}))\, p_1' diag(\mathbf{S})^{-\frac{1}{2}} diag(\mathbf{S})\, diag(\mathbf{S})^{\frac{1}{2}} p_1 \\ &= tr(diag\,(\mathbf{S}))\, p_1' p_1 \\ &= tr(diag\,(\mathbf{S})). \end{aligned}$$

(d) **Estimating translation vectors**

An easy way of dealing with translation vectors u_i $(i = 1, \ldots, k)$ is performing the estimation/minimization algorithm concerning only with X_{ci}, i.e., centering the X matrices at each step. After convergence, the translation vectors u_i can be estimated by the average of the columns of $W_i * X_i - X_{ci}$ and then, the estimator of the translation matrix is

$$1\hat{u}_i'. \tag{2.34}$$

2.2.3 Orthogonal transformation

When the transformation matrices T_i $(i = 1, \ldots, k)$ are restricted to be orthogonal (say, Q_i), the expressions tend to become simpler due to the general properties of the orthogonal matrices.

(a) Estimating missing values

In the procedure described by Gower & Dijksterhuis (2004), making $T_i = Q_i$ turn things simpler. When no correction for translation is done in the missing cells, the estimator $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$ simplifies for $\boldsymbol{x} = \boldsymbol{y}$. However, when the translation correction is done, the estimator of missing cells remains $\boldsymbol{x} = \boldsymbol{A}^{*-1}\boldsymbol{y}$, because though $T = A = I_p$, T^* and consequently A^* are not identities necessarily.

In the procedure proposed here, the estimator expression for X_{ui}^r do become simpler as well because T reduces for identity ($T = Q_i Q_i' = I_p$) and Y reduces for $G_i Q_i' - s_i X_{ci}$.

Therefore, the new estimator for X_{ui}^r $(i = 1, \ldots, k)$ is

$$\hat{X}_{ui}^r = Y = G_i Q_i' - s_i X_{ci} \tag{2.35}$$

or

$$\hat{X}_{ui}^r = W_i * [(I - N)^{-1}(G_i Q_i' - s_i X_{ci})], \tag{2.36}$$

when translation correction is allowed.

(b) Estimating rotation matrices

When one wants T_i to be orthogonal, producing a rigid body rotation in X_{ci}, the solution is well known.

Consider (2.17) setting $T_i = Q_i$, then

$$\sum_{i=1}^{k} ||s_i X_{ci}^* Q_i - G_i|| = tr\left(s_i^2 Q_i' X_{ci}^{*'} X_{ci}^* Q_i - 2 s_i Q_i' X_{ci}^{*'} G_i - G_i' G_i\right)$$
$$= tr\left(s_i^2 X_{ci}^{*'} X_{ci}^* - 2 s_i G_i' X_{ci}^* Q_i - G_i' G_i\right). \qquad (2.37)$$

Since G_i does not depend on Q_i, and remembering that s_i and X_{ci}^* are considered to be fixed at this step, minimizing (2.37) is maximizing $tr\,(G_i' X_{ci}^* Q_i)$. Remember that one wants Q_i to lead the function above to its minimum and to be orthogonal. Then, for $i = 1, \ldots, k$, expressing $G_i' X_{ci}^*$ in terms of its singular value decomposition $G_i' X_{ci}^* = U_i \Sigma_i V_i'$, where U and V are orthogonal matrices and Σ is diagonal, gives

$$tr\left(G_i' X_{ci}^* Q_i\right) = tr\left(U_i \Sigma_i V_i' Q_i\right)$$
$$tr\left(G_i' X_{ci}^* Q_i\right) = tr\left(\Sigma_i V_i' Q_i U_i\right)$$
$$= tr\left(\Sigma_i H_i\right),$$

where $H_i = V_i' Q_i U_i$, being the product of orthogonal matrices, is itself ortogonal. Considering the elements of Σ_i and H_i, respectively, $\Sigma_i = \sigma_j$ and $H_i = h_{lj}$, one has that

$$tr(\Sigma_i H_i) = \sum_{j=1}^{p} h_{jj} \sigma_j,$$

which, because the singular values σ_j are non-negative, is maximum when $h_{jj} = 1$ for $j = 1, \ldots, p$, the maximal value attained by the elements of an ortogonal matrix. Thus at the maximum $H_i = I$ giving, $I = V_i' Q_i U_i$ and $Q_i = V_i U_i'$.

Therefore, the estimator $\hat{Q}_i$ that minimizes the Procrustes problem and guarantees Q_i to be orthogonal is:

$$\hat{Q}_i = VU' \qquad i = 1, \ldots, k. \qquad (2.38)$$

(c) **Estimating scaling factors**

Scaling factors are obtained from the same standard eigenvalue problem (2.33), but its argument matrix is modified,

$$\hat{s} = tr(diag\,(\boldsymbol{A}))^{\frac{1}{2}}\, diag(\boldsymbol{A})^{-\frac{1}{2}} p_1, \tag{2.39}$$

where $\boldsymbol{A}$ is slightly different from $\boldsymbol{S}$, i.e., it is function only of the centered configurations $\boldsymbol{A}_{ii'} = \{tr(X_{ci}'^{*} X_{ci'}^{*})\}$.

(d) **Estimating translation vectors**

Following the same principle of centering X_i in each step of the algorithm and estimating u_i at the end, the estimator $\hat{u}_i$ remains the same (2.34) if one is performing an orthogonal transformation in the configurations.

2.2.4 Algorithm

Empirical tests have shown that the order of the steps in the modified EM algorithm is responsible for a slower or faster converge. Moreover, estimating all parameters within each i or, for estimating within each parameter all $i = 1, \ldots, k$, also seems to make a great effect. However, further works must be done to comprove and explain such phenomenon. Following algorithm is already in the suggested order:

1. Insert putative values in the missing cells of matrices X_{ui} and compute X_i^*, $i = 1, \ldots, k$.
2. Centre matrices X_i^* obtaining X_{ci}^*, $i = 1, \ldots, k$.
3. Normalize X_{ci}^*, $i = 1, \ldots, k$.
4. Set initial matrices T_i to identities and s to unit vectors.

5. Compute the *ith*-excluded configurations G_i, $i = 1, \ldots, k$.
6. Compute the initial residual sum of squares (Procrustes loss).
7. Estimate updated transformation matrices T_i or Q_i, $i = 1, \ldots, k$.
8. Estimate the missing values and update X^*_{ci}, $i = 1, \ldots, k$.
9. Recentre X^*_{ci}, $i = 1, \ldots, k$.
10. Update G_i, $i = 1, \ldots, k$.
11. Estimate new scaling factors s_i, $i = 1, \ldots, k$.
12. Recompute the Procrustes loss and compare with its previous value. If their different suits a desired converge tolerance, than the process is finished. Otherwise, return to step 7.

2.2.5 Conclusions (B)

Algorithms for estimating missing data in Procrustes problems are still a unfinished field although well discussed in the recent literature. Determining a suitable (optimum) algorithm (order of steps) is a field to be explored.

On the other hand other constraints should be investigated in order to provide estimates closer to *reality*, i.e., closer to the cells that are missing.

For non-ortogonal transformations, two-sided eigenvalue problems with both matrices positive semi-definite must be solved in order to guarantee such the convergence of the estimation process to plausible values. Moreover, practical interpretation of such transformations are still waited for a closer relation with sensory experiments.

3 STATISTICAL INFERENCE AND GPA

3.1 Introduction

Generalised Procrustes analysis (GPA) is considered a descriptive multivariate statistical method (Ferreira, 2004). A graphical method that allows relative statements concerning a set of objects in its answer-plan. Therefore, one of the results of GPA is an answer plan containing the two first dimensions of maximal variation. When G is referred to that plan, its n points (n objects) represent the consensus configuration. In sensory analysis it represents the 'mean' scores, along the assessors, to all analyzed objects. However, each assessor configuration can be referred as well, for instance, to check their distance from the group average. That plan is generated by two axes that represent linear combinations of the attributes (dimensions) used to describe the objets.[1]

However, GPA presents some limitations. Though any assumption is necessary to be made about the multivariate distribution of the scores given to objects' attributes, the current methodology does not allow the statistical inference about a population of possible consumers of a product, for instance. In other words, there is no statistical inference in Procrustes analysis. It does not allow neither an interval estimation of the position of an object in a multivariate sensorial space nor hypothesis tests with a level

[1]It is worth noting that the number of principal axes investigated may vary from 1 to p, though it is most common to restrict the inference to the first two axes (the plan).

of significance associated. The only possible conclusions are relative ones.

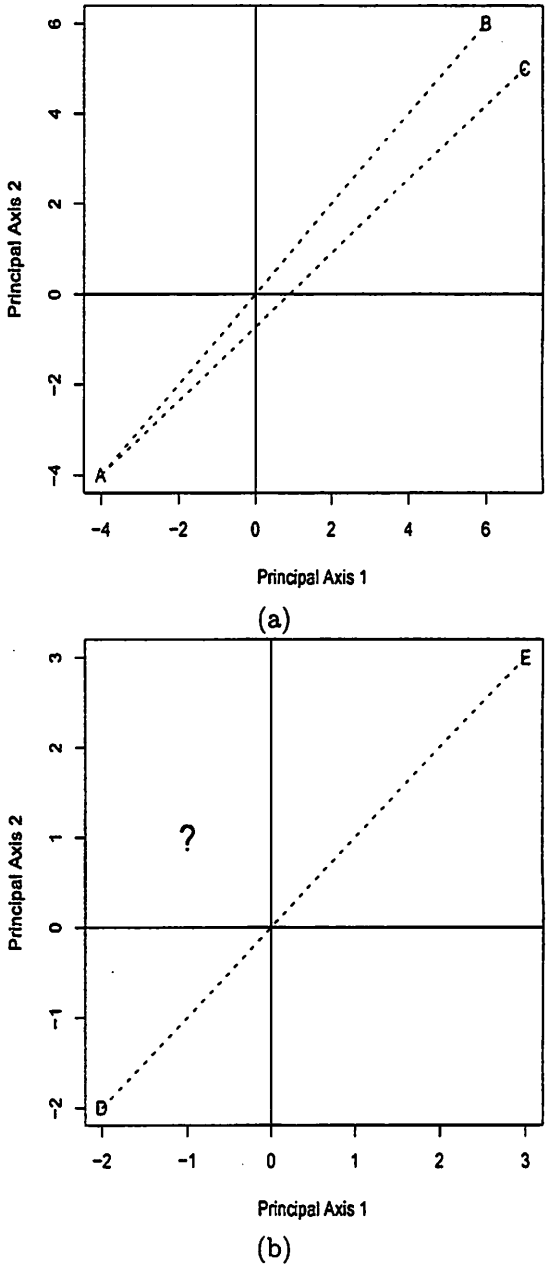


FIGURE 11: (a) Relative euclidian distances between objects A, B e C. (b) Euclidian distances between objects D and E.

For instance, let Figure 11a represent the group average configuration of a panel about three objects, say, A, B and C. Therefore, the relative euclidian distance between the points can be understood as the magnitude of the differences perceived to exist among the objects, according to the group average configuration. Here, one can infer that the judges agree that objects B and C are more similar than objects A and C, for instance (11a). That statement is itself informative, but what about the question: "are A and C distinct (statistically different)?", or "are the judges capable of distinguishing between objects B and C?". Besides, what about performing an experiment to describe (compare) only two objects (Figure 11b)? What can one infer in this situation? What does the magnitude of the distance mean if there is no minimum significant distance known? Nothing at all.

It never was the role of GPA to test the equality of objects' means. It is clear from the quotations of use of translations that the first thing GPA does is *throwing away* the means. This work proposes estimating and testing the relative euclidian distance between objects associating uncertainty to the configurations. But, since it does not make much sense to construct confidence regions for configurations, it is done for their vertices, the objects' locations.

Inference in Generalised Procrustes Analysis is a challenge to be faced. It was not so until today probably due to two main reasons: (i) the multivariate distribution of the scores given by the judges is unknown (and would be very strong the assumption of a multivariate normal distribution); (ii) GPA is an iterative algorithm, therefore, even if the original distribution of the scores was known before GPA, it would be very difficult to know it after its transformations. Besides, it is worth noting that reason (i) only makes sense to occur if all the scores matrices are composed by the *same* p variables. However, one of the main advantages using GPA is the ability of dealing with data coming from Free Choice Profiling (FCP), for instance,

i.e., each matrix can contain its own set of variables. Remember that GPA is methodology that compares configurations, and those configurations are formed by *dimensions* rather than *variables*.

Therefore, intensive computational techniques are suitable for handling that situation. Here, mainly *bootstrap* techniques are used to estimate the empirical distribution of the desired parameters' estimators (whichever is the original scores distribution). Thus, such a methodology can enrich GPA's potentialities allowing estimation and decision theories to be stated.

3.2 Sampling in Sensory Sciences

Some discussion was already done in section 1.4, but here one intend to stress some considerations and address some recommendations.

Generally, statistical tests require a sample to be selected randomly from the population it represents. This may be a complex matter but there are additional points of design which are worth mention here.

When sampling, one must select the sample in an unbiased manner. In practice, a so-called random sample is rarely completely random. Subjects are usually volunteers and volunteers are not necessarily representative of the population from which they were drawn. Also, an experimenter's method of choosing subjects is hardly likely to be random; he or she will tend to select those who are friendly and enthusiastic, as well as those who are conveniently near the testing location. The importance of such difficulties is more a matter of philosophy than mathematics; it is certainly hard to gauge. Unfortunately, such questions are rarely considered; experimenters seem to hope that they are unimportant. For instance, one of the common errors in sensory work with foods is to sample from the wrong population (O'Mahony, 1985).

Further, members of the sample may also tell lies. This can occur when question of a personal nature (money, sex, etc) are asked. The tendency to please the experimenter and give the 'right' answer is strong among human subjects.

Control groups are often used to distinguish what is been estimated (treatments, blocks, variables, etc) for any from any other effect that might be happening. The response of non-treated experimental units is the expression of the so called *placebo effect*. By comparing treated units and control a measure can be obtained, independent of any suggestion. Control groups are often used in behavioral research and could sometimes be useful in the sensory analysis of foods.

Generalised Procrustes analysis deals with data coming from experts, trained, semi trained, untrained, consumers, etc. Furthermore, GPA is sufficiently dynamic to treat scores produced by different scales, containing different sorts of bias; each assessor can use how many and whatever variables they want to measure the same object. By definition, GPA is not concerned in comparing means but configurations. GPA adjusts configurations and identifies mains agreements and disagreements among judges. For those reasons, GPA is more important in a Free Choice Profile (FCP) context (Williams & Langron, 1984). For any other, say, 'balanced' situation, 'classical' multivariate statistical tools, such as Principal Components Analysis (PCA), Factor Analysis (FA), Cluster Analysis (CA), Partial Least Squares Regression (PLSR), Multivariate Analysis of Variance (MANOVA), etc, tend to be more adequate. Therefore, due to the huge freedom to vary of FCP, semi trained and untrained assessors can be drawn from correspondent populations, leading that to be a random effect.

3.3 Generalised Procrustes Analysis

Lets X_i $(i = 1, \ldots, k)$ be k data matrices, representing k assessors, with dimensionality $n \times p_i$, where n is the number of evaluated objects $(l = 1, \ldots, n)$ and p_i is the number of dimensions or attributes used by assessor i $(j = 1, \ldots, p_i)$. In each of those matrices, the lth row represents the coordinates of a point $S_l^{(i)}$ in the space $\mathbb{R}^{p_i}$. In a sensory analysis context, it represents the lth object scored by assessor i. Therefore, each judge has its own space of attributes where each object is scored; the estimate of the object's parameter location is represented by $S_l^{(i)}$. Then, each judge is a set of points in a vectorial space (Ferreira, 2004).

In a more general case, assessors can evaluate the objects according their particular set of attributes, even when vocabularies differ from assessor to assessor. Therefore, judge i can use p_i attributes (number of columns of X_i). However, for enabling Generalised Procrustes analysis to operate well, is better to have the scores matrices with columns forming groups of similar attributes and is advisable to have the same number of columns. Possible solutions are: (i) estimating the missing columns; (ii) padding them with zeros (in centered configurations) or; (iii) modeling the problem to allow for different matrices dimensions. Lets consider the number of attributes as a constant p derived from $p = Max(p_i)$ [2].

Each matrix X_i is geometrically represented by n points (n objects). When a line joints the points referring to one object, through all assessors, the appearing polygon or configuration is called *object configuration.* Figure 12 illustrates an object configuration highlighting the polygons of four objects $(O_1, \ldots, O_4)$, evaluated by three judges.

On the other hand, when joining points referring to a same assessor,

[2] If, for instance $p = Min(p_i)$ was used, one would reject information, though turning possible matrix calculations.

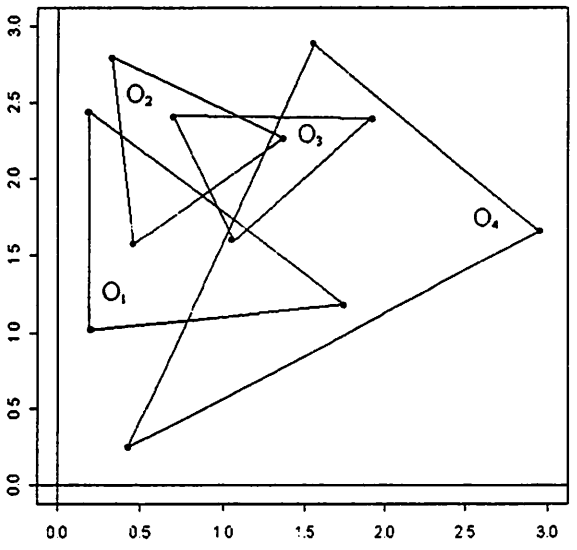


FIGURE 12: Configurations for four objects $(O_1, \ldots, O_4)$.

along all objects, that polygon is called *assessor configuration.* Figure 13 illustrates the assessor configuration highlighting the same points of Figure 12, but with different polygons drawn, i.e., polygons of three assessors (X_1, X_2, X_3), evaluating four objects.

3.4 Describing an object

Describing an object is when a set of assessors scores a single object (continuous scores, in a more general case) for the intensity of its important characteristics or attributes[3]. Table 5 shows a series of scores $\{x_{ij}\}$ given by k judges $(i = 1, \ldots, k)$ for a given object under p attributes $(j = 1, \ldots, p)$.

Here, the objective is just characterize an object, i.e., settle it in a space formed by the axes that represent the scored attributes. Lets call

[3]Important attributes are those characteristics inherent to the object that are essential for its description. For instance, one would hardly describe a concentrated orange juice without mentioning its color, sourness and sweetness.

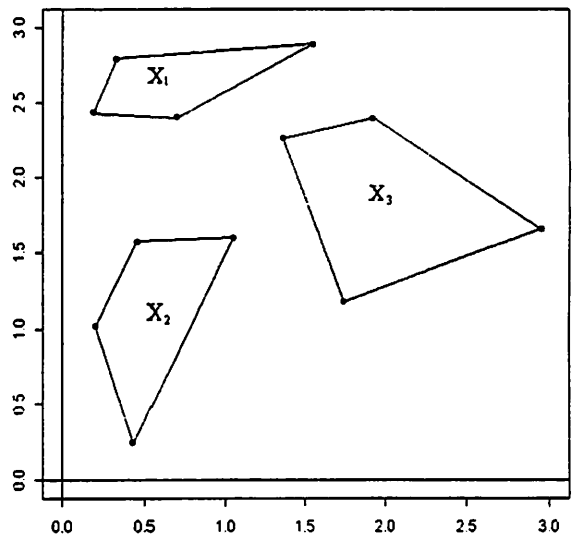


FIGURE 13: Configurations of three assessors (X_1, X_2 e X_3).

TABLE 5: Tabular representation of scores from k assessors to one object according to p attributes (x_{ij}).

	Attributes			
Assessors	1	2	...	p
1	x_{11}	x_{12}	...	x_{1p}
2	x_{21}	x_{22}	...	x_{2p}
⋮	⋮	⋮	⋱	⋮
k	x_{k1}	x_{k2}	...	x_{kp}

that space *sensorial space* (Ω_Ψ). That space is a particular vectorial space (with all required properties) where the dimensions are sensorial attributes.

For instance, one object was evaluated by five judges, who scored a pair of main attributes using the same scale (say, scoring from 0 to 10). Then, consider that each attribute is an axe and the scores are points coordinates in a sensorial space generated by those attributes (Figure 14).

As was said above, each one of the points is an estimate of the real location of that object in that sensorial space Ω_Ψ. Lets call the parametric location *Sensu* and denote by Ψ. Under a classic statistical point of view, the parameter Ψ (fixed and unknown) can be understood as a consensus or average location of that object when scored in that space by all possible assessors, i.e., by everyone that comprises the population under study. Therefore, each single score is an estimate $\hat{\Psi}_i$ of that sensu Ψ itself and, of course, each possible consensus is a better estimate of Ψ (Figure 14).

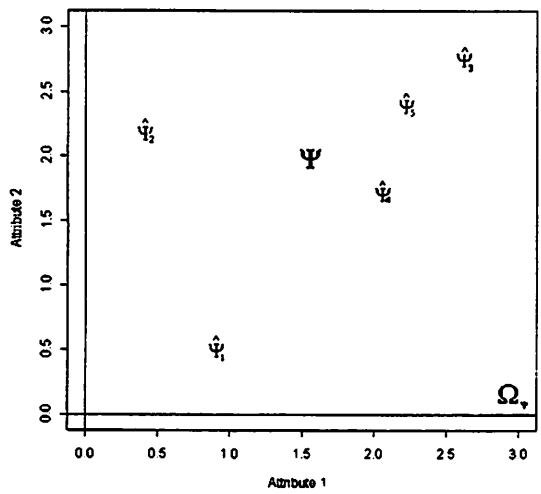


FIGURE 14: Scores of five assessors to one object according to a pair of attributes: generation of a sensorial space (Ω_Ψ), Sensu of the object (Ψ) and its estimates ($\hat{\Psi}_1$, $\hat{\Psi}_2$, $\hat{\Psi}_3$, $\hat{\Psi}_4$ e $\hat{\Psi}_5$).

It is worth noting that this section only didactically introduces an inference discussion since GPA is not able to be performed with only one object. GPA is a method of adjustment of assessors configurations, and it does by minimizing a criterion of simultaneously reducing the sum of squares of the distances among similar points (objects). Of course, it would turn out superimposed points (sum of squares zero) trying to do that with just one product. Therefore, GPA is a method of *comparison* and of taking *relative* conclusions.

3.5 Comparing objects

In a second case, assessors compare a set of objects scoring their main attributes in a similar way.

Let $\{x_{ijl}\}$ represent the scores of k judges. Suppose they are organised in matrices (illustrated in tabular form in Table 6), in which the rows refer to objects ($l = 1, \ldots, n$), the columns refer to the descriptor attributes ($j = 1, \ldots, p$) and each matrix contains the scores of a judge ($i = 1, \ldots, k$). Scores for the *ith* judge are in Table 6.

TABLE 6: Tabular representation of scores for n objects, under p attributes, given by the *ith* judge.

	Attributes			
Object	1	2	...	p
1	x_{11i}	x_{12i}	...	x_{1pi}
2	x_{21i}	x_{22i}	...	x_{2pi}
⋮	⋮	⋮	⋱	⋮
n	x_{n1i}	x_{n2i}	...	x_{npi}

For instance, consider that three objects were evaluated by four assessors, who scored their two main attributes using the same scale (e.g.

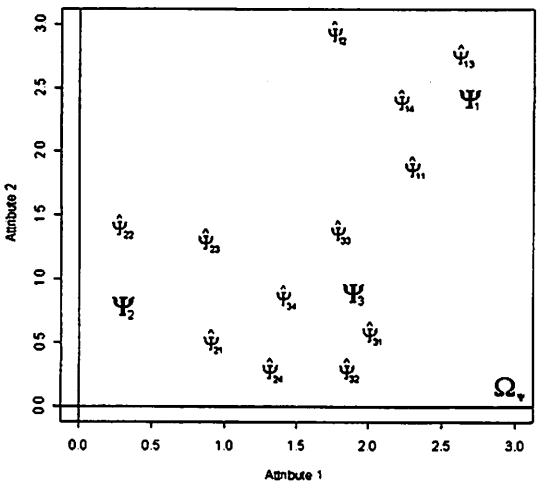


FIGURE 15: Score of four assessors ($i = 1, \ldots, 4$) to three objects ($l = 1, \ldots, 3$), with respect to two main attributes: generation of a sensorial space (Ω_Ψ), Sensu of the three objects (Ψ_l) and their respective estimates ($\hat{\Psi}_{li}$).

0 to 10). Then, assuming each attribute as a axe, the scores for each object can be seen as coordinates of points in a space spanned by those two attributes (Figure 15).

However, as can be also seen in Figure 15, one can suspect that both Sensus could be considered statistically equal, since their estimates are merged (mainly regarding objects 2 and 3). In order to decide if they are equal, one should determine the smaller distance between them that keep they been considered distinct.

Note that one is neither talking about Procrustes analysis nor any other statistical method yet. That is just the scene supposed that leads to a similar reasoning when one has an answer plane coming from GPA. In that case, one would have two (or more) axes, understood as linear combinations of the p scored attributes and, in a large sense, k assessors evaluating n objects.

3.6 GPA and Inference

Prior to GPA, let X_i $(i = 1, \ldots, k)$ denote a $(n \times p)$ matrix containing information (for instance, scores) about n products in n dimensions, given by the *ith* assessor. So, one could say that X_i *represents* the *ith* assessor, since all its entries come from him/her.

On the other hand, when one defines

$$Y_l = \sum_{i=1}^{k} e_i e_l' X_i, \tag{3.1}$$

Y_l $(k \times p)$ refers to the information for the *lth* product, each row coming from one assessor; e_i (i=1,...,k) and e_l $(l = 1, \ldots, n)$ are vectors of *zeros* with *one* in the *ith* and *lth* position $(i = 1, \ldots, k$ and $l = 1, \ldots, n)$.

Now, reconsider the Procrustes problem (2.3),

$$f(s_1, \ldots, s_k, u_1, \ldots, u_k, T_1, \ldots, T_k) = \sum_{i<i'}^{k} ||s_i(X_i - 1u_i')T_i - s_{i'}(X_{i'} - 1u_{i'}')T_{i'}||,$$

where X_i is a known $n \times p$ data matrix; s_i is a scaling factor; u_i is a translation vector; and T_i is a general transformation matrix $(i < i' = 1, \ldots, k)$.

Lets call X *transformed*, X_i^t, the information matrix X_i after all transformation set in (2.3),

$$X_i^t = s_i(X_i - 1u_i')T_i.$$

In turn, one could think about the Y_l matrix for transformed con-

figurations.

$$Y_l^t = \sum_{i=1}^{k} e_i e_l' X_i^t$$
$$= \sum_{i=1}^{k} s_i e_i e_l' (X_i - 1u_i')T_i. \qquad (3.2)$$

In practical situations, the probability distribution of the rows of Y_l is not known. Even if it was, it is not straightforward deriving the distribution for the rows of Y_l^t. Though Y_l^t is function of X_i just like Y_l, the transformations of translation, scaling, rotation, reflection etc, are customized per row of X_i, i.e. per assessor. Then, it is not possible to rewrite explicitly Y_l^t as function of Y_l. Therefore, even if the distribution of the Y_l rows was known, the distribution of Y_l^t rows was not. That is a great motivation for using intensive computational techniques.

Automatically, it is impossible to derive algebraic expressions for the distributions of the consensus (G), i.e. of the rows (estimates of products locations), which is the estimator one is interested in to infer about,

$$G = \frac{1}{k}\sum_{i=1}^{k} s_i(X_i - 1u_i')T_i = \frac{1}{k}\sum_{i=1}^{k} X_i^t.$$

In fact, one would want to derive simultaneously one probability distribution for each G row, in order to make inference (estimation and decision) for the n products at once. Since it is not possible, or at least not straightforward, to do that algebraically, resampling techniques can be used through intensive computational methods.

3.7 Monte Carlo simulation

Monte Carlo simulation is usually used as a first evaluation of new strategies and decision rules before take the risk to try them in real situa-

tions (Miranda, 2004).

Monte Carlo utilizes computational resources to generate samples according to theoretical known distributions in order to evaluate the behaviour of statistical techniques for validate them, i.e., enable them to be used in practical situations (Dachs, 1988).

3.8 Permutation tests

Permutation tests are computational efforts largely used to empirically validate methods and/or test hypothesis. According to Gower & Dijksterhuis (2004), the idea of permutation tests is to destroy the correspondence between the rows of the data matrices. In the GPA framework this means that the rows of the matrices X_i are permuted. The number of variables p in a matrix X_i is of no importance for the permutation tests since only the rows are permuted; it is immaterial whether the number p_i of variables in each X_i is constant. The set of all possible permutations of N rows of k matrices contains $(n!)^k$ permutations which will often be far too large a number to handle, so a much smaller random subset of permutations is used. Each permuted data set has a particular RSS, say σ_s, associated with it (s = 1, ..., 100, say). The value associated with the unpermuted data set is called the *empirical value*, σ_0. A certain significance value, can be attributed to the empirical value obtained. With RSS, the probability $P(\sigma_s > \sigma_0)$ gives the significance.

The null hypothesis is that σ_0 is not different from the other values σ_s. Gower & Dijksterhuis (2004) say that when this is true it means that GPA did not find any correspondence between the matrices X_i, but from another point of view, it can be considered that permutating the rows within matrices means to impose that all products are the same, i.e. H_0: $\sigma_0 = \sigma_s$ means H_0: $\Psi_1 = \Psi_2 = \ldots = \Psi_n$, where Ψ_l stands for the *lth* product.

When $p - value$ is small, say less than 5%, the empirical value σ_0 lies in the tail of the distribution of σ_s values. In this case it may be judged that there is likely to be a correspondence between the matrices detected by GPA and/or also means to admite that, at least, two objects are different.

It is worth noting that the same investigation about the equality of objects can be done, in a very similar way, through *bootstrap* procedures. Virtual panels can be easily resampled from the original data "mixing" the rows for imposing equality of objects and producing a similar distribution of RSS and computing $P(\sigma_s > \sigma_0)$. The same interpretations follow.

3.9 *Bootstrapping*

The technique of *bootstrapping* was first considered in a systematic manner by Efron (1979), although the generality of the method means that it was used in some particular circumstances before that time. The essence of *bootstrapping* is the idea that in the absence of any other knowledge about a population, the distribution of values found in a random sample of size n $(X_1, \ldots, X_k)$ from the population is the best guide to the distribution in the population. Therefore, to approximate what would happen if the population was resampled it is sensible to resample the sample. In other words, the infinite population that consists of the n observed values, each with probability 1/n, is used to model the unknown real population. The sampling is with replacement, which is the only difference in practice between *bootstrapping* and randomization in many applications (Manly, 1998).

Let θ be a parameter of interest and $\hat{\theta}_r$ its point estimate, called here *reference estimate*. To apply the *bootstrap* technique one draw several (e.g. thousands) of *bootstrap* estimates of the particular parameter $\hat{\theta}_b$, $b = 1, \ldots, B$, with replacement. Such *bootstrap* estimates generate the so called *bootstrap* distribution or empirical distribution of the estimator $\hat{\theta}$.

Such a distribution can be used to perform *bootstrap* hypothesis tests and construct *bootstrap* confidence intervals for the parameter or functions of the parameter of interest (Efron, 1993).

3.9.1 Confidence regions estimation

The simplest method for obtaining *bootstrap* confidence limits is called the standard *bootstrap* method (Manly, 1998). The principle here is that if an estimator $\hat{\theta}$ is normally distributed with mean θ and standard deviation σ, then there is a probability of $1-\alpha$ that the statement

$$P\left(\theta - z_{\frac{\alpha}{2}}\sigma < \hat{\theta} < \theta + z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha,$$

holds for any random value of $\hat{\theta}$, where $z_{\frac{\alpha}{2}}$ is the value that is exceeded with probability $\frac{\alpha}{2}$ for the standard normal distribution. This statement is equivalent to

$$P\left(\hat{\theta} - z_{\frac{\alpha}{2}}\sigma < \theta < \hat{\theta} + z_{\frac{\alpha}{2}}\sigma\right) = 1-\alpha,$$

which therefore holds with the same probability, i.e., determining a $100(1-\alpha)\%$ confidence interval.

With the standard *bootstrap* confidence interval, σ is estimated by the standard deviation of estimates of a parameter θ that are found by *bootstrap* resampling of the values in the original sample of data,

$$\hat{\sigma} = \sqrt{\frac{\sum_{b=1}^{B}\hat{\theta}_b^2 - \frac{\left(\sum_{b=1}^{B}\hat{\theta}_b\right)^2}{B}}{B-1}}.$$

The interval is then

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \times (\textit{Bootstrap} \text{ Standard Deviation})$$

The requirements for this method to work are that:

(a) $\hat{\theta}$ has an approximately normal distribution;

(b) $\hat{\theta}$ is unbiased so that its mean value for repeated samples from the population of interest is θ;

(c) *bootstrap* resampling gives a good approximation to σ (Efron, 1979).

Percentile method (Efron, 1979) determines a $100(1-\alpha)\%$ confidence interval setting as upper and lower limits the *bootstrap* estimates that contain between them $100(1-\alpha)\%$ of the ordered *bootstrap* estimates.

If a increasingly monotonic function exists, such that the transformed estimates $f(\hat{\theta})$ are normally distributed with mean $f(\theta)$ and standard deviation 1, then the mathematical form of the transformation does not need to be known.

Therefore,

$$P\left(f(\theta) - z_{\frac{\alpha}{2}} < f(\hat{\theta}) < f(\theta) + z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

and then,

$$P\left(f(\hat{\theta}) - z_{\frac{\alpha}{2}} < f(\theta) < f(\hat{\theta}) + z_{\frac{\alpha}{2}}\right) = 1 - \alpha.$$

If the function $f(\cdot)$ is known, one just need to apply the inverse to obtain the confidence interval for θ. However, that function is usually unknown. Then, one way of deriving the confidence interval for $f(\theta)$ is to resample the original sample and find the limits that exceed $\frac{\alpha}{2}$ (lower limit) and $1-\frac{\alpha}{2}$ (upper limit) of the transformed values.

However, due to the monotonic nature of the transformation function, the order of the transformed *bootstrap* estimates $f(\hat{\theta}_b)$ is the same as non-transformed $\hat{\theta}_b$. Therefore, $\alpha/2$ and $1-\alpha/2$ limits can be set on the ordered *bootstrap* estimates and no transformation is necessary anymore.

According to Miranda (2004), there are so many other rules for constructing *bootstrap* confidence intervals, like Hall's (1992), which is based on the same arguments of Efron (1979), but deals with residuals distribution.

3.9.2 Hypothesis testing

According to Manly (1998), *bootstrap* tests of significance have not been as well studied as *bootstrap* confidence intervals, although they represent an obvious application of the basic idea. Indeed, one way to carry out a *bootstrap* test of the hypothesis that the parameter θ takes the particular value θ_0 involves simply calculating a *bootstrap* confidence interval for θ and seeing whether this includes θ_0. This appears to indicate that the extension of the theory of *bootstrapping* to tests of significance is trivial. However, there is an important difference between the two applications that needs to be recognized. This is that when carrying out a test of significance it is important to obtain accurate estimates of critical values of the test statistic even if the null hypothesis is not true for the population from which the sample being bootstrapped came from. Basically it is a question of deciding exactly how the null hypothesis being tested should influence the choice of the statistic being bootstrapped.

Whatever test statistic S is used, a *bootstrap* test involves seeing whether the value of S for the available data is sufficiently extreme, in comparison with the *bootstrap* distribution of S, to warrant rejecting the null hypothesis. Generally the test can operate in a similar way to a randomization test. Thus suppose that large values of S provide evidence against the null hypothesis. Then the observed data provide a value S_1 of S, and *bootstrap* resampling of the data produces another $B-1$ values $S_2, S_3, \ldots, S_B$. All B test statistics are from the same distribution if the null hypothesis is true. Hence, S_1 is significantly large at the $100\alpha\%$ level if it is one of the largest $100\alpha\%$ of the B test statistics. Or, to put it a

different way, the significance level for the data is $p = \frac{m}{B}$, where m is the number of the statistics S_1 to S_B that are greater than or equal to S_1.

This argument applies with any number of *bootstrap* samples but generally a large number is better than a small number in order to reduce the effect of the random sampling from the *bootstrap* distribution. The key requirement for the test to be valid is that the *bootstrap* distribution really does mimic the distribution of S when the null hypothesis is true.

3.10 Methodology

Statistical software *R*, from version 2.4.0 to 2.6.0 (R Development Core Team, 2007), was used for all computations, estimations, tests, programming and analysis. Its worth noting that no significant differences were observed in the used functions, along such versions. Specific routines were programmed for every step of the data analysis (Appendix B).

Due to the relative the conclusions yielded by GPA and intending to associate statistical confidence to them, a *bootstrap* approach of statistical inference, for both estimation and decision processes, is suggested.

Comparing distances between possible location estimates for objects is an issue that can be seen through either an estimation or a decision point of view. In general, assuming a minimum significative distance around an object estimate is delimiting a sort of *sensorial confidence region* (in sensory analysis context) in $\mathbb{R}^{p^*}$ space, where p^* is the desired dimension. Then, each pairwise distance could be evaluated and/or a confidence region by itself could be an estimate of the "real" location of an object.

There are, at least, two ways of looking to this problem, they are:

(i) for each product, compute the confidence region around the consensus point (point estimate) in order to comprise $100\gamma\%$, say 95%, of the

possible point estimates generated by resampling the actual scores and then performing GPA (Figure 16a);

(ii) for each possible pair of point estimates (products), test the significance of the sample (reference) euclidian distance, by comparing it with quantiles of a distribution of euclidian distances under H_0 (Figure 16b).

For both situations, intensive computational techniques (as *bootstrapping*) are suitable due to the difficult to derive algebraically the distribution of such distances since the scores of each assessor suffer particular transformations during the match process of GPA (Section 3.6).

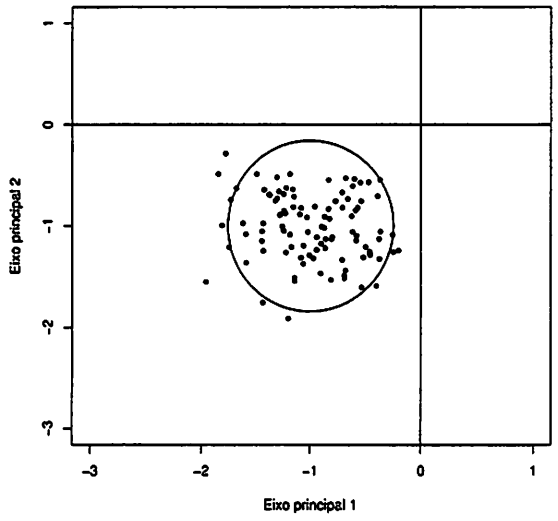
However, the huge number of possible *bootstrap* tests ($C_{n,2}$) required by the approach (ii) can turn it very hard to be performed; since it is usual to evaluate 12, 16 or even 60 objects in an experiment (Dijksterhuis, 1995), yielding 66, 120 and 1770 tests, respectively.

For that reason, it is preferred estimating the *bootstrap* confidence regions for the sensus estimates and perform tests for the euclidian distances only between those products for which the equality is suspected, i.e. for which the confidence regions touch each other.

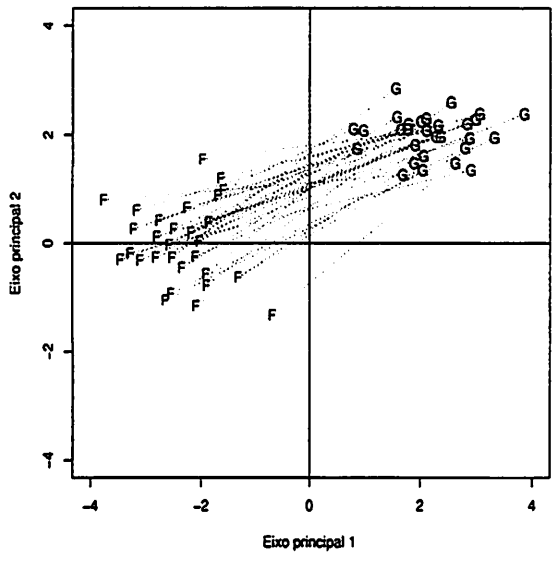
That recommendation is made since the worst error in Sensory Analysis is usually the Type II error, in the following pair of hypothesis

$$\begin{cases} H_0: \text{ Both objects are equal} \\ H_1: \text{ Both objects are different.} \end{cases}$$

Under a rigorous point of view, i.e. at infinity precision, any pair of objects is different. Therefore, committing Type I error, i.e. affirm that the objects are different when they are not, is not a serious error. However, asserting that two objects are the same, or at least undistinguishable according under a specific vocabulary, requires one to be very sure about that.



(a)



(b)

FIGURE 16: (a) Sensorial confidence region for one object comprising $100\gamma\%$ of the data generated by resampling one group average configuration ('mean' scores through assessors). (b) Illustration of the *bootstrap* distance between supposed objects F and G.

It is useful and important to predict the (empirical, in this case) size of the test (α) and its power ($1-\beta$) in some basic situations. That can be done using Monte Carlo methods of simulation.

One can compute several simulated random samples of scores and perform *bootstrap* inference (estimation and decision) of GPA accounting for the percentage of times it agrees with the "truth". Let α be the probability of incurring in Type I error. An estimator of α ($\hat{\alpha}$) can be the average percentage of times that Type I error occurs, i.e. the null hypothesis (all objected are the same) is rejected when it is true. Consequently, that is an estimator of the size of that test as well. Let β be the probability of incurring in Type II error. An estimator of β, say $\hat{\beta}$, can be the average percentage of times that, when accepting H_0 that decision is found to be false, i.e. Type II error occurs. Consequently, one an estimator of the power of that test is $1-\hat{\beta}$.

3.10.1 Estimation

One can estimate the confidence region for each point (object) via *bootstrap*. The idea is to resample, with replacement, the information (e.g. scores) matrices, each one representing a subject.

A discussion on the ideal resample unit is pertinent. Here, one is considering the whole data matrix X_i as the resample unit. That agrees with Husson et al. (2005) when estimating confidence ellipses for sensory profiles obtained by principal component analysis by *bootstrapping*. In fact, that resample unit seems to provide suitable estimates. However, more work is required in this issue. For instance, when one experiment has a small number of assessors, resampling the whole matrix X_i can lead to several identical estimates since many identical virtual panels are going to be formed. It can compromise the reliability of *bootstrap* results. Of course, the main cause for a poor reliability in this situation is the small

size of original sample. However, in those cases a more efficient resample unit must be searched. When the assessors scores have high correlation among each other, e.g. they were equally well trained, resampling rows or matrices produces nearly the same result. That happens because the new virtual assessor formed is very similar to a real one. If one resample rows in a low between-assessors correlation context, the virtual assessor formed would be far different from the rest of the panel, leading to inadequate estimates. In situations where the between-assessors correlation is *high enough*, the practice of resampling rows would round the low degrees of freedom problem.

Resampling with replacement one builds up a *bootstrap* sample (virtual panel) with the same size of the original one. That process is repeated several times and, for each *bootstrap* sample set up, one Generalised Procrustes Analysis is performed and the group average configuration computed. After produced several virtual panels, and therefore group average configurations, an ultimate GPA for adjusting all group (*bootstrap* + sample) averages and referring everyone to the same coordinate system. After that, one can see a cloud of points "around" the sample estimates.

For those points referred to the same axes, one can determine a $100\gamma\%$ confidence *bootstrap* confidence region, i.e. containing $100\gamma\%$ of the estimates for each product. When those points follow a symmetric bivariate distribution (for instance, Gaussian distribution), it is straightforward estimating that region through an *bootstrap* confidence ellipse (Figure 17a)[4].

Note that, if a vector $\mathbf{Z}$ contains p coordinates of a point that follows a multivariate Normal distribution $N(\mu, \Sigma)$, with mean vector μ and

[4]R package called *ellipse* brings a set of functions for estimating ellipses. One of them, with the same name, fits perfectly the present situation.

variance-covariance matrix Σ, then

$$(\mathbf{Z}-\mu)'\Sigma^{-1}(\mathbf{Z}-\mu) \sim \chi^2(p),$$

where $\chi^2(p)$ is the Chi-square distribution with p degrees of freedom (Johnson & Wichern, 1998). Therefore the approximated $100\gamma\%$ confidence ellipse (in $\mathbb{R}^2$) expressed in terms of sample quantities $\bar{x}$, $\bar{y}$, sample means and $\mathbf{C}$, sample variance-covariance matrix, is given by the equation[5]

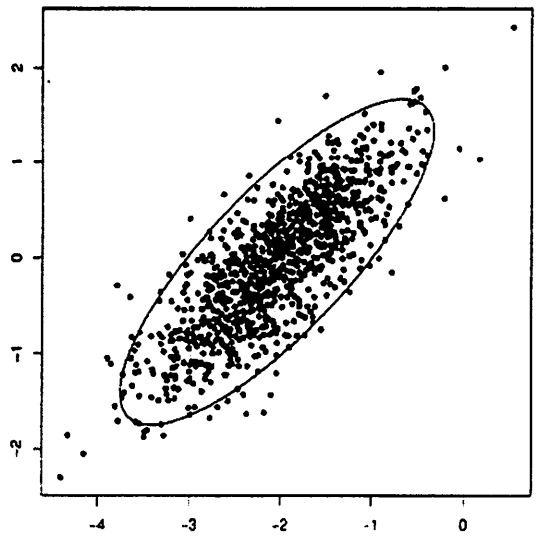
$$\frac{(x-\bar{x})^2}{det(\mathbf{C})/s_y^2} + \frac{(y-\bar{y})^2}{det(\mathbf{C})/s_x^2} - 2\frac{(x-\bar{x})(y-\bar{y})}{det(\mathbf{C})/s_{xy}} \leq \chi^2_\alpha(2),$$

where $\chi^2_2(\alpha)$ is the $100\alpha\%$ superior quantile of the Chi-square distribution with 2 degrees of freedom; and s_x^2, s_y^2 and s_{xy} are the sample variance of x, y and the covariance between x and y, respectively.

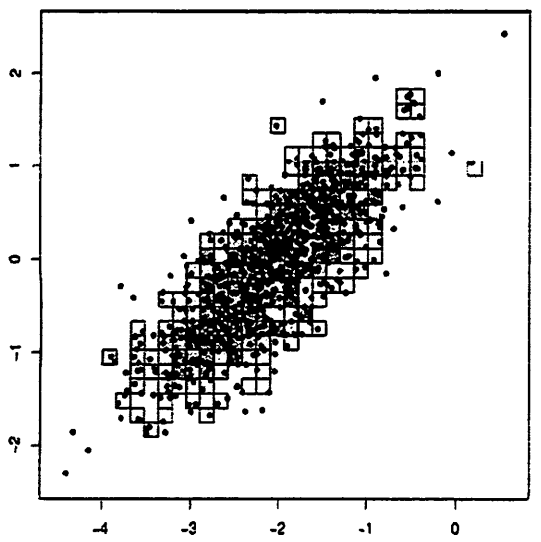
In more general cases, like asymmetric clouds of points, the ellipse fails to be a good estimator for the desired *bootstrap* confident region. Thus, a routine was developed and programmed in R language in order of approximate a suitable region for those cases. A very simple algorithm allows constructing a bivariate histogram delimiting a $100\gamma\%$ confidence region on the cloud of points, i.e. retaining $100\gamma\%$ of the points in, on the plane. A fictitious example is give on Figure 17b for illustrating both types of estimation.

As mentioned before, when detected intersection of regions, it is suggested to perform *bootstrap* hypothesis test for deciding the significance of the euclidian distance between the pair of points. In that test, the null hypothesis $H_0 : \Psi_l = \Psi_{l'}$, where Ψ_l and $\Psi_{l'}$ stand for any pair of objects, is imposed by 'mixing' only that pair of products through all assessors and forming virtual panels that score only those two products. One GPA is performed for each virtual panel and the euclidian distances under the null hypothesis are recorded. The empirical distribution of the *bootstrap* euclid-

[5]Derivation is given in A.13.



(a)



(b)

FIGURE 17: (a) Estimate of a 95% confidence region base on bivariate normal distribution. (b) Estimate of a 95% confidence region base on bivariate histogram estimator; for the same hypothetic cloud of points.

ian distances is displayed highlighting the position of the sample euclidian distance. After that, the sample euclidian distance is considered significant if $100\gamma\%$ of the *bootstrap* euclidian distances are less than it. The decision process is detailed in the next section.

It is worth stressing that GPA conclusions under the proposed *bootstrap* inference method do *not* loose their relative feature. During the whole method, what is estimated or tested is the relative location of objects and the distance between objects in a generally scaled R^p coordinate system. In other words, the scale is immaterial for conclusions.

3.10.2 Decision

Let δ be the euclidian distance between a pair of products. The sample provides an reference estimate ($\hat{\delta}_r$) of the parametric euclidian distance, according to a set of assessors.

Based on that, some hypothesis can rise. Consider the following plausible pair of hypothesis H_0 and H_1:

$$\begin{cases} H_0 : \delta = 0 \\ H_1 : \delta > 0 \end{cases} . \tag{3.3}$$

Hypothesis H_0 states that the pair of evaluated products are equal. Of course, δ will hardly be exactly zero. Statement (3.3) turns clear that $\delta = 0$. In other words, the reference euclidian distance is statistically equal to those found when the products are imposed to be the same.

That can be done by "mixing" the scores of a specific pair of objects all together and drawing them with replacement for composing virtual panels, new scores matrices. By scores to be mixed is meant the entire rows of the actual matrices of scores, i.e. the vector of coordinates for a product, along all assessors.

After computing, say, B *bootstrap* estimates $\hat{\delta}_b$ $(b = 1, \ldots, B)$ are ordered and compared with the reference sample value $\hat{\delta}_r$ (Figure 18). If the number of *bootstrap* estimates n_b that is greater than the reference sample value is small enough, the decision to reject H_0 is taken with significance $\hat{\alpha}_b = \frac{n_b}{B}$. Otherwise, H_0 is accepted. A *small enough* value is every value less than a critical value. The critical value is determined by the test and is as small as more rigorous is the test, since it is proportional to the probability to incur in Type I error.

This way, one can identify the magnitude of the ordinary distances or random distances between the objects (according to a specific panel) when the products are the same. Therefore, a sample distance greater than the most *bootstrap* distances indicates that the evaluated products are distinguished (considered different) by those assessors.

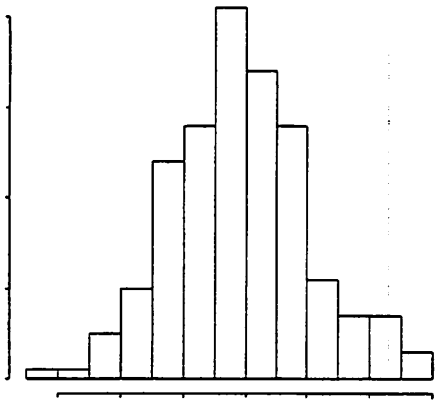


FIGURE 18: Illustration of an empirical distribution (set of ordered values $\hat{\delta}_b$) and highlight on a possible location of the sample reference value ($\hat{\delta}_r$).

In order to summarize the proposed methodology, Figure 19 shows a flowchart that displays the suggested ways of deciding to accept or reject the hypothesis that certain pair of objects can be considered equal. It goes from the sensory data matrices X_i to the final decision, passing through *bootstrap* resampling for confidence regions and/or hypothesis testing. Ba-

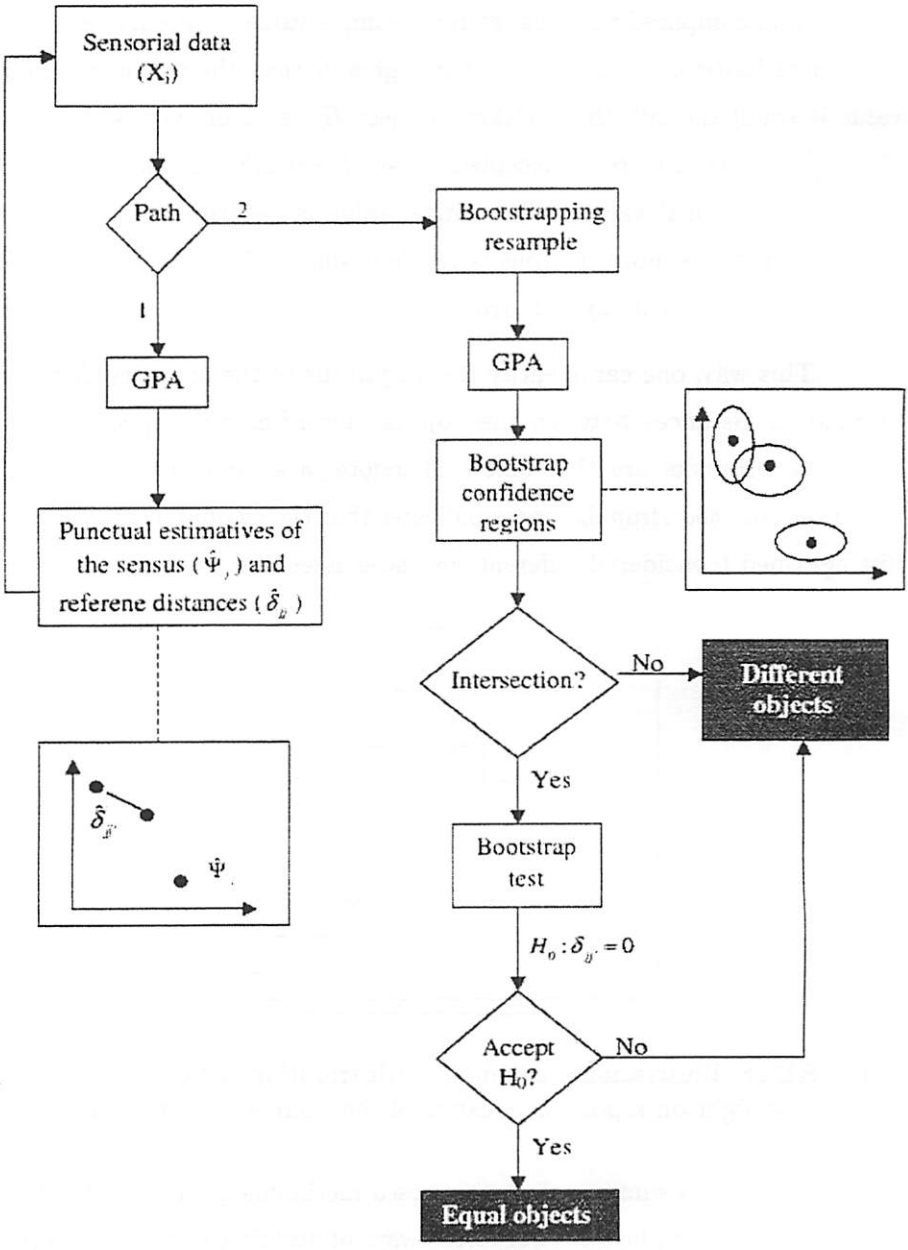


FIGURE 19: Flowchart for the proposed methodology.

sically, the flowchart is formed by two (didactically distinct) parts: (1) point estimation of the location of the evaluated objects in a sensory space; and (2) estimation of sensory confidence regions and *bootstrap* tests. When one decides going on path 1, conventional GPA is performed and, obligatorily one has to return to initial data for path's 2 procedure. On path 2, *bootstrap* resampling provides the generation of several virtual panels which determine the *bootstrap* confidence regions estimates. Then, visual inspection for intersection of confidence regions can be done. In visual inspections for each pair of objects is made a decision whether their are different or further investigation can be done. When the later option is chosen, a *bootstrap* hypothesis test is performed for each pair of objects for which the equality is suspected. It is worth noting that there is no impediment to test the equality hypothesis even when there is no intersection of the confidence regions, however that action does not counts for optimizing the process.

3.10.3 Algorithm

This section aims to summarize the proposed method in form of algorithm. Steps are as follow:

(1) Group the information about n objects, under p dimensions, given by k assessors in k matrices $n \times p$;

(2) Perform GPA on sample data and compute the group average configuration G_s;

(3) Resample the matrices of original data forming a virtual panel. Perform GPA in that panel and compute the group average configuration G_{b1};

(4) Repeat (3) a *high enough* number of times B yielding $G_{b1}, G_{b2}, \ldots, G_{bB}$;

(5) Perform an ultimate GPA on $\{G_s, G_{b1}, \ldots, G_{bB}\}$ and reduce the dimensionality to a low number (e.g. two) of dimensions, in order to concentrate the variation for plotting;

(6) Construct $100\gamma\%$ confidence regions (not necessarily ellipses) for the *ith* object ($i = 1, \ldots, n$);

(7) If no ellipses intercept each other the process stops and (besides usual conclusions) the objects are said to be distinguishable. Otherwise, go to (8);

(8) Test each pair of intercepted ellipses by imposing H_0 of equality. It is done "mixing" only two such rows among assessors and producing virtual panels. Then, for each virtual panel produced, a GPA is performed and the negligible distance between such pair of objects is computed. After a sufficient number of distance computed, the sample distance is compared with the empirical distribution of distances under H_0. Percentage of values greater than the sample distance is understood as an exact p-value. Normal decisions are made confirming or refuting the suspect of equality.

3.10.4 Practical experiments

Two practical data example were used to illustrate both confidence region estimation and some hypothesis tests.

a) The first one was kindly available by Fernando A. R. Magalhães, from EPAMIG/Instituto de Laticínios Cândido Tostes. It is regarded to sensory evaluation of gorgonzola cheese, from two distinct technologies. Nine assessors scores eight different cheeses (4 ripening ages from each technology) in a Quantitative Descriptive Analysis (QDA) context, according to ten attributes: appearance (Ap), amount of mold (AM), texture (Tx), characteristic odor (CO), aroma (Ar), characteristic flavor (CF), salty taste (ST), acid taste (AT), bitter taste (BT), residual flavor (RF). Cheeses from 1 to 4 come from technology 1; and from 5 to 8 come from technology 2. Ripening ages are, respectively 30, 45, 60 and 70 days. Magalhães (2002) explore these data via univariate analysis of variance (ANOVA) and Ferreira (2004) analysed the same data set through multivariate analysis of variance (MANOVA) and Generalised

Procrustes Analysis (GPA). However, conclusions can still be enhanced with inference techniques.

b) The second example was kindly available by Wageningen Centre for Food Sciences (WCFS) through Garmt Dijksterhuis, Pieter Vereijken and Eric Dransfield. It is a two experiment data set were 90 naïve assessors were asked to evaluate 19 (9+10) commercial semi-solid food products purchased in The Netherlands, basically custards, yoghurts, mayonnaises (experiment 1) and fresh creams, soft cheeses and creams (experiment 2). They were evaluated according to a nine attribute vocabulary: dairy taste (DT), thickness mouth-feel (TMF), airy mouth-feel (AMF), melting mouth-feel (MMF), heterogeneity mouth-feel (HMF), creamy soft mouth-feel (CMF), fatty mouth-feel (FMF), sticky mouth-feel (SMF) and astringent mouth-feel (AsMF).

3.10.5 Simulation studies

Inspired by practical data sets two simulation studies were held to validate the described method and estimate the power and type I error rates of the estimation and decision procedures.

a) Based on the Gorgonzola data set, virtual panels were simulated from multivariate Gaussian distributions. Their parameters (mean vectors and variance-covariance matrices) were the sample mean vectors and sample variance-covariance matrices from the practical data set. Data set had 720 records from an experiment of 8 products × 9 assessors × 10 attributes. The reason of simulating from Normal distributions was that it is the more used and well known distribution and it has the *good* properties.

$$X_i \sim N(\mu_i = \bar{x}_i, \Sigma_i = S_i), \quad i = 1, \ldots, 8.$$

Five hundred virtual panels were simulated under complete alternative hypothesis $H_1 : \Psi_1 \neq \Psi_2 \neq \ldots \neq \Psi_N$ and the *bootstrap* inference method was performed in order to evaluate the power rate. Percentage of wrong decisions was accounted in each step (estimation and decision) and in the whole process.

b) In order to better explore the behaviour of the described method, another simulation study was held. Six products were carefully chosen from the Dutch dairy sensory experiment: products 1, 2, 4 and 6 from the first experiment and products 5 and 7, from the second one. From now on, they are referred as products 1 to 6, respectively.

In this simulation study, the partial $H_0 : \Psi_1 \neq \Psi_2 \neq \Psi_3 = \Psi_4 \neq \Psi_5 = \Psi_6$ was imposed. For products 1 e 2, were considered the sample mean vectors and the sample variance-covariance matrices. For pairs $(3,4)$ and $(5,6)$, were considered the average of the sample mean vectors and the average of the sample variance-covariance matrices, for each pair. Therefore, three distinct groups were formed: (i) $\Psi_1 \neq \Psi_2$; (ii) $\Psi_3 = \Psi_4$; and (iii) $\Psi_5 = \Psi_6$ (Figure 20).

$$
\begin{aligned}
W_1 &\sim N(\mu_1 = \bar{x}_1, \Sigma_1 = S_1) \\
W_2 &\sim N(\mu_2 = \bar{x}_2, \Sigma_2 = S_2) \\
W_3, W_4 &\sim N(\mu_{3,4} = \bar{\bar{x}}_{3,4}, \Sigma_{3,4} = \bar{S}_{3,4}) \\
W_5, W_6 &\sim N(\mu_{5,6} = \bar{\bar{x}}_{5,6}, \Sigma_{5,6} = \bar{S}_{5,6})
\end{aligned}
$$

where W_l $(l = 1, 2, \ldots, 6)$ are random vectors of length $p = 9$.

For sample sizes of $k = 5, 10, 15, 20, 50, 90$ assessors, five hundred replications were drawn and the described method was performed in order to evaluate power and type I error rates.

Eight rates were measured at each replication for evaluating the behaviour of the described method:

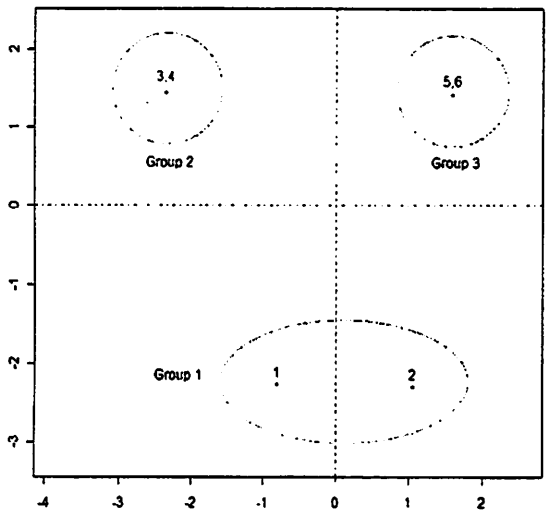


FIGURE 20: Illustration scheme of the second simulation study.

i) *"Power" within group 1 (R_1)*: in fact, it was estimated the probability to do not move to the second stage of the method (testing), that is, identify that confidence regions of products 1 and 2 do not touch each other and then they are said to be different without the need of a formal test. It was estimated by 1 minus the ratio between the number of interception of regions over the number of replications.

$$R_1 = 1 - \frac{\#\text{ of interceptions}(1,2)}{\#\text{ of replications}} \tag{3.4}$$

ii) *"Type I error rate" within group 2 (R_2)*: probability of considering products 3 and 4 different at the first stage (estimation), that is, percentage of wrong decisions, when the confidence regions 3 and 4 do not touched each other.

$$R_2 = 1 - \frac{\#\text{ of interceptions}(3,4)}{\#\text{ of replications}} \tag{3.5}$$

iii) *"Type I error rate" within group 3 (R_3)*: probability of considering products 5 and 6 different at the first stage (estimation), that is,

percentage of wrong decisions, when the confidence regions 5 and 6 do not touched each other.

$$R_3 = 1 - \frac{\#\text{ of interceptions}(5,6)}{\#\text{ of replications}} \tag{3.6}$$

iv) *"Power" between groups (R_4)*: probability of considering the three groups to be different at the first stage (estimation), that is, percentage of correct decisions, when the confidence regions of the groups do not touched each other.

$$R_4 = 1 - \frac{\#\text{ of interceptions between groups}}{\#\text{ of replications}} \tag{3.7}$$

v) *Type I error rate within group 2 (R_5)*: since one moved to the second stage of the method (testing), percentage of times it rejects the null hypothesis $H_0 : X_3 = X_4$ ($\alpha = 5\%$).

$$R_5 = 1 - \frac{\#\text{ of rejections of } H_0 : X_3 = X_4}{\#\text{ of tests performed}} \tag{3.8}$$

vi) *Type I error rate within group 3 (R_6)*: since one moved to the second stage of the method (testing), percentage of times it rejects the null hypothesis $H_0 : X_5 = X_6$ ($\alpha = 5\%$).

$$R_6 = 1 - \frac{\#\text{ of rejections of } H_0 : X_5 = X_6}{\#\text{ of tests performed}} \tag{3.9}$$

vii) *Whole type I error rate within group 2 (R_7)*: summation of type one error rates of the first (estimation) and second (testing) stages of the inference method, for group 2.

$$R_7 = R_2 + R_5 \tag{3.10}$$

viii) *Whole type I error rate within group 3 (R_8)*: summation of type one error rates of the first (estimation) and second (testing) stages

of the inference method, for group 3.

$$R_8 = R_3 + R_6 \tag{3.11}$$

3.11 Results

According to the enumeration of the examples described in the Methodology, the results and discussion follow. First, two practical experiments are discussed: the Gorgonzola experiment with well trained assessors (a) and the two Dutch commercial semi-solid foods experiments with 90 assessors (b).

3.11.1 Practical experiments

(a) Gorgonzola experiment (Ferreira et al., 2007)

Figure 21, adapted from Ferreira (2004), shows the two principal axes after Generalised Procrustes Analysis for eight gorgonzola cheeses (1 to 8), evaluated by nine assessor (A to I). Ferreira (2004) argues that, sometimes, the location estimates for some products are so close that one could think they are undistinguishable by the group consensus. Though there are no so close estimates in this case one could wonder whether some distances are significant (for instance, products 1 and 6 or 3 and 7). To solve that problem one can, for instance, determine the confidence region for each cheese based on their consensus estimates. Thus, 95% confidence regions were built based on bivariate normal distributions. $B = 100$ virtual panels were formed drawing B *bootstrap* samples, with replacement, from the original data set, performing B GPA's and computing their group average configurations (Figure 22). Figure 22a shows 100 *bootstrap* estimates (black numbers) and the sample reference estimate (white number), while Figure 22b displays the same scene highlighting the *bootstrap* confidence re-

gions and the sample reference estimates. Both biplots bring the attributes (described in Section 3.10.4) positioning.

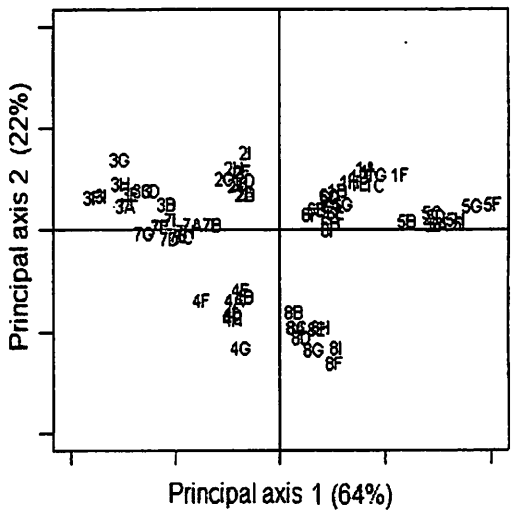
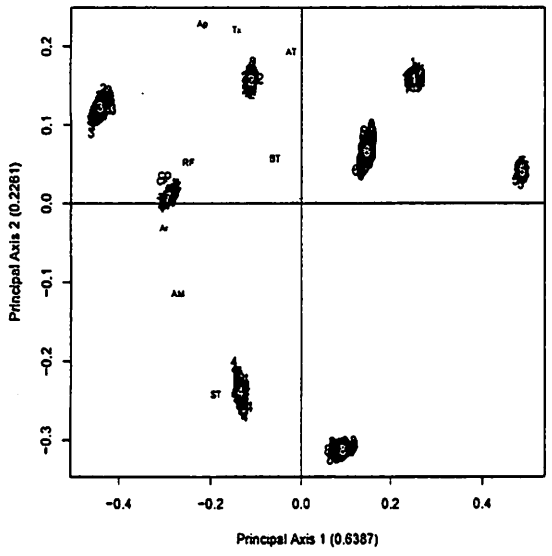


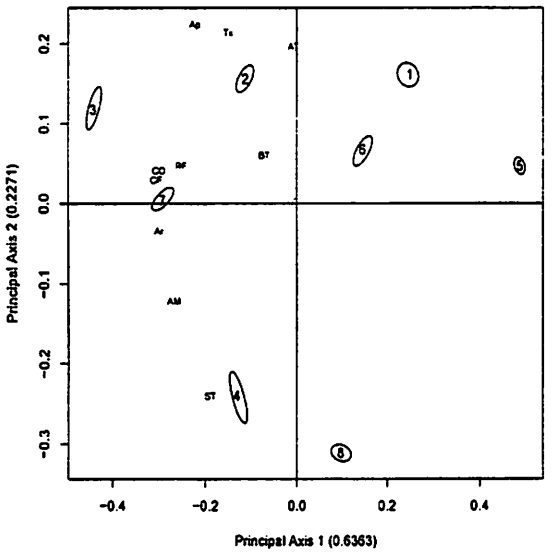
FIGURE 21: Transformed scores for eight products (1 to 8), by nine assessor (A to I), on the response-plane of GPA.

As was said, looking to Figure 21, one could suspect that this panel would confound products 1 and 6 or 3 and 7. That is not the same impression one has looking to Figure 22. Since there is no intersection between ellipses, according to the proposed methodology (Figure 19), one could conclude all cheeses are distinguishable, with at least, 95% of confidence.

Although conclusions for this case seem to be clear, suppose one insists in performing a test for the equality between the pairs of cheeses $(1, 6)$ and $(3, 7)$. Figure 23 shows empirical distributions of euclidian distances between such pairs of cheeses, under the null hypothesis. They were computed from 500 ordered *bootstrap* observations of null euclidian distances. The euclidian distance between the pair $(1, 6)$, for them to be considered undistinguishable, seems to be something around 0.09. Since the sample euclidian distance is more than 0.19 and according to the empirical distribution (a) of Figure 23, it is highly advisable do not to accept H_0: $\delta_{1,6} = 0$

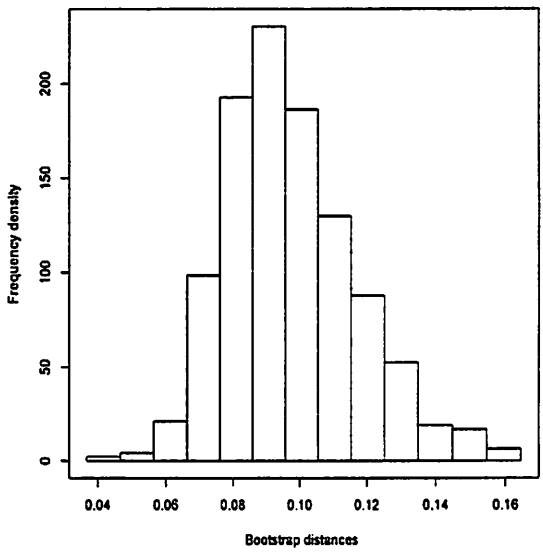


(a)

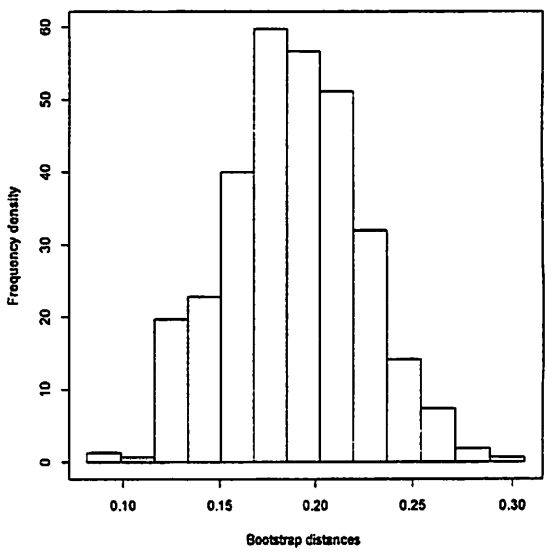


(b)

FIGURE 22: (a) Estimated *bootstrap* confidence regions for the 8 cheeses. *Bootstrap* estimates are in black and the sample estimate in white. (b) Same scene hiding the *bootstrap* estimates and highlighting the 95% confidence region. Attributes position are highlighted.



(a)



(b)

FIGURE 23: Empirical distribution of the euclidian distances based on 500 *bootstrap* estimates under the null hypothesis. (a) $H_0 : \delta_{1,6} = 0$ and; (b) $H_0 : \delta_{3,7} = 0$

(p-value= 0), i.e. cheeses 1 and 6 are not close enough to be equal.

Figure 23b highlights that the minimum significant distance between products 3 and 7 is settled around 0.17, while their sample distance is 0.33. Therefore, for cheeses 3 and 7, one can not accept $H_0 : \delta_{3,7} = 0$, in a 5% significance *bootstrap* test. Those conclusions play an important role since products 3 and 7 are cheeses of the same ripening age, but coming from different technologies, i.e. cheeses from those technologies seem to be differently perceived at that age (60 days). On the other hand, cheese 1 is the 30 days old of technology 1 while cheese 6 is the 45 day-old, from technology 2. They are of course supposed to be perceive in different ways, but it surprises the fact that cheese 1 is closer to 6 than to cheese 5 (30 days, technology 2). Summarizing and noting the location of the desirable attributes, one can conclude that the 8 chesses are clearly distinguished under that vocabulary and technology 1 do present higher scores during all the ripening period (30 to 75 days).

In Section 3.10.1, a discussion about resample unit was initiated. Due to fact that in this data set the assessors were well trained, and then highly correlated, a *bootstrap* estimation process based on *rows* as resample unit gave the same results as matrices (not published). In order to assess panel consonance a method proposed by Dijksterhuis (1995) was used. The idea is simply perform a PCA on matrices containing in columns all assessor, about the same variable, and expect the most variance to be explained by the first dimension. C statistic was proposed by him for measuring consonance, given by

$$C = \frac{\lambda_1}{\sum_{i=2}^{k} \lambda_i},$$

where λ_i is the ith eigenvalue of $X_j'X_j$, and X_j is a $(n \times k)$ matrix of information for the jth attribute.

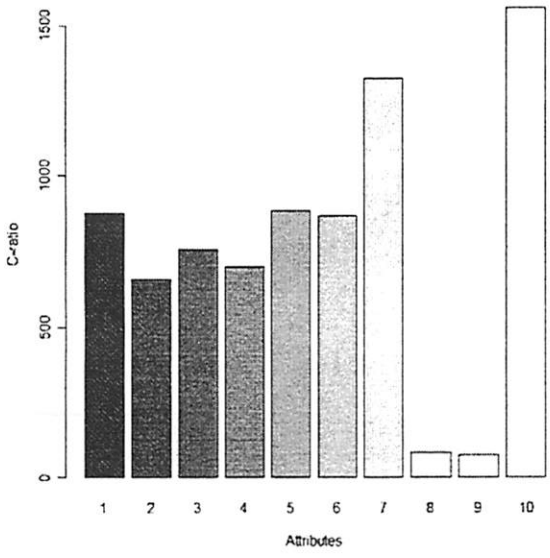
Figure 24 shows a bar plots of C statistic for all 1-10 attributes

used in Gorgonzola evaluation. However, note that Figure 24a is a kind of misleading. It is important to say that the numbers 1 to 10 for attributes are related to the order presented above, i.e. 1– appearance (Ap), 2– amount of mold (AM), ..., 10– residual flavor (RF). Its smallest bar (attribute 9) seems to indicate a variable poorly interpreted by the assessors, but its first eigenvalue is 70 times greater than summation of the other ones! There is a scale "problem" here, since the attribute 10 has a first eigenvalue 1500 times the sum of the other ones. In this practical example, all attributes are incredibly well understood and well scored by the assessors. It seems to be a weak point for Dijksterhuis' C statistic. When a attribute gets close to the ideal situation, be explained in 1 dimension, $\sum_{i=2}^{k} \lambda_i$ is tiny, then C tends to infinity. In order to turn them less misleading, Figure 24b shows the same C statistics in a logarithm scale.

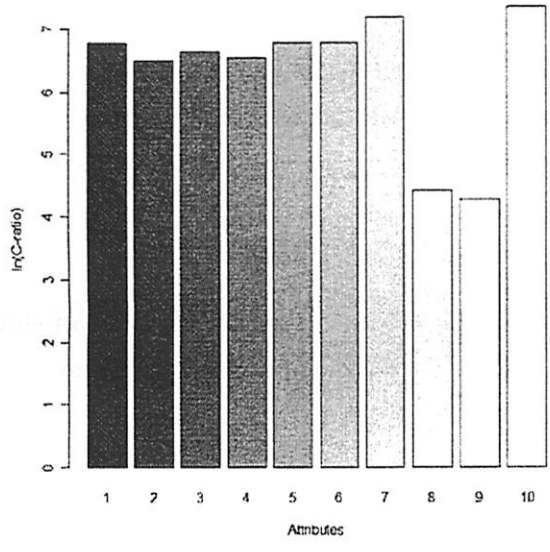
Another tools suggested by Dijksterhuis (1995) are a common scree plot for all attributes and plots of assessors space in order to search for bad trained. As this situation shows to be close to ideal and all attributes were well scored, scree plots are expected to be similar to each other and well designed, as can be seen on Figure 25. For all of them, the first dimension explains very far the most variance and their explanation is very similar.

In Figure 25 it is really hard to check that attribute 10 behaves better and attribute 9 is the worst. In order to explore in more details the assessors behaviour per attribute is shown on Figure 26a and Figure 26b the positions of all assessors in attributes 10 and 9 (better and worst, respectively). Even in the worst consonance, the panel scored quite well, i.e. Figure 26b shows assessors very close to the first dimension, indicating a extraordinary overall panel consonance.

Another as interesting as expected feature is the variance accounted for on the axes. As can be seen almost all the variations is explained in the first component, what confirms good training.



(a)



(b)

FIGURE 24: Bar plots for attributes 1 to 10 of Gorgonzola experiment. (a) Highlights C statistics and (b) their logarithm.

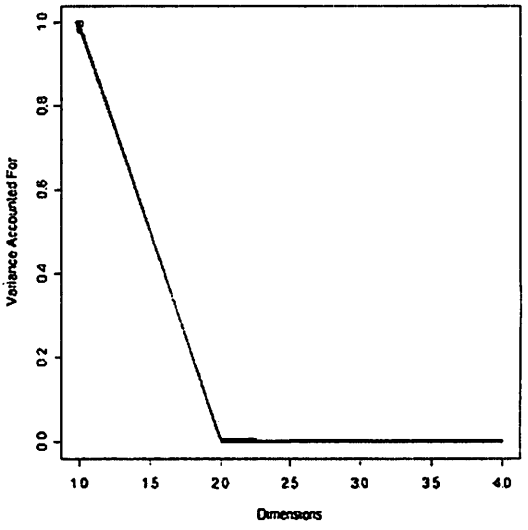


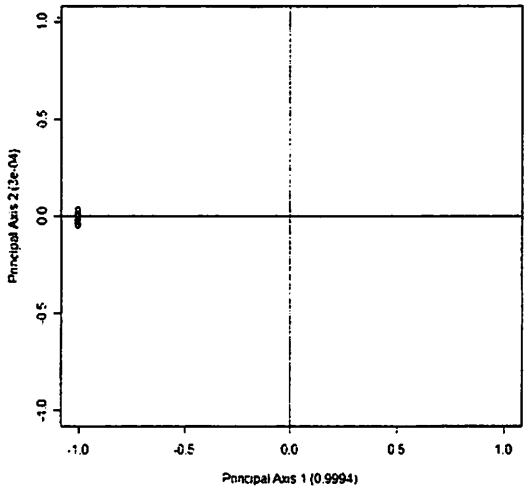
FIGURE 25: Scree plot for attributes 1 to 10 in Gorgonzola experiment.

(b) Commercial semi-solid foods in The Netherlands

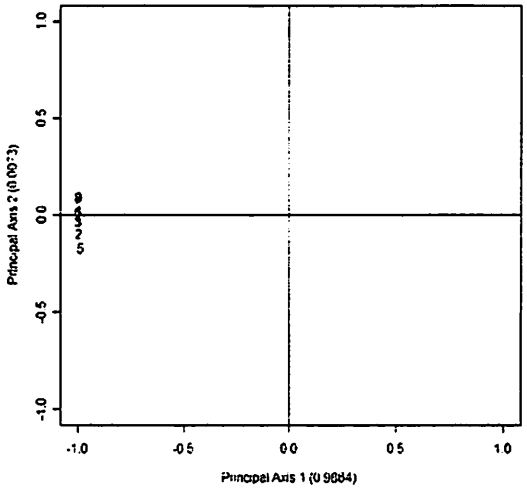
Estimating confidence regions for products' locations can be done as a first step of the analysis, that is, it is not necessary to perform conventional GPA and then *bootstrapping* after a diagnostic. Figure 27a shows the products of experiment 1 with their 95% confidence ellipse and the nine attributes, in a biplot.

According to the proposed method, the pairs of ellipses that touch each other are now tested for significance of euclidian distances between them. Then, pairs $(1,7)$, $(3,7)$, $(4,6)$, $(5,6)$, yielding, respectively p-values: 0.46, 0.58, 1 and 0.24.

It is clear in Figure 27a that assessors perceive 3 groups of products differently. Products numbered 1, 3 and 7 are custards; products 2, 8 and 9 are yoghurts; and products 4, 5 and 6 are mayonnaises. It is interesting to note that yoghurts are well distinguished between them, while mayonnaises are a kind of confounded, as well as custards.



(a)



(b)

FIGURE 26: Assessors space for attributes 10 - residual flavor (a) and 9 - bitter taste (b).

In turn, Figure 27b shows fresh creams, soft cheeses and creams. Clusters here are not so well define as in Figure 27a, but the adicional interpretation is quite interesting. Products 1 and 2 are fresh creams. Note that they would not be well distinguished by a possible population those assessors come from. Additionally, products 10, a soft cheese, is confounded with them. Products 7, 8 and 9 are creams. Note that 7 and 9 are much more similar than 8. A very far-from-everyone is product number 5, a Mozzarella, that share few attributes with the other products. Finally, 3, 4 and 6 are cheeses probably with intermediate characteristics.

Again, products suspect of equality were tested by the *bootstrap* hypothesis test algorithm. Proposed method was efficient for identifying pairs $(1,2)$, $(2,10)$ and $(1,10)$, as similar and testing them, yielding p-values 1 (100%) for all them. That confirms that these three products are perceived the same way according to a specific vocabulary.

Investigating panel consonance, one can see a scene less "ideal" than the Gorgonzola experiment. That is, assessors are less trained and therefore agree less, in this case. It can be seen clearly on Figures 28 and 29.

In experiment 1, attribute 3 is the more unidimensional one. On the other hand, attribute 6 presents itself a little bidimensional (Figure 28a), even though its C is almost 5, that is, its first dimension explains almost five times the sum of the other dimensions. Figure 28b shows the variance explained by the first four dimensions for the nine attributes. Attribute 3 shows a high first dimension and a very low second one, what turns it the better understood one. However, attribute 6, besides presenting the lowest first dimension, its shows the higher second dimension, then it can be considered to have the poorest understanding.

Figure 30a and 30b show the assessors' space for attributes 3 and 6, respectively. Most unidimensional attribute 3 shows most of the assessors scoring quite well. However it can be identified assessors 74 and 82 scoring

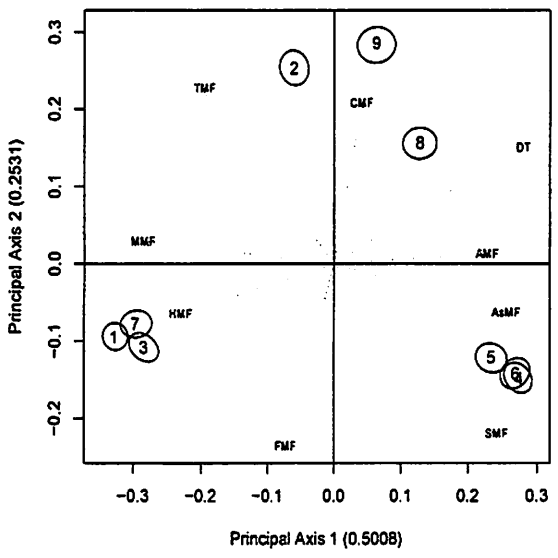
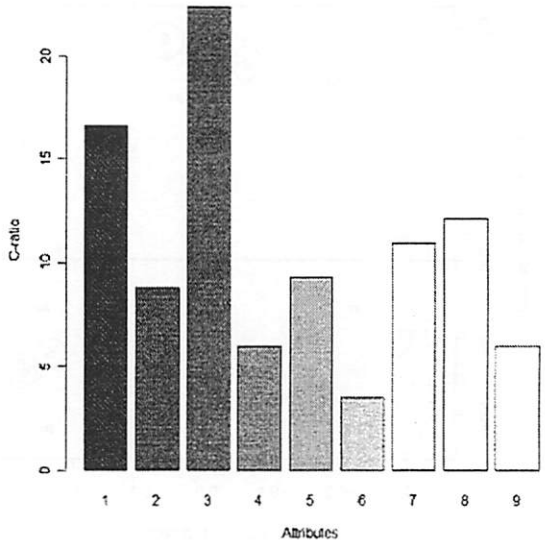
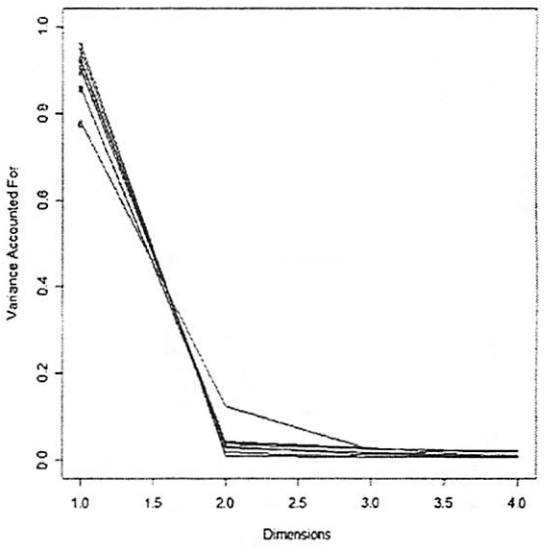


FIGURE 27: Two first principal axis, highlighting 95% confidence ellipses and attributes, for experiments (a) and (b) with comercial dairy products in The Netherlands.

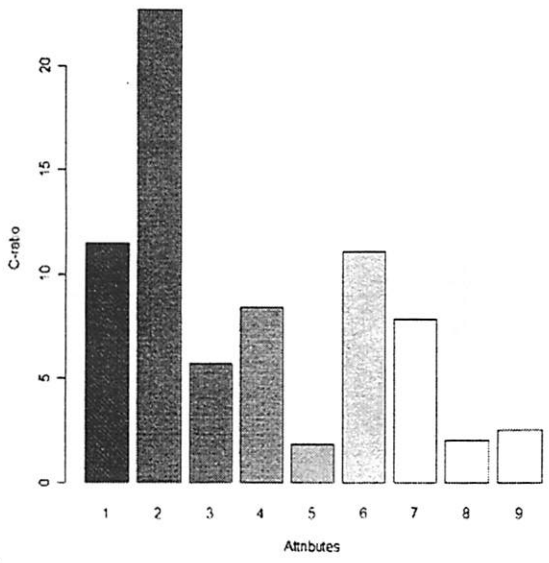


(a)

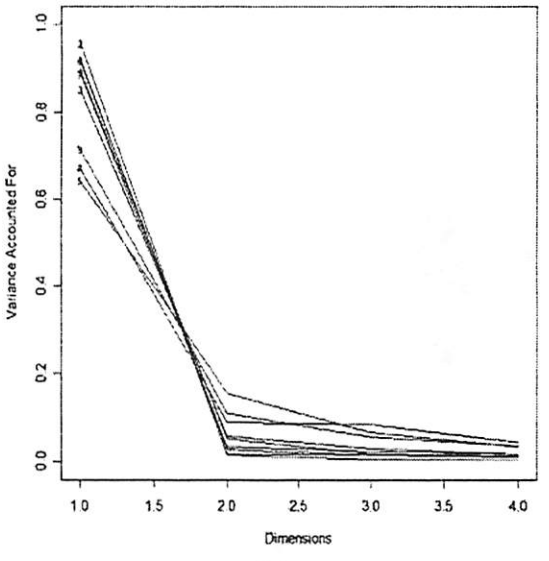


(b)

FIGURE 28: (a) Bar plot of the statistic C and; (b) scree plot for the first four dimensions, of all nine attributes of experiment 1 of Dutch comercial products.

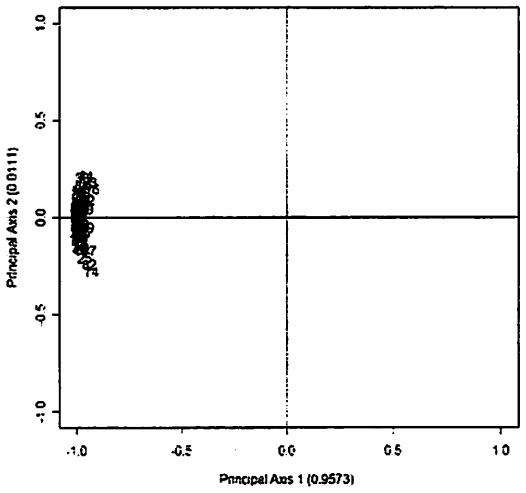


(a)

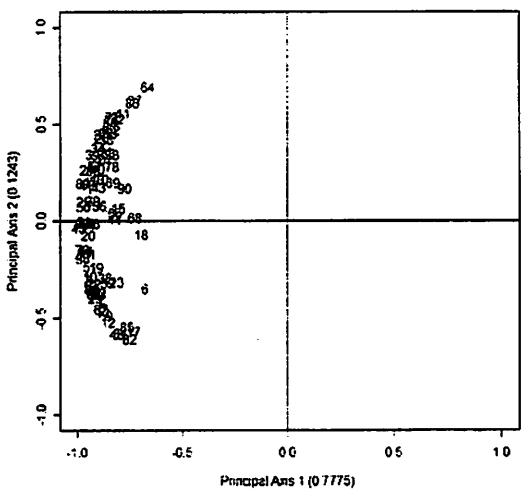


(b)

FIGURE 29: (a) Bar plot of the statistic C and; (b) scree plot for the first four dimensions, of all nine attributes of experiment 2 of Dutch comercial products.



(a)



(b)

FIGURE 30: Assessors' space for the most and the less unidimensional attribute of Dutch experiment 1, respectively attributes 3 (a) and 6 (b).

worst, that is, more bidimensional.

On the other hand, on Figure 30b can be seen that several assessors are scoring attribute 6 in two dimensions. Assessors 64 and 82 (again) are highlighted as the most discrepant ones.

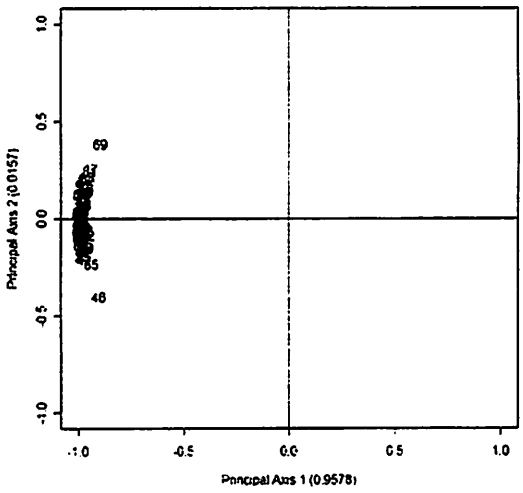
In experiment 2, attribute 2 is by far the better understood, the better expressed in one dimension. Its C is greater than 20, i.e. its first dimension accounts more than 20 times the other ones (Figure 29a). The same behaviour of attribute 2 is reproduced by Figure 29b, where it can be seen it has the highest first dimension and the lowest second dimension. On the other hand, although Figure 29a shows attribute 5 to have the lowest first dimension, Figure 29b reveals attribute 9 to leave the great amount of variation to be explained by the second dimension. Therefore, attribute 9 can be considered the more two dimensional attribute.

For the most unidimensional attribute (2), Figure 31a shows assessors 46 and 69 as the most bidimensional ones when scoring that attribute. In turn, when scoring attribute 9, the most bidimensional one (Figure 31b), several assessors seem to understand it wrongly, highlighting assessors 26, 61 and 78.

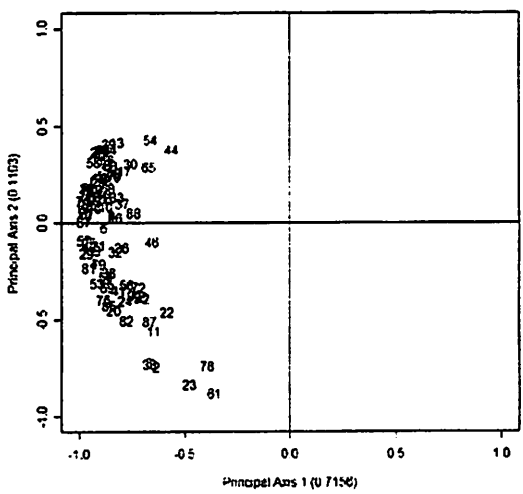
It is interesting comparing Figure 26b and Figure 31b. Note how assessors are better trained in Gorgonzola's experiment. By the way, in experiment with comercial Dutch products, the assessors are not specifically trained. Even if they were, training 90 assessors would be a very harder task than training nine.

3.11.2 Simulation studies

a) Along 500 drawings from multivariate Normal distributions, with parameter coming from the Gorgonzola data set, the behaviour of the described method was stable and well defined. Since the scores have come



(a)



(b)

FIGURE 31: Assessors' space for the most and the less unidimensional attribute of Dutch experiment 2, respectively attributes 2 (a) and 9 (b).

from well trained assessors, mean vectors were almost unbiased and variances (and covariances) were naturally small.

In such study, only power could be evaluated, since simulations were under alternative hypothesis $H_1 : \Psi_1 \neq \Psi_2 \neq \ldots \neq \Psi_n$. Practical variance was so tiny that, after 500 drawings, no wrong decision has been made, yielding 100% of power estimate. Figure 32a shows a biplot randomly chosen to give an idea of the distances between products, sizes of variances and the decision made.

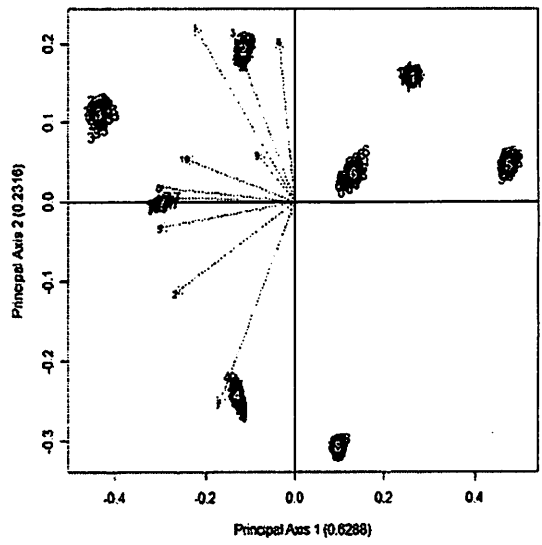
Another interesting aspect to pay attention is the behaviour of the Variance Accounted For (VAF) the components along the 500 replications. Figure 32b displays box plots of the VAF for all components and the sample variance accounted for the estimates. Again, the standard error of that estimator was so small that the box plots looked like bars, but still contain the sample estimate, just like expected.

Future works can contemple higher variances in order to insert uncertainty to the process and induct some type II error rates in simulations under complete H_1. However, it is worth knowing that the proposed method presents high power rates in a good training practical-like context. Such high power rates can, however, induce type I error rate greater than the nominal for cases under H_0.

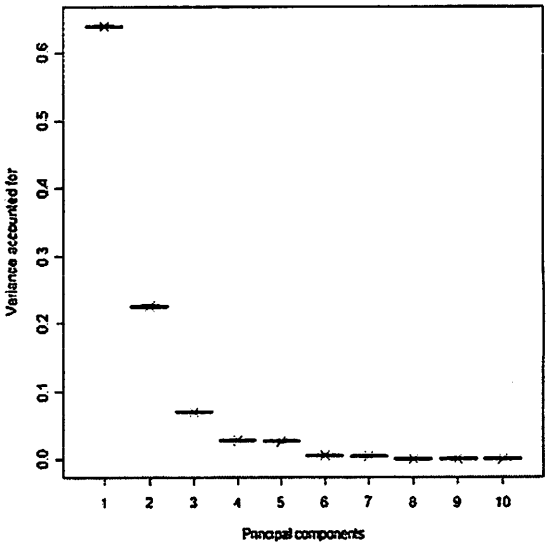
b) Partial H_0 is a usually a tough situation for tests. It is not rare for a test to present type I error rates higher than the nominal. For that reason, the *bootstrap* inference method was submitted to such demanding situation in order to evaluate its behaviour.

Table 7 presents power and type I error rates for both estimation and decision stages (R_1 to R_8) of the described method for the case under partial H_0, for panels of sizes 5, 10, 15, 20, 50 and 90 assessors.

For this simulation study, *bootstrap* hypothesis test revealed to be



(a)



(b)

FIGURE 32: (a) Randomly chosen estimation biplot for illustrating the replication of the simulation process based on Gorgonzola parameters estimates. (b) Box plots for variance accounted for (VAF) for 10 principal components and sample estimate, represented by a $\times$.

Table 7: Power and type I error rates (R_1 to R_8) for both estimation and decision stages, under partial H_0, for virtual panels of sizes 5, 10, 15, 20, 50 and 90.

	Panel size					
Rate	**5**	**10**	**15**	**20**	**50**	**90**
R_1	0.8200	0.8320	0.8700	0.8981	0.9923	1.0000
R_2	0.4660	0.3460	0.2980	0.2731	0.3359	0.3333
R_3	0.5320	0.4800	0.4220	0.4596	0.3359	0.3333
R_4	0.9640	0.9980	1.0000	1.0000	1.0000	1.0000
R_5	0.0150	0.0183	0.0085	0.0238	0.0174	0.0000
R_6	0.0000	0.0154	0.0104	0.0178	0.0174	0.0000
R_7	0.4740	0.3580	0.3040	0.2904	0.3475	0.3333
R_8	0.5320	0.4880	0.4280	0.4692	0.3475	0.3333

conservative[6], since it presented type I error rates (R_5 and R_6) under the nominal level (5%). In addition, only random variations seemed to happen along the sample sizes, what suggest that the test respects the nominal level (and is conservative) even for small sample sizes.

On the other hand, decisions based on the estimation stage revealed to be very liberal[7], that is, power rates are extremely high (R_1 and R_4), easily reaching 100%. For that reason, type I error rates of this same stage are very greater than the nominal level (R_2 and R_3). However, type I error rates are usually greater than the nominal level under partial H_0 (Borges, 2002; Santos, 2000). In addition, power rates increased and type I error rates decreased as the sample size (k) increased, as expected.

Rates R_7 and R_8 consider the whole type I error rate, that is, error in the first and (plus) in the second stage. Note that, though those rates decrease along sample increment, they are still too high at the biggest sample size ($k = 90$). It suggests that the whole inference method is liberal, that is, its type I error rates are greater than its type II error rates. That

[6]**Conservative test:** which presents type I error rate less than the nominal level (α).
[7]**Liberal test:** which presents type I error rate greater than the nominal level (α).

is not a problem itself, since type II error is the worst type or error in most cases. It is due to the fact that two products are never *exactly* equal.

However, some attitudes can try to round such problem. If the confidence regions were set to higher confidence, more tests will be required, reducing the type I error rate. Further studies must be performed associating a kind of protection to the confidence regions. For instance, the Bonferroni protection, where the confidence are set as $1 - \alpha^*$, where $\alpha^* = \alpha / C_{n,2}$.

Associating estimation and testing stages for decision can be quite interesting since the prior seems to be liberal and the second, conservative. Nevertheless, empirical tests (not published) suggest that the *bootstrap* test remains conservative even when performed for pair of products that do not have interception of 95% confidence regions. Therefore, associating estimation and testing stages enables one to reduce the number of possible tests ($C_{n,2}$) and prevents the whole test to be conservative.

In order to evaluate the stability of the process, the variance accounted for principal components, as well as their standard deviations, were computed along the replications for all considered sample sizes (Table 8). Estimates of the variance accounted for were coherent since the smallest sample size. Standard deviation of such estimates tended to decrease as the sample size increased, as expected.

Sample size 20 yielded the more discrepant and variable estimates of the variances accounted for principal components. Such estimates were taken as randomly diverging from the expected values.

Due to small empirical standard deviations and coherent estimates, described method was considered efficient to estimate variances accounted for principal components.

Table 8: Variances accounted for the first five principal components and their standard deviations along the simulated sample sizes.

	Variance Accounted For (VAF)									
k	PC1		PC2		PC3		PC4		PC5	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
5	0.47	0.05	0.24	0.04	0.14	0.02	0.09	0.02	0.05	0.01
10	0.47	0.03	0.23	0.03	0.14	0.02	0.09	0.01	0.06	0.01
15	0.47	0.03	0.23	0.02	0.14	0.01	0.09	0.01	0.07	0.01
20	0.43	0.14	0.20	0.07	0.13	0.04	0.08	0.03	0.06	0.02
50	0.47	0.02	0.22	0.01	0.14	0.01	0.09	0.01	0.07	0.01
90	0.47	0.01	0.23	0.01	0.14	0.01	0.09	0.01	0.08	0.00

3.12 Discussion

Practical data sets were suitably analyzed. Gorgonzola cheeses showed to be well distinguished by the expert panel, both by technology and ripening age. *Bootstrap* hypothesis tests confirmed the products are distinct and more confident management decisions were allowed to be taken. Assessors agreed more about residual flavor of the cheeses and had the poorest performance scoring bitter taste.

Two Dutch experiments of commercial foods were analyzed successfully as well. Experiment 1 distinguished groups of yoghurts, custards and mayonnaises, while experiment 2, with the same assessors and attributes, grouped fresh creams, soft cheese, creams, the Mozzarella and other cheeses. For products evaluated in experiment 1, the 90 assessors scored more unidimensionally attribute airy mouth-feel, while they were less consistent scoring the creamy soft mouth-feel. On the other hand, the same panel had a different performance evaluating products of experiment 2. They agreed more about thickness mouth-feel and had the poorest performance scoring astringent mouth-feel. In both experiments, some assessors can be identified as "outliers" and must be re-trained or removed from the panel.

From simulation studies, described method showed to combine a

liberal first stage (estimation) with a conservative second one (testing), in the decision process. According to need of protecting against type I or type II error rates, one can set the confidence of value of the regions suitably or ever suppress the estimation stage in order to allow more or less tests. Further works must be done about that issue.

Variances accounted for the principal components were precisely estimated, suggesting stability, even in small sample sizes.

One can note a clear behaviour along all sample sizes: high power in both estimation and testing stages; high type I error rates in estimation stage (liberal) and small type I error rates in testing stage (conservative). Such behaviour seems to be slightly improved along the sample size increment. However, a protection against the liberality of the estimation stage might yield better results and keep avoiding testing all possible combinations of products. Front of those results there is not a clear cut-line for the sample size, though the big sample, the better. Well trained assessors (less bias and small variances, as in Gorgonzola experiment) are as important as big sample sizes. However, the method seems to behave well for expert panels and samples sizes greater than 20 to semi-trained people.

For all cases, the best sample *bootstrap* resample unit was the whole matrix. However, when the number of panelists is very small, rows can be treated as resample units in order to increase the number of possible virtual panels, but only in those cases where the variance between assessors is equivalent to the variance within assessors.

Bootstrap GPA inference routines programmed in R language showed up to be very computational demanding. For instance, each replication of the simulation study with 90 assessors, has demanded, in average, 240 hours (in a Pentium III, 512Mb RAM and Windows operational system). Further works must be done in order to improve the efficiency of such routines.

4 FINAL CONSIDERATIONS

After several studies (simulations, algebraic and practical experiments) performed in Procrustes analysis for estimating missing cells and inferring about the population of focus, it is clear that Procrustes problems are a rich field to be explored and can contribute greatly to Sensometrics.

Putative values can lead to local minima in missing values estimation and therefore must be suitably chosen, for instance setting to zero for centered matrices. Algorithm efficiency and more suitable constraints also must be well studied for estimating missing cells. Two-sided eigenvalue problems with both matrices positive semi-definite seems to be a border to be passed for a better development of Procrustes problems with general transformation. As important as such algebraic derivation is the practical interpretability of such transformation, what can contribute even more with sensory experiments.

For the first time an inference procedure is described for Procrustes analysis. However, it has to be discussed and improved both in the estimation/decision process and in the programming efficiency. Optimization techniques must be used to refine the codes and turn the whole process *lighter*. After that, ready-to-use computational packages will be able to help several sensometricians, statisticians and food engineers world wide (preferably in open code languages). At least, (another) first step has been done in order to improve and extend such a classical analysis as Procrustes.

APPENDIX A - PROOFS

(A.1): Leat squares estimator of T under no restriction is $\hat{T} = \left(X_1' X_1\right)^{-1} X_1' X_2$.
Proof: Let

$$X_2 = X_1 T + \varepsilon,$$

where ε is a random vector of residuals.[1]

Then, if ε is distributed according to some multivariate distribution with mean vector Φ (vector of zeros) and covariance matrix $\sigma^2 I$, and letting $X_1 T$ be constant, follows that

$$\begin{aligned} E[X_2] &= E[X_1 T + \varepsilon] \\ &= E[X_1 T] + E[\varepsilon] \\ &= X_1 T. \end{aligned}$$

One estimator of T can be derived by minimizing the residuals sum of squares in relation to $E[X_2]$:

$$\begin{aligned} \|X_2 - E[X_2]\| &= (X_2 - E[X_2])(X_2 - E[X_2])' \\ &= (X_2 - X_1 T)(X_2 - X_1 T)' \\ &= (X_2 - X_1 T)(X_2' - T' X_1') \\ &= X_2 X_2' - X_2 T' X_1' - X_1 T X_2' + X_1 T T' X_1' \\ &= f(T) \end{aligned}$$

[1] In that case, X_1 is a fixed (given) matrix and X_2 is random matrix.

$$\frac{\partial f(T)}{\partial T} = 0 \Rightarrow$$

$$\Rightarrow -X_1'X_2 - X_1'X_2 + 2X_1'X_1\hat{T} = 0 \Rightarrow$$

$$\Rightarrow 2X_1'X_1\hat{T} = 2X_1'X_2 \Rightarrow$$

$$\Rightarrow \hat{T} = \left(X_1'X_1\right)^{-1} X_1'X_2. \qquad \text{(A.1)}$$

(A.2):

$$\sum_{i<i'}^{k} \left\| s_i(X_i - 1u_i')Q_i - s_{i'}(X_{i'} - 1u_{i'}')Q_{i'} \right\| = k \sum_{i=1}^{k} \| s_i(X_i - 1u_i')Q_i - G \|$$

Proof: Let $s_i(X_i - 1u_i')Q_i = X_i^*$ and $s_{i'}(X_{i'} - 1u_{i'}')Q_{i'} = X_{i'}^*$. Since $\sum_{i'=1}^{k} X_{i'}^* = \sum_{i=1}^{k} X_i^*$, then

$$\begin{aligned}
\sum_{i<i'}^{k} \|X_i^* - X_{i'}^*\| &= \frac{\sum_{i=1}^{k}\sum_{i'=1}^{k} \left\|(X_i^* - X_{i'}^*)\right\|}{2} \\
&= \frac{k\sum_{i=1}^{k}\|X_i^*\| - 2\sum_{i=1}^{k}\sum_{i'=1}^{k} tr(X_i^{*'}X_{i'}^*) + k\sum_{i'=1}^{k}\left\|X_{i'}^*\right\|}{2} \\
&= \frac{2k\sum_{i=1}^{k}\|X_i^*\| - 2\sum_{i=1}^{k}\sum_{i'=1}^{k} tr(X_i^{*'}X_{i'}^*)}{2} \\
&= k\sum_{i=1}^{k}\|X_i^*\| - \sum_{i=1}^{k}\sum_{i'=1}^{k} tr(X_i^{*'}X_{i'}^*) \\
&= k\sum_{i=1}^{k}\|X_i^*\| - tr\left(\sum_{i=1}^{k} X_i^* \sum_{i'=1}^{k} X_{i'}^*\right) \\
&= k\sum_{i=1}^{k}\|X_i^*\| - \sum_{i=1}^{k}\|X_i^*\| \\
&= k\left[\sum_{i=1}^{k}\|X_i^*\| - \frac{\sum_{i=1}^{k}\|X_i^*\|}{k}\right] \\
&= k\sum_{i=1}^{k}\left\|X_i^* - \bar{X}^*\right\|
\end{aligned}$$

$$= k \sum_{k=1}^{k} ||s_i(X_i - 1u_i')Q_i - G|| \qquad (A.2)$$

(A.3): $\sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G|| = \left(\frac{k-1}{k}\right)^2 \sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G_i||$

Proof: From (2.5) and (2.7), yields

$$\begin{aligned} G - G_i &= \frac{1}{k}\sum_{i=1}^{k} s_i(X_i - 1u_i')T_i - G_i \\ &= \frac{\sum_{i=1}^{k} s_i(X_i - 1u_i')T_i - kG_i}{k} \\ &= \frac{\sum_{i' \neq i}^{k} s_{i'}(X_{i'} - 1u_{i'}')T_{i'} + s_i(X_i - 1u_i')T_i - kG_i}{k} \\ &= \frac{(k-1)G_i + s_i(X_i - 1u_i')T_i - kG_i}{k} \\ &= \frac{s_i(X_i - 1u_i')T_i - G_i}{k}. \end{aligned}$$

Using such result,

$$\begin{aligned} \sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G|| &= \sum_{i=1}^{k} ||[s_i(X_i - 1u_i')T_i - G_i] + (G_i - G)|| \\ &= \sum_{i=1}^{k} ||[s_i(X_i - 1u_i')T_i - G_i] - (G - G_i)|| \\ &= \sum_{i=1}^{k} \left|\left|[s_i(X_i - 1u_i')T_i - G_i] - \left[\frac{s_i(X_i - 1u_i')T_i - G_i}{k}\right]\right|\right| \\ &= \sum_{i=1}^{k} \left|\left|\left(1 - \frac{1}{k}\right)[s_i(X_i - 1u_i')T_i - G_i]\right|\right| \\ &= \left(1 - \frac{1}{k}\right)^2 \sum_{i=1}^{k} ||s_i(X_i - 1u_i')T_i - G_i||. \qquad (A.3) \end{aligned}$$

(A.4): $||(X_i - X_{ui})T_i - G_i|| = ||X_{ui}T_i - (X_iT_i - G_i)||$. *Proof*:

$$\begin{aligned} ||(X_i - X_{ui})T_i - G_i|| &= ||X_iT_i - X_{ui}T_i - G_i|| \\ &= ||X_{ui}T_i - X_iT_i + G_i|| \\ &= ||X_{ui}T_i - (X_iT_i - G_i)|| \end{aligned} \tag{A.4}$$

(A.5): $||X_{ui}T_i - (X_iT_i - G_i)|| \propto tr[(X'_{ui}X_{ui})(T_iT'_i) - 2X'_{ui}(X_iT_i - G_i)T'_i]$.

Proof:

$$\begin{aligned} &||X_{ui}T_i - (X_iT_i - G_i)|| = ||T'_iX'_{ui} - T'_iX'_i + G'_i|| \\ &= tr[(X_{ui}T_i - X_iT_i + G_i)(T'_iX'_{ui} - T'_iX'_i + G'_i) \\ &= tr[X_{ui}T_iT'_iX'_{ui} - X_{ui}T_iT'_iX'_i + X_{ui}T_iG'_i - X_iT_iT'_iX'_{ui} + X_iT_iT'_iX'_i \\ &- X_iT_iG'_i + G_iT'_iX'_{ui} - G_iT'_iX'_i + G_iG'_i] \\ &\propto tr[X_{ui}T_iT'_iX'_{ui} - X_{ui}T_iT'_iX'_i + X_{ui}T_iG'_i - X_iT_iT'_iX'_{ui} + G_iT'_iX'_{ui}] \\ &\propto tr[X_{ui}T_iT'_iX'_{ui} - 2X_iT_iT'_iX'_{ui} + 2G_iT'_iX'_{ui}] \\ &\propto tr[(X'_{ui}X_{ui})T_iT'_i - 2X'_{ui}X_iT_iT'_i + 2X'_{ui}G_iT'_i] \\ &\propto tr[(X'_{ui}X_{ui})T_iT'_i - 2X'_{ui}(X_iT_i - G_i)T'_i] \end{aligned} \tag{A.5}$$

(A.8): Minimization of $tr[(X'_{ui}X_{ui})T - 2X'_{ui}Y]$ over x_r, subject to $X_{ui} = \sum_{m=1}^{M} x_m e_{i_m} e'_{j_m}$ yields $y_r = \sum_{m=1}^{M} x_m (e'_{i_r} e_{i_m})(t_{j_r, j_m})$, $r = 1, \ldots, M$.

Proof: First, let $X_{ui}T = Z$.

$$\begin{array}{c} \\ i_r \end{array} \begin{array}{c} \begin{array}{cccc} & & j_r & \end{array} \\ \begin{pmatrix} \sum_{i=1}^{p} x_{1i}t_{i1} & \cdots & \sum_{i=1}^{p} x_{1i}t_{ip} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{p} x_{ni}t_{i1} & \cdots & \sum_{i=1}^{p} x_{ni}t_{ip} \end{pmatrix} \end{array} = Z$$

So,

$$tr(X'_{ui}X_{ui}T) = tr(X'_{ui}Z) = x'_{ui}z =$$

$$= \begin{pmatrix} x_{i_1j_1} & x_{i_2j_2} & \dots & x_{i_Mj_M} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} z_{i_1j_1} \\ z_{i_2j_2} \\ \vdots \\ z_{i_Mj_M} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= x_{i_1j_1}z_{i_1j_1} + x_{i_2j_2}z_{i_2j_2} + \ldots + x_{i_Mj_M}z_{i_Mj_M} = \sum_{m=1}^{M} x_{i_mj_m}z_{i_mj_m}.$$

Since,

$$\begin{aligned}
\frac{\partial}{\partial x_r}(x_{i_rj_r}z_{i_rj_r}) &= \frac{\partial}{\partial x_r}(x_{i_rj_r}\sum_{p=1}^{p} x_{i_rp}t_{j_pj_r}) \\
&= \frac{\partial}{\partial x_r}(x_{i_rj_r}x_{i_r1}t_{j_1j_r} + \ldots + x^2_{i_rj_r}t_{j_rj_r} + \ldots + x_{i_rj_r}x_{i_rp}t_{j_pj_r}) \\
&= x_{i_r1}t_{j_1j_r} + \ldots + 2x_{i_rj_r}t_{j_rj_r} + \ldots + x_{i_rp}t_{j_pj_r} \\
&= \sum_{p=1}^{p} x_{i_rp}t_{j_pj_r} + x_{i_rj_r}t_{j_rj_r}
\end{aligned}$$

and since $\sum_{m=1}^{M} x_{i_mj_m}z_{i_mj_m} = \sum_{m=1}^{M} x_{i_rj_m}z_{i_rj_m}$ for the missing cells that are in the same row of $x_{i_rj_r}$, for each m holds,

$$x_{i_rj_m}z_{i_rj_m} = x_{i_rj_m}\sum_{p=1}^{p} x_{i_rj_p}t_{j_pj_m}$$

$$j_p = j_r$$

$$\frac{\partial}{\partial x_r}(x_{i_rj_m}x_{i_rj_r}t_{j_rj_m}) = x_{i_rj_m}t_{j_rj_m}.$$

Therefore,

$$\frac{\partial}{\partial x_r}\sum_{m=1}^{M} x_{i_m j_m} z_{i_m j_m} = \sum_{p=1}^{p} x_{i_r p} t_{j_p j_r} + x_{i_r j_r} t_{j_r j_r} + \sum_{m \neq r}^{M} x_{i_r j_m} t_{j_r j_m}.$$

Since T is symmetric, that is, $t_{j_m j_r} = t_{j_r j_m}$, and using the indicator function $(e'_{i_r} e_{i_m})$, yields

$$\frac{\partial}{\partial x_r} tr(X'_{ui} X_{ui} T) = \frac{\partial}{\partial x_r}\sum_{m=1}^{M} x_{i_m j_m} z_{i_m j_m} = 2\sum_{m=1}^{M} x_{i_r j_m}(e'_{i_r} e_{i_m}) t_{j_r j_m}. \quad \text{(A.6)}$$

On the other hand,

$$2tr(X'_{ui} Y) = 2x'_{ui} y =$$

$$= 2\begin{pmatrix} x_{i_1 j_1} & x_{i_2 j_2} & \cdots & x_{i_M j_M} & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} y_{i_1 j_1} \\ y_{i_2 j_2} \\ \vdots \\ y_{i_M j_M} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= 2\sum_{m=1}^{M} x_{i_m j_m} y_{i_m j_m}.$$

Then, the derivative in relation to $x_{i_r j_r}$, or just x_r is

$$\frac{\partial}{\partial x_r} 2tr(X'_{ui} Y) = \frac{\partial}{\partial x_r} 2\sum_{m=1}^{M} x_{i_m j_m} y_{i_m j_m} = 2y_{i_r j_r}.x_{i_r j_m}(e'_{i_r} e_{i_m}) t_{j_r j_m}. \quad \text{(A.7)}$$

Finally, from (A.6) and (A.7), yields

$$\frac{\partial}{\partial x_r} tr(X'_{ui} X_{ui} T - X'_{ui} Y) = 0$$

$$2y_{i_r j_r} = 2 \sum_{m=1}^{M} x_{i_r j_m} (e'_{i_r} e_{i_m}) t_{j_r j_m}$$

$$y_{i_r j_r} = \sum_{m=1}^{M} x_{i_r j_m} (e'_{i_r} e_{i_m}) t_{j_r j_m}, \quad \text{or just,}$$

$$y_r = \sum_{m=1}^{M} x_m (e'_{i_r} e_{i_m}) t_{j_r j_m}. \tag{A.8}$$

(A.9): The least squares estimator of (2.18) is $\hat{X}^r_{ui} = YT^{-1}$.

Proof:

$$f(X^r_{ui}) = tr[(X^{r'}_{ui} X^r_{ui})T - 2X^{r'}_{ui} Y]$$

$$\frac{\partial f(X^r_{ui})}{\partial X^r_{ui}} = \frac{\partial (X^{r'}_{ui} X^r_{ui})T}{\partial X^r_{ui}} - 2\frac{\partial X^{r'}_{ui} Y}{\partial X^r_{ui}} = 0$$

$$2X^r_{ui} T - 2Y = 0$$

$$\hat{X}^r_{ui} = YT^{-1} \tag{A.9}$$

(A.10):

Minimizing $||s_i(X_{ci} + X^r_{ui})T_i - G_i||$ is minimize $tr[(X^{r'}_{ui} X^r_{ui})T - 2X^{r'}_{ui} Y]$.

Proof:

$$||s_i(X_{ci} + X^r_{ui})T_i - G_i|| =$$

$$= ||s_i X_{ci} T_i + s_i X^r_{ui} T_i - G_i||$$

$$= ||s_i X^r_{ui} T_i - (G_i - s_i X_{ci} T_i)||$$

$$= tr[(s_i X^r_{ui} T_i - (G_i - s_i X_{ci} T_i))'(s_i X^r_{ui} T_i - (G_i - s_i X_{ci} T_i))]$$

$$= tr[(s_i T'_i X^{r'}_{ui} - G'_i + s_i T'_i X'_{ci})(s_i X^r_{ui} T_i - G_i + s_i X_{ci} T_i)]$$

$$= tr[s_i^2 T_i' X_{ui}^{r'} X_{ui}^r T_i - s_i T_i' X_{ui}^{r'} G_i + s_i^2 T_i' X_{ui}^{r'} X_{ci} T_i - s_i G_i' X_{ui}^r T_i +$$
$$+ G_i' G_i - s_i G_i' X_{ci} T_i + s_i^2 T_i' X_{ci}' X_{ui}^r T_i - s_i T_i' X_{ci}' G_i + s_i^2 T_i' X_{ci}' X_{ci} T_i]$$

From now on, lets consider only those term that depend on X_{ui}^r, since minimizing them is minimizing the whole expression.

$$\propto tr[s_i^2 T_i' X_{ui}^{r'} X_{ui}^r T_i - s_i T_i' X_{ui}^{r'} G_i + s_i^2 T_i' X_{ui}^{r'} X_{ci} T_i - s_i G_i' X_{ui}^r T_i + s_i^2 T_i' X_{ci}' X_{ui}^r T_i]$$
$$= tr[s_i^2 X_{ui}^{r'} X_{ui}^r T_i T_i' + 2 s_i^2 X_{ui}^{r'} X_{ci} T_i T_i' - 2 s_i X_{ui}^{r'} G_i T_i']$$
$$= tr[s_i^2 (X_{ui}^{r'} X_{ui}^r) T_i T_i' - 2 s_i X_{ui}^{r'} (G_i - s_i X_{ci} T_i) T_i']$$
$$= tr[(X_{ui}^{r'} X_{ui}^r) T_i T_i' - 2 X_{ui}^{r'} (s_i^{-1} G_i - X_{ci} T_i) T_i']$$
$$= tr[(X_{ui}^{r'} X_{ui}^r) T - 2 X_{ui}^{r'} Y]. \qquad \text{(A.10)}$$

(A.11): $\hat{T}_i = \frac{1}{s_i} (X_{ci}' X_{ci})^{-1} X_{ci}' G_i$.

Proof:

$$f(T_i) = ||s_i X_{ci} T_i - G_i||$$
$$= tr\left[(s_i X_{ci} T_i - G_i)'(s_i X_{ci} T_i - G_i)\right]$$
$$= tr\left[(s_i T_i' X_{ci}' - G_i')(s_i X_{ci} T_i - G_i)\right]$$
$$= tr\left[s_i^2 T_i' X_{ci}' X_{ci} T_i - s_i T_i' X_{ci}' G_i - s_i G_i' X_{ci} T_i - G_i' G_i\right]$$
$$= tr\left[s_i^2 T_i' X_{ci}' X_{ci} T_i - 2 s_i T_i' X_{ci}' G_i - G_i' G_i\right]$$

$$\frac{\partial f(T_i)}{\partial T_i} = 0$$
$$-2 s_i X_{ci}' G_i + 2 s_i^2 X_{ci}' X_{ci} T_i = 0$$
$$s_i X_{ci}' G_i = s_i^2 X_{ci}' X_{ci} T_i$$
$$\hat{T}_i = \frac{1}{s_i} (X_{ci}' X_{ci})^{-1} X_{ci}' G_i \qquad \text{(A.11)}$$

(A.12): $\sum_{i<i'}^{k} ||s_i X_{ci}^* T_i - s_i' X_{ci'}^* T_i'|| = s'(kDiag(\mathbf{S}) - \mathbf{S})s$

Proof:

$$\sum_{i<i'}^{k} ||s_i X_{ci}^* T_i - s_{i'} X_{ci'}^* T_{i'}|| =$$

$$= k \sum_{i=1}^{k} ||s_i X_{ci}^* T_i - G||$$

$$= k \sum_{i=1}^{k} tr(s_i^2 T_i' X_{ci}'^* X_{ci}^* T_i - 2 s_i T_i' X_{ci}'^* G + G'G)$$

$$= k \sum_{i=1}^{k} s_i^2 tr(T_i' X_{ci}'^* X_{ci}^* T_i) - 2k \sum_{i=1}^{k} s_i tr(T_i' X_{ci}'^* G) + k \sum_{i=1}^{k} tr(G'G)$$

$$= k \sum_{i=1}^{k} s_i^2 tr(T_i' X_{ci}'^* X_{ci}^* T_i) - 2 \sum_{i=1}^{k} s_i tr\left(T_i' X_{ci}'^* \sum_{i'=1}^{k} s_{i'} X_{ci'}^* T_{i'} \right) +$$

$$+ tr\left[\left(\sum_{i'=1}^{k} s_{i'} X_{ci'}^* T_{i'} \right)' \left(\sum_{i'=1}^{k} s_{i'} X_{ci'}^* T_{i'} \right) \right]$$

$$= k \sum_{i=1}^{k} s_i^2 tr(T_i' X_{ci}'^* X_{ci}^* T_i) - 2 \sum_{i=1}^{k} \sum_{i'=1}^{k} s_i s_{i'} tr(T_i' X_{ci}'^* X_{ci'}^* T_{i'}) +$$

$$+ \sum_{i=1}^{k} \sum_{i'=1}^{k} s_i s_{i'} tr(T_{i'}' X_{ci'}'^* X_{ci'}^* T_{i'})$$

$$= k \sum_{i=1}^{k} s_i^2 tr(T_i' X_{ci}'^* X_{ci}^* T_i) - \sum_{i=1}^{k} \sum_{i'=1}^{k} s_i s_{i'} tr(T_i' X_{ci}'^* X_{ci'}^* T_{i'})$$

$$= \sum_{i=1}^{k} [k s_i^2 tr(T_i' X_{ci}'^* X_{ci}^* T_i) - \sum_{i'=1}^{k} s_i s_{i'} tr(T_i' X_{ci}'^* X_{ci'}^* T_{i'})]$$

$$= \sum_{i=1}^{k} [(k-1) s_i^2 tr(T_i' X_{ci}'^* X_{ci}^* T_i) - \sum_{i' \neq i}^{k} s_i s_{i'} tr(T_i' X_{ci}'^* X_{ci'}^* T_{i'})]$$

$$= s_1^2(k-1)tr(T_1'X_{c1}'^*X_{c1}^*T_1) - \sum_{i'\neq 1}^{k} s_1 s_{i'} tr(T_1'X_{c1}'^*X_{ci'}^*T_{i'}) + \ldots$$

$$+ s_k^2(k-1)tr(T_k'X_{ck}'^*X_{ck}^*T_k) - \sum_{i'\neq k}^{k} s_k s_{i'} tr(T_k'X_{ck}'^*X_{ck}^*T_{i'})$$

$$= s'(kDiag(\mathbf{S}) - \mathbf{S})s \qquad \text{(A.12)}$$

(A.13): $\dfrac{(x-\bar{x})^2}{det(\mathbf{C})/s_y^2} + \dfrac{(y-\bar{y})^2}{det(\mathbf{C})/s_x^2} - 2\dfrac{(x-\bar{x})(y-\bar{y})}{det(\mathbf{C})/s_{xy}} \leq \chi_\alpha^2(2)$

Proof:

$$(\mathbf{z}-\mu)'\Sigma^{-1}(\mathbf{z}-\mu) \sim \chi^2(2)$$

$$\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}' \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} \sim \chi_\alpha^2(2)$$

$$\frac{1}{det(\Sigma)} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}' \begin{pmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{pmatrix} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} \sim \chi_\alpha^2(2)$$

$$\frac{1}{det(\mathbf{C})} \begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix}' \begin{pmatrix} s_y^2 & -s_{xy} \\ -s_{xy} & s_x^2 \end{pmatrix} \begin{pmatrix} x-\bar{x} \\ y-\bar{y} \end{pmatrix} \dot{\sim} \chi_\alpha^2(2)$$

$$\frac{1}{det(\mathbf{C})}(x-\bar{x})^2 s_y^2 + (y-\bar{y})^2 s_x^2 - 2(x-\bar{x})(y-\bar{y})s_{xy} \leq \chi_\alpha^2(2)$$

$$\frac{(x-\bar{x})^2}{det(\mathbf{C})/s_y^2} + \frac{(y-\bar{y})^2}{det(\mathbf{C})/s_x^2} - 2\frac{(x-\bar{x})(y-\bar{y})}{det(\mathbf{C})/s_{xy}} \leq \chi_\alpha^2(2) \qquad \text{(A.13)}$$

APPENDIX B – R FUNCTIONS

Main R functions used along the thesis are briefly described below. R codes are not published because the need of optimization in performance, but they will be turn public as soon as such improvements occur.

- *bootGPA.cr (X, ortho=TRUE, sf=TRUE, resmpl=100, conf=0.95)*

  Performs generalised (or ordinary) Procrustes analysis and builds $1-\alpha$ *bootstrap* confidence regions for all evaluated objects. It allows orthogonal or general transformations, with or without scaling and enables the user to set the resample size and the confidence of the region.

- *bootGPA.tst (X, a, b, resmpl=500, plot.result=TRUE)*

  Performs generalised (or ordinary) Procrustes analysis and a α *bootstrap* test between two products a and b ($H_0 : a = b$). It allows an empirical histogram to be constructed and the empirical p-value is shown on the console output.

- *missingGPA (X, ortho = TRUE, sf = TRUE, tol = 1e-4, plot.result = FALSE, plot.iterat = FALSE)*

  Performs iteratively generalised (or ordinary) Procrustes adjustment and missing cells estimation (estimation/minimisation). It allows orthogonal or general transformations, with or without scaling, enable

the user to set the convergence tolerance, and plot either the GPA biplot or the residual sum of squares convergence through iterations.

- *hist.biv (a, boot = TRUE, level = .95, ellip = FALSE, point = FALSE, pty = 20, xn=c("X"), yn=c("Y"))*

  Builds bivariate histograms for data sets form both *bootstrapping* or not. The confidence can be set. An ellipse can be draw on the plot if desired. The original points can be draw or not and the character can be chosen, as well as the labels of the axes.

- *bootGPA (X, attrib = c(seq(1:dim(X)[2])), ortho = TRUE, sf = TRUE, resmpl = 100, conf = 0.95, rows = FALSE, plot.result = TRUE)*

  Performs generalised (or ordinary) Procrustes analysis and the whole described inference procedure. It allows orthogonal or general transformations, with or without scaling and enables the user to set the resample size, the confidence of the region, if rows must be the resample units, and draws the final biplot with original attributes.

- *p.cons (X, attrib=seq(1:dim(X)[2]), plot.pr=FALSE, plot.as=FALSE, scree=FALSE, bar.C=TRUE)*

  Performs Dijksterhuis' panel consonance test. It plots either product space, assessors space, scree plot or the bar plot of statistic C, with original attribute labels.

REFERENCES

AGUIAR, R. L. **Produção e caracterização de massa de pizza pré-assada e congelada.** 1995. 88 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Federal de Viçosa, Viçosa.

AKCA, M. D. **Generalized procrustes analysis and its applications in photogrammetry.** 2003. 23 p. Zurich: Swiss Federal Institute of Technology, Institute of Geodesy and Photogrammetry, ETH-Hoenggerberg. Disponível em: <http://e-collection.ethbib.ethz.ch/ecol-pool/bericht/bericht_363.pdf>. Acesso em: 28 out. 2004.

ALLCROFT, D. J.; GLASBEY, C.A.; PAULO, M. J. A latent gaussian model for multivariate consumption data. **Food Quality and Preference**, Cidade, v. 3, n. 18, p. 508-516, 2007.

AMERINE, M. A.; PANGBORN, R. M.; ROESSLER, E. B. **Principles of sensory evaluation of food.** 1. ed. San Francisco: Academic Press, 1965. 602 p.

ANZALDUA-MORALES, A. **La evaluación sensorial de los alimentos en la teoría y la práctica.** Zaragoza: Editorial Acribia S.A., 1994. 198 p.

AQUINO, A. A. **Aproveitamento do tegumento da semente do cacau em ração para coelhos em crescimento.** 1992. 59p. Dissertação (Mestrado em Zootecnia) - Universidade Federal de Viçosa, Viçosa.

ASTRUC, C.; BLUMENTHAL, D.; DELARUE, J.; DANZART, M.; SIEFFERMANN, J. M. How to construct a global driving procedure for dynamic hedonic tests? In: THE SENSOMETRICS MEETING, 8., 2006, Ås, Norway. **Proceedings...**, 2006. p. 24.

BORGES, L.C. **Poder e taxas de erro tipo I dos testes Scott-Knott, Tukey e Student-Newman-Keuls sob distribuições Normal e não Normais dos resíduos.** 2002. 94 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

CARDOSO, R. L. **Estabilidade de geléia de jambo-vermelho (Eugenia malaccensis, *Lin*) em copo de vidro.** 1994. 157 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Estadual de Campinas, Campinas.

CHAGAS, S. J. R. **Caracterização química e qualitativa de café de alguns municípios de três regiões produtoras de Minas Gerais.** 1994. 83 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras.

CHREA, C.; VALENTIN, D.; SULMONT-ROSSÉ, C.; LY MAI, H.; HOANG NGUYEN, D.; ABDI, H. Culture and odor categorization: agreement between cultures depends upon the odors. **Food Quality and Preference**, Cidade, v. 7-8, n. 15. p. 669-679, 2004.

CLIFF, N. Orthogonal rotation to congruence. **Psychometrika**, Cidade, v. 31, p. 33-42, 1966.

COMMANDEUR, J. J. F. **Matching configurations**. Leiden: DSWO Press, 1991. 171 p.

DACHS, J. N. W. **Estatística computacional: uma introdução em Turbo Pascal**. Rio de Janeiro: LTC, 1988. paginas p.

DAEMS, V.; DELVAUX, F. Multivariate analysis of descriptive sensory data on 40 commercial beers. **Food Quality and Preference**, Cidade, v. 5-6, n. 8, p. 373-380, 1997.

DAIROU, V.; SIEFFERMANN, J.; PRIEZ, A.; DANZART, M. **Sensory evaluation of car brake systems - the use of flash profile as a preliminary study before a convention profile.** Disponível em: <http://www.perception-sensorielle.com/Congress2003/Dairou_SAE.pdf>. Acesso em: 2 ago. 2004.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood from incomplete data via the em algorithm (with discussion). **Journal of Royal Statistical Society B**, v. 39, p. 1-38, 1977.

DIJKSTERHUIS, G. Assessing panel consonance. **Food Quality and Preference**, v. 6, p. 7-14, 1995.

DRYDEN, I. L.; MARDIA, K. V. **Statistical Shape Analysis.** Chichester: John Wiley, 1998. 376 p.

EFRON, B. Bootstrap methods: Another look at the jackknife. **Annals of Statistics**, v. 7, p. 1-26, 1979.

EFRON, B.; TIBSHIRANI, R. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993. 311 p.

FERREIRA, E. B. **Análise Generalizada de Procrustes via R: Uma aplicação em laticínios**. 2004. 216 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

FERREIRA, E.B.; MAGALHÃES, F.A.R.; OLIVEIRA, M.S.; FERREIRA, D.F. Sensory profile of the gorgonzola type cheese through bootstrap confidence regions in generalized procrustes analysis. **Revista do Instituto de Laticínios Cândido Tostes**, v. 62, p. 535-540, 2007.

GOWER, J. C. Statistical methods for comparing different multivariate analysis of the same data. In: HODSON, J.R.; KENDALL, D.G.; TAUTU, P. **Mathematics in the Archaeological and Historical Sciences.** Edinburgh: Edinburgh University Press, 1971. p. 138-149.

GOWER, J. C. Generalised procrustes analysis. **Psychometrika**, v. 1, n. 40, p. 33-51, 1975.

GOWER, J. C.; DIJKSTERHUIS, G. B. **Procrustes Problems**. Oxford: Oxford University Press, 2004. 256 p.

GREEN, B. F. The orthogonal approximation of an oblique structure in factor analysis. **Psychometrika**, v. 17, p. 429-440, 1952.

HALL, P. **The bootstrap and edgeworth expansions.** New York: Springer-Verlag, 1992. paginas.

HEYMANN, H. Sensory science and sensometrics: yesterday, today and tomorrow. In: THE SENSOMETRICS MEETING, 8., 2006, Ås, Norway. **Proceedings...** 2006. p. 2.

HUMMER, S. Application of multivariate analysis of variance. **Food Quality and Preference**, v. 3, n. 9, p. 83-85, 1998.

HURLEY, J. R.; CATTLE, R. B. The procrustes program: producing direct rotation to test a hypothesized factor structure. **Behavioral Science**, v. 7, p. 258-262, 1962.

HUSSON, F.; LÊ, S.; PAGÈS, J. Confidence ellipse for the sensory profiles obtained by principal component analysis. **Food Quality and Preference**, v. 16, p. 245-250, 2005.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** 4. ed. New Jersey: Prentice Hall, 1998. 816 p.

JORGE, N. **Curvas de secagem de cogumelo (*Agaricus bisporus*) e avaliação sensorial do produto após diferentes períodos de armazenagem.** 1989. 80 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Federal de Viçosa, Viçosa.

KRISTOF, W.; WINGERSKY, B. Generalization of the orthogonal procrustes rotation procedure for more than two matrices. **Proceedings of the 79th Annual Convention of the American Psychological Association**, v. 6, p. 89-90, 1971.

LELE, S. R.; RICHTSMEIER, J. T. **An invariant approach to the Statistical Analysis of Shape.** Boca Raton: Chapman and Hall, 2001. paginas.

LINGOES, J. C.; BORG, I. Alternative measures of fit for the schönemann-carroll matrix fitting algorithm. **Psychometrika**, v. 43, p. 423-427, 1978.

MACDOUGALL, D. B. Coloring of food, drugs, and cosmetics. **Food Quality and Preference**, v. 3, n. 11, p. 255, 2000.

MAGALHÃES, F. A. R. **Estudo sobre a evolução de parâmetros físicoquímicos e sensoriais durante a maturação do queijo tipo gorgonzola.** 2002. 88 p. Tese (Doutorado em Ciência dos Alimentos) - Universidade Federal de Lavras, Lavras.

MALVINO MADRID, R. M. **Efeito da salinidade da água de mar e da alimentação na composição de solutos musculares e de algumas propriedades sensoriais do camarão de água doce, Macrobrachium rosenbergii (De Man, 1879) (*Crustacea Palaemonidae*).** 1994. 191 p. Dissertação (Mestrado em Ciência dos Alimentos) - Universidade Estadual de Campinas, Campinas.

MANLY, B. F. J. **Randomization, bootstrap and Monte Carlo methods in biology**. 2. ed, London: Chapman & Hall, 1998. 399 p.

MIRANDA, V. F. L. **Avaliação Monte Carlo de testes assintóticos e de Bootstrap para autocorrelação residual.** 2004. 77 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

MIRONOV, A.; PEVZNER, P.; GELFAND, M. **Procrustes: Gene recognition via spliced alignment.** 2004. Disponível em: <http://www.hto.usc.edu/software/procrustes/index.html#salign>. Acesso em: 29 fev. 2004.

MORAES, D. A. **A morfometria geométrica e a 'revolução na morfometria': localizando e visualizando mudanças na forma dos organismo.** 2004. Disponível em: <http://www.bioletim.hpg.ig.com.br/III-3/Artigos/moraes.htm>. Acesso em: 17 fev. 2004.

MUELLER, C. G. **Psicologia sensorial.** Rio de Janeiro: Zagar Editores, 1966. 177 p.

NORONHA, J. F. **Apontamentos de análise sensorial.** Coimbra: ESAC, 2003. Disponível em: <http://pwp.netcabo.pt/0331491102/A.S/Apontamentos%5

Csebenta_v_1_0.pdf>. Acesso em: 2 ago. 2004.

O'MAHONY, M. **Sensory evaluation of food.** New York: Marcel Dekker, Ico., 1985. 487 p.

PEAY, E. R. Multidimensional rotation and scaling of configurations to optimal agreement. **Psychometrika**, v. 53, p. 199-208, 1988.

R DEVELOPMENT CORE TEAM, version 2.6.0: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. 2007.

RANGARAJAN, A.; CHUI, H.; BOOKSTEIN, F. L. **The softassign matching algorithm.** 2003. Disponível em: <http://citeseer.nj.nec.com/rangarajan97softassign.html>. Acesso em: 18 dez. 2004.

RODRIGUES, L. A.; SANTOS, V. F. **Morfometria geométrica - aplicações em paleobiologia de Dinossauros.** 2004. Disponível em: <http://www.mnhn.ul.pt/dinos/public_html/Luis_A_Rodrigues/

ABstractCONGGEo.htm>. Acesso em: 17 fev. 2004.

SANTOS, C. dos. **Novas alternativas de testes de agrupamento avaliadas por meio de simulação Monte Carlo.** 2000. 85 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

SCHÖNEMANN, P. H. A generalized solution of the orthogonal procrustes problem. **Psychometrika**, v. 1, n. 31, p. 1-10, 1966.

SCHÖNEMANN, P. H.; CARROL, R. M. Fitting one matrix to another under choice of a central dilation and a rigid motion. **Psychometrika**, v. 3, n. 35, p. 245-256, 1970.

SINESIO, F.; MONETA, E.; SABA, A. Comparison of multivariate methods of analysis to evaluate panelists' performance. **Food Quality and Preference**, v. 4, n. 3, p. 201-208, 1991-1992.

TEN BERGE, J. M. F. Orthogonal procrustes rotation for two or more matrices. **Psychometrika**, v. 42, p. 267-276, 1977.

TEN BERGE, J. M. F.; KNOL, D. L. Orthogonal rotations to maximal agreement for two or more matrices of different column orders. **Psychometrika**, v. 49, p. 49-55, 1984.

THYBO, A. K.; MARTENS, M. Analysis of sensory assessors in texture profiling of potatoes by multivariate modeling. **Food Quality and Preference**, v. 4, n. 11, p. 283-288, 2000.

WILLIAMS, A. A.; LANGRON, S. P. The use of free choice profiling for the evaluation of commercial ports. **Journal of Science and Food Agriculture**, v. 5, n. 35, p. 558-568, 1984.

WORTEL, V. A. L.; WIECHERS, J. W. Skin sensory performance of individual personal care ingredients and marketed personal care products. **Food Quality and Preference**, v. 1-2, n. 11, p. 121-127, 2000.