doi: http://dx.doi.org/10.5892/ruvrv.2012.102.128135

Comparison between packages *Shapes* and *Genstat*® in ordinary procrustes analysis for sensory profiling of food products

Eric Batista FERREIRA[1]
Marcelo Silva de OLIVEIRA[2]

[1]Professor do Instituto de Ciências Exatas. Universidade Federal de Alfenas. eric.ferreira@unifal-mg.edu.br.
[2]Professor do Departamento de Ciências Exatas. Universidade Federal de Lavras. marcelo.oliveira@dex.ufla.br.

Recebido em: 16/10/2012 - Aprovado em: 12/12/2012 - Disponibilizado em: 30/12/2012

ABSTRACT: In modern management of companies, the *quantitative* knowledge is determinant for defining marketing strategies and product images. Particularly for industries of products that strongly stimulate the human senses, the statistical methods that allow quantitative evaluations are joined under the science chunk called Sensometrics. One of those methods is the Ordinary or Classical Procrustes Analysis (OPA), that consists in finding the best way one matrix can be fitted to a reference matrix by minimizing the sum of squared distances. Since this methodology begun to be use in sensory evaluation many specific and non-specific software have been developed. The high costs involved to purchase such programs lead some researchers to look for free software. R is a free statistical and mathematical package that already has an implementation of OPA and GPA (Generalized Procrustes Analysis) in its *Shapes* package (a statistical shape analysis package). The aim of this paper is to interpret, describe and compare R output to a non-specific software, GenStat®**,** normally used in sensory analysis. Under simulated data the software were compared and lead to same conclusions and interpretations. Therefore, package *Shapes* from R can be used in sensory analysis, rather than Statistical Shape analysis.

Key words: Ordinary Procrustes Analysis, software R, data simulation, Food products.

INTRODUCTION

In the modern management of companies, the *quantitative* knowledge of the consumer preferences is determinant for defining marketing strategies and product images, impacting also in the definition of competitive advantage by production (SLACK; CHAMBERS; JOHNSTON, 1997). Further, knowing such preferences is strictly related to the implemented programs and politics, aims and objectives of the quality management of the company.

Particularly for industries of products that strongly stimulate the human senses (food, cosmetics and perfumes industry, among others), the statistical methods that allow quantitative evaluations are joint under the science chunk called Sensometrics, which deals with several categories of problems like consumer preference, sensory profile of a product and assessors training (FERREIRA; OLIVEIRA, 2007).

One of the greatest needs of Food Sensorial Analysis, especially in dairy industries, is a satisfactory statistical analysis of the panelist's scores. For these reason, along the years, a lot of tools have been tested. The multivariate techniques had been demonstrated to be efficient to provide good answers about consumer's preference and food quality (FERREIRA et al., 2008).

This paper deals with the particular problem of identifying the possible consensus of a trained assessor and an expert. It is reasonable to assume that if the assessor agrees with the expert, his/her training can be considered satisfactory. Cheese sensory data is used here for such evaluation.

One of the suitable multivariate techniques for such evaluation is the Procrustes Analysis. Particularly, the Ordinary Procrustes Analysis (OPA) (or Classical Procrustes Analysis) is a multivariate technique that is used when one want to find the best way one matrix can fit another. For instance, how similar one score matrix is to another. While the OPA allow to match just two score matrices, with Generalized Procrustes Analysis (GPA) we can match m matrices at the same time (GOWER, 1975).

The minimization problem of transforming one given matrix X_1 by a matrix, say Q, such that best fits a given target matrix X_2 is called a Procrustes Problem since Procrustes was an evil personage of the Greek Mythology which used to fit his guests in a so called magical bed by sawing or stretching them. The term Procrustes Problem is due to Hurley and Cattel (1962) that suggested for the first time the problem of transforming one matrix into another by minimizing the Residual Sum of Squares (RSS)

$$||sX_1Q - X_2||. \qquad (1)$$

Here, Q refers to a orthogonal matrix (SCHÖENEMANN, 1966), and s is a scaling factor that enables stretching or shrinking the configuration (GOWER, 1975).

In dairy industry is natural to suppose that an expert panelist train a beginner. In order to verify the efficiency of the training an OPA must be performed (fitting the trainee score matrix to the expert score matrix) to evaluate the residual sum of squared distances between these two matrices. Also, agreements and disagreements between the two panelists can be identified in some two axes graphics (FERREIRA et al., 2007).

There are a lot of specific and non-specific software that perform this analysis satisfactorily, but all of the commonly used are so expensive. R is a free statistical and mathematical program. It has some basic packages (default) and many others packages to be loaded in accordance with the user needs (R DEVELOPMENT CORE TEAM, 2011). In R there isn't a package specific for Procrustes analysis, but there is a package called *Shapes*, for statistical shape analysis, that brings OPA and GPA algorithms implemented (DRYDEN, 2009).

GenStat® is statistical software that is able to perform the Procrustes analysis, enabling many graphic tools (GENSTAT, 2003). Despite its high purchase cost, it is commonly used in sensory analysis (GAINS; THOMSON, 1990; MCEWAN; SCHLICH, 1992).

This proposed situation was performed in the programs R and GenStat® (7th edition) in order to compare the outputs. We also built graphics with R and got some practical information.

METHODOLOGY

Ordinary Procrustes Analysis (OPA) consists in find the best way one matrix fits a reference matrix by minimizing the sum of squared distances between the points (lines) in a fixed matrix and a shifted and rotated to-be-fitted-one (GOWER, 1975).

A reference matrix (expert scores) was established and a data matrix (trained assessor) was simulated in R (like demonstrated below).

It was considered four brands of Gorgonzola cheese, each line of the matrix representing a brand. The judges scored these four samples at two attributes (appearance and characteristic flavor), each column of the matrix representing an attribute. So, this simulated sensory analysis occurred in a fixed vocabulary context. The reason we choose analyze just two attributes was to facilitate the building of graphics.

Data simulation

Graybill (1976) presents an easy manner to generate data from a bivariate Normal distribution (X, Z), where $Z|X=x$ is found by a linear relation of X.

This way, to simulate data from a bivariate Normal distribution one has to assume known five parameters, since

$$(X, Z) \sim N_2(\mu_Z, \mu_X, \sigma_{ZZ}, \sigma_{XX}, \sigma_{ZX}) \qquad (2)$$

where μ_Z, μ_X are the means of Z e X; σ_{ZZ}, σ_{XX} the variances of Z e X and σ_{ZX} the covariance between Z e X; and where

$$\sigma_{ZX} = \rho\sigma_Z\sigma_X.$$

The linear relationship between Z e X is the mathematical expectation of the conditional $Z|X = x$

$$E[Z \mid X = x] = \mu_Z + \frac{\sigma_{ZX}}{\sigma_{XX}}(x - \mu_X) = \beta_0 + \beta_1 x \qquad (3)$$

and with variance

$$V[Z|X = x] = \sigma_{ZZ}(1-\rho^2) \qquad (4)$$

So, one value of the variable $Z|X=x$ is given by

$$Z|X = \mu_{Z|X=x} + e(x) \qquad (5)$$

where $e(x) \sim N(0, \sigma_{ZZ}(1-\rho^2))$.

Data simulation was necessary only to get the trained assessor's scores. We assumed that the trained assessor scores were random values from four bivariate

Normal distributions (*X, Z*), where *X* is the appearance and *Z* is the characteristic flavor. The means of these four bivariate Normal distributions are the expert scores.

The expert scores were arbitrarily established and it was considered a positive correlation between the attributes (appearance and characteristic flavor). We assumed the following covariance matrix

$$R = \begin{pmatrix} 100 & 120 \\ 120 & 225 \end{pmatrix}$$

Therefore, there was a positive correlation between attributes (ρ=0,8) but not between cheese brands.

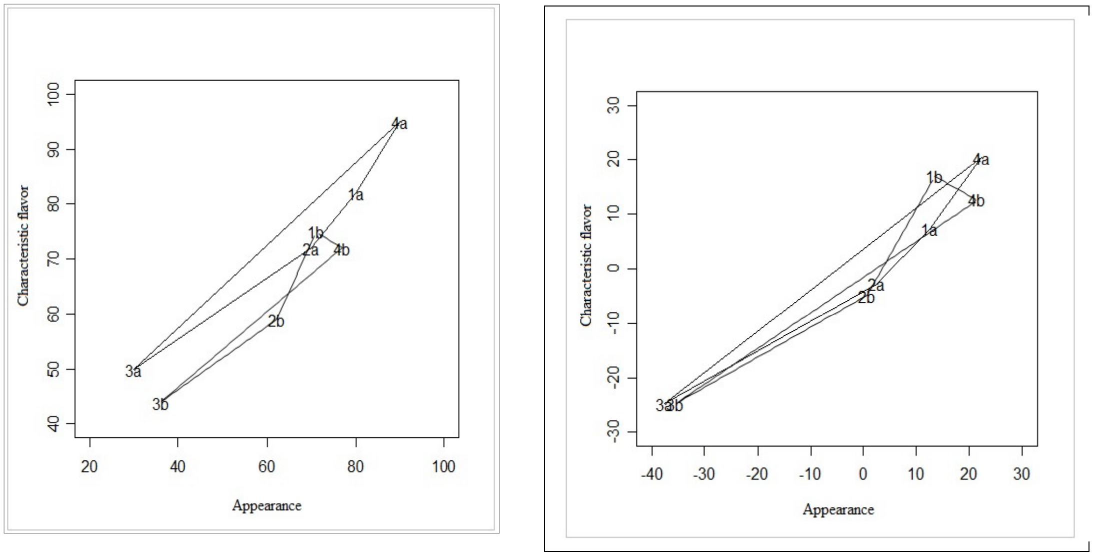
MAIN RESULTS

After data simulation, the established (A) and simulated matrices (B) (score matrices from expert and trained assessor, respectively) were:

$$A = \begin{pmatrix} 80\,82 \\ 70\,72 \\ 30\,50 \\ 90\,95 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 71\,75 \\ 62\,59 \\ 36\,44 \\ 77\,72 \end{pmatrix}$$

Later, the configuration graphics were made utilizing R. Figure 1 presents the configurations of the both assessors, before and after Procrustes analysis. In Figure 1 one can observe how OPA allowed a better view of the agreements and disagreements of such assessors.

Looking at the right hand side of Figure 1 one can observe that both assessors agree more about brands 2 and 3 than about brands 1 and 4. Considering first the appearance axis, for both assessors, the decreasing relation of brands is ordered like 4, 1, 2 and 3; but when one consider the characteristic flavor axis, the decreasing relation of brands for expert is ordered like 4, 1, 2 and 3; and for the trained assessor the relation of brands is ordered like 1, 4, 2 and 3. So, concerning to characteristic flavor, they disagree in first and second positions. However, both agree that brand 4 is the better whole sample and brand 3 is the worst.



(a) (b)

FIGURE 1. Configurations of the both assessors: expert (1a, 2a, 3a, 4a) e trained assessor (1b, 2b, 3b, 4b), before (left) and after (right) Ordinary Procrustes Analysis.

One can compare the R and GenStat® outputs analyzing Table 1. This table shows the element name given by each software and the numerical result of the rotation matrix, scaling factor, fixed configuration A, Procrustes fitted configuration B, and Ordinary Procrustes residual sum of squares.

In Table 1 one also can note that the both software numerical results are identical, what suggests that such R algorithm is the same algorithm implemented in GenStat®. Hence, the *Shapes* package OPA algorithm can be used in food sensory analysis without loss.

TABLE 1. Comparison between R and GenStat® outputs for Ordinary Procrustes Analysis. Software's element names and numerical results of rotation matrix, scaling factor, fixed configuration A, Procrustes fitted configuration B, and Ordinary Procrustes residual sum of squares.

	R	**GenStat®**
Name	**$R**	**Orthogonal Rotation**
Rotation matrix	$\begin{pmatrix} 0.99980923 & -0.01953201 \\ 0.01953201 & 0.99980923 \end{pmatrix}$	$\begin{pmatrix} 0.99981 & -0.01953 \\ 0.01953 & 0.99981 \end{pmatrix}$
Name	**$s**	**Least-squares Scaling factor**
Scaling factor	**1.375420**	**1.3754**
Name	**$Ahat**	**Xout**
Fixed configuration A	$\begin{pmatrix} 12.5 & 7.25 \\ 2.5 & -2.75 \\ -37.5 & -24.75 \\ 22.5 & 20.25 \end{pmatrix}$	$\begin{pmatrix} 12.50 & 7.25 \\ 2.50 & -2.75 \\ -37.50 & -24.75 \\ 22.50 & 20.25 \end{pmatrix}$
Name	**$Bhat**	**Yout**
Procrustes fitted configuration B	$\begin{pmatrix} 13.3998104 & 16.934260 \\ 0.5935525 & -4.826485 \\ -35.5635272 & -24.755373 \\ 21.5701643 & 12.647598 \end{pmatrix}$	$\begin{pmatrix} 13.40 & 16.93 \\ 0.59 & -4.83 \\ -35.56 & -24.76 \\ 21.57 & 12.65 \end{pmatrix}$
Name	**$OSS**	**Sum of squares**
Ordinary Procrustes residual sum of squares	**164.9520**	**Fitted Configuration 2992.7980** **Residual 164.9520** **--------------------------------------** **Fixed Configuration 157.7500**

FINAL REMARKS

In the proposed dairy context situation, brands 2 and 3 were better-characterized, i.e. expert and trained assessor better agreed brands 2 and 3. They also agreed that the better brand is number 4 and the worst is brand 3. Concerning to appearance, the panelists agreed about the order of the brands; and concerning to characteristic flavor, they didn't agreed about two positions.

R and GenStat® outputs were numerically identical, but with different element names (notations). Hence, R is a useful tool to perform Ordinary Procrustes Analysis in food sensory analysis context, leading to the same conclusions that GenStat® and also enabling graphics building.

Those quantitative methods will allow managers of companies of products highly dependent of human senses to objectively analyze their data, leading to more secure decisions, avoiding subjectivity of opinions, which can put under risk the goal of an management action in a competitive market culture.

ACKNOWLEDGMENTS

We would like to thank Professor Ann Noble (University of California), professor John Clifford Gower (Open University – UK) and the financial agencies Capes and CNPq.

REFERENCES

DRYDEN, I. **Shapes: Statistical shape analysis**. R package version 1.1-3. 2009. http://CRAN.R-project.org/package=shapes.

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.

FERREIRA, E. B.; OLIVEIRA, M. S. **Sensometria**: uma abordagem com ênfase em Procrustes**.** Santa Maria: UFSM, 2007. 71p.

FERREIRA, E. B.; OLIVEIRA, M. S. de; FERREIRA, D. F.; MAGALHÃES, F. A. R. Sensory Profile of Gorgonzola via Generalized Procrustes Analysis Using R. **Boletim do Centro de Pesquisa e Processamento de Alimentos**, v. 26, p. 151-159, 2008.

GAINS, N.; THOMSON, D.M.H. Sensory profiling of canned lager beers using consumers in their own homes. **Food Quality and Preference**, v. 2, p. 39-47. 1990.

GENSTAT. **Genstat for Windows**, Release 7.1.0.198, 7th edn. Oxford, UK: VSN International Ltd. 2003.

GOWER, J.C. Generalized Procrustes Analysis. **Psychometrika**, v. 40, p. 33-51. 1975.

GRAYBILL, F.A. **Theory and Application of the Linear Model**. Duxbury Press:Boston, Massachusetts. 1976. 703 p.

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

MCEWAN, J.A.; SCHLICH, P. Correspondence Analysis in sensory evaluation. **Food Quality and Preference**. v. 3, p. 23-36. 1992.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2011. URL http://www.R-project.org.

SCHÖENEMANN, P.H. A generalized solution of the Orthogonal Procrustes Problem. **Psychometrika**, v. 31, n. 1, p. 1-10. 1966.

SLACK, N.; CHAMBERS, C. H.; JOHNSTON, R. **Administração da Produção.** São Paulo: Atlas. 1997. 726 p.