SENSORY PROFILE OF GORGONZOLA VIA GENERALIZED PROCRUSTES ANALYSIS USING R

ERIC BATISTA FERREIRA*
MARCELO SILVA DE OLIVEIRA**
DANIEL FURTADO FERREIRA***
FERNANDO ANTONIO RESPLANDE MAGALHAES****

Sensory data analysis in Brazil used to be performed through univariate Analysis of Variance (ANOVAs) hence one looses valuable intra variable information. To solve this problem one may use multivariate tools. Generalized Procrustes Analysis (GPA), often performed in expensive software, is a multivariate exploratory data analysis method that aims: (i) remove scores' bias; (ii) resume the assessors' agreement, usually, on a plan; (iii) show the objects relative differences. An application on sensory profile of Gorgonzola cheese is provided. Data was analyzed by uni, multivariate ANOVAs and GPA, on the free statistical software R. ANOVAs detected significant differences among cheeses and GPA provided the consensus configuration displaying their relative differences and sensory evolution. Assessors preferred cheeses with a higher fat content. GPA can be reasonably performed by free software, say R, and should be utilized like a complement of Analysis of Variance.

KEY-WORDS GENERALIZED PROCRUSTES ANALYSIS; SENSORY ANALYSIS; MULTIVARIATE STATISTICS; GORGONZOLA.

* Doutorando em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras (UFLA), Lavras, MG (e-mail: ericbferreira@netscape.net).

** Professor Adjunto, Departamento de Ciências Exatas, UFLA, Lavras, MG (e-mail: marcelo.oliveira@ufla.br).

*** Professor Adjunto, Departamento de Ciências Exatas, UFLA, Lavras, MG (e-mail: danielff@ufla.br).

**** Pesquisador, Empresa de Pesquisa Agropecuária de Minas Gerais/Centro de Treinamento/Instituto de Lacticínios Candido Tostes (e-mail: fernando.magalhães@epamig.br).

1 INTRODUCTION

Competition amongst food industries looking for a greater market share requires the improvement of quality and also the development of statistical methodologies concerned with fidelity and suitable decision-making.

Sensory evaluation of food allows that assessors or judges evaluate, through their basic senses, one or more attributes of a food. With that purpose in mind, they use mainly the taste, the smell and the touch (MORAES, 1988). A sensory panel is constituted by a set of people, trained or not to detect differences amongst foods, determine its profile, score the acceptation, identify the most positive and/or the most negative attribute, or verify if it is within the range of some pre-established quality limits. According to TEIXEIRA, MEINERT & BARBETTA (1987), sensory evaluation is applied even in new product development by the food industry.

Generalized Procrustes Analysis (GPA) is considered a multivariate exploratory data analysis since it is not concerned with inference (FERREIRA, 2004). When applied to sensory evaluation, it's main focus is to provide graphical interpretation of the between-products distances.

In the Greek mythology, Procrustes, the son of Poseidon, kept an inn benefiting from what he claimed to be a wonderful all-fitting bed. He lopped off excess height from tall guests and either flattened short guests by hammering or stretched them by racking. That is why such an adjusting data procedure was honored with his name (GOWER & DIJKSTERHUIS, 2004).

Procrustes analysis was generalized by GOWER (1975), allowing the multiple comparison of, say, *m* matrices or assessors. Since GOWER (1975) such methodology became popular and it's use is not restricted to sensory analysis. One can find GPA applied to, besides food sciences, paleontology (RODRIGUES & SANTOS, 2004), medicine (RANGARAJAN, CHUI & BOOKSTEIN, 2003), ecology (MORAES, 2004), molecular genetics (MIRONOV, PEVZNER & GELFAND, 2004), automotive industry (DAIROU et al., 2004) and Photogrametry (AKCA, 2003), etc.

In sensory analysis of food, GPA can be found in paper from abroad like DIJKSTERHUIS (1994), PASTOR et al. (1996), SINESIO & MONETA (1997) e DELAHUNTY et al. (1997).

In Brazil, its use is still limited. One can address BENASSI, DAMÁSIO & CECCHI (1998), which characterized the profile of national italic Riesling wines.

Usually, specific packages (McEWAN, 1989) or routines (de JONG, HEIDEMA & VARDEN KNAAP, 1998) are required to perform GPA. In both, a great amount of money is invested on such purchases. Therefore, this study aimed to explore the use a free statistical software, namely R, to play the same role.

2 MATERIAL AND METHODS

Let X_i (i = 1, 2, ...,m) be matrices n x p_i. In sensory context, these matrices contain scores of *m* assessors, to *n* foods, in p_i attributes or dimensions. Note that the number of attributes may vary from one assessor to another. In each matrix, the j-th row (j = 1, ..., n) means the coordinates of a point $P_j^{(i)}$ in p_i orthogonal axes, i.e., in $\square^{p_i}$. These are the scores of the i-th assessor to the j-th food accordingly to their arbitrary p_i attributes. Therefore, each assessor describes the same set of *n* foods as *n* points in their Euclidian space. The number of rows (p_i) in each matrix can differ according to the assessor's wish but here, to simplify, they are considered equal, i.e., $p_i = p \ \forall i$ (FERREIRA, 2004).

According to GOWER (1975), when joined, the points of one same assessor form a polygon called the *assessor's configuration*. GPA's main idea is to best fit the *m* configurations by motions of rotation, reflection, translation, stretching and shrinking; preserving unaltered the relative distances

between points of a same configuration. One criterion that can be used is minimizing the Residual Sum of Squares (S),

$$S = \sum_{i<j}^{m} \left\| X_i^* - X_j^* \right\| \quad (1)$$

i.e., the sum of squared distances among the transformed configurations X_i^* and X_j^* Without loss of generality for X_j^* ,

$$X_i^* = \rho_i X_i H_i + T_i \quad (2)$$

where ρ_i is a scaling factor, X_i is the original matrix of scores from assessor *i*, H_i is an orthogonal matrix and T_i is a matrix containing, in each row, the column means of X_i. In this context, ρ_i performs the shrinkage or stretches, H_i performs the rotation and T_i performs the translation to coincide all centroids at the origin of the Euclidian space.

It can be demonstrated that the right hand side of (1) equals

$$m \sum_{i=1}^{m} \left\| X_i^* - G \right\| \quad (3)$$

where $G = \left(\sum_{i=1}^{m} X_i^* \right) / m$ is called *group average configuration* (GOWER & DIJKSTERHUIS, 2004). Obviously, when we find ρ_i , H_i and T_i ($i = 1, 2, \ldots, m$) that minimize $\sum_{i=1}^{m} \left\| X_i^* - G \right\|$ it will be at the minimum of , that is, the best fit.

Looking for these parameters might be a tricky task and requires an iterative algorithm that, in each step, looks for the best value of one parameter treating the others as constants. GOWER & DIJKSTERHUIS (2004) compile many of those algorithms. Here is considered that one that deals only with orthogonal matrices H_i

After a satisfactory convergence, results can be displayed in a plot by referring *G* and/or X_i ($i = 1, 2, \ldots, m$) to the principal axes of *G*, that is, *GH* and $X_i^* H$ ($i = 1, 2, \ldots, m$), where:

$$G'G = H \Lambda H' \quad (4)$$

is the spectral decomposition of *G'G* and, obviously, *H* contains it's eigenvectors. GOWER (1975) found 10^{-4} to be a nice convergence, so that limit was used here as well.

MAGALHÃES (2002) studied the sensory profile of Gorgonzola cheeses from two distinct technologies along their ripening, via Quantitative Descriptive Analysis (QDA). Data analysis was performed through univariate analysis of variance (ANOVA). That author gently conceded the full data set to be further analyzed. Eight cheeses, 2 technologies (Tec) in 4 ripening ages (Ag), have been scored by 9 experts (Exp), along 10 sensory attributes, in 4 repetitions. Cheeses from technology 1 had its fat level higher than those from technology 2. Ripening ages were 30, 45, 60 e 75 days; and the sensory attributes were 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), and residual flavor (RF). Experiment was carried out in random blocks design and the data set was analyzed in factorial scheme 2x4.

Univariate ANOVAs were performed in Sisvar v4.3 (FERREIRA, 2000) and R v2.0.1 (R DEVELOPMENT..., 2004); a multivariate ANOVA was performed in R v2.0.1 e Minitab v13.20; and the Generalized Procrustes analysis was performed through function *procGPA* of the R package *Shapes*. That function provided the transformed configurations (X_i^*) and the group average configuration (*G*). Additional required operations were programmed in R language. Means of the 4 repetitions for each cheese were used as main entry in GPA.

3 RESULTS AND DISCUSSION

Table 1 displays an abstract of the ten univariate ANOVAs. It is clear that, except for appearance, all the interactions *technology x age* are significant. That is, the differences between technologies are constant along the time.

Multivariate analysis of variance (MANOVA) was performed to search for joint significance for the mean's differences between treatments. Table 2 displays a summary of the MANOVA for that data set. According to Table 2, there is a joint significance amongst treatments and interactions. Therefore, a single set of univariate ANOVAs certainly would loose lots of information. That is a strong reason for associating a multivariate technique, as GPA.

TABLE 1 - MEAN SQUARES OF THE ANOVA'S SOURCES OF VARIATION

SV	Ap	AM	Tx	CO	Ar	CF	ST	AT	BT	RF
Exp	0.2	0.22	0.08	0.13	0.06	0.243	0.05	0.26	0.06	0.05
Tec	41.06*	68.7*	53.2*	14.59*	6.19*	19.2*	2.13*	0.52*	0.33*	0.07
Ag	53.11*	66.4*	10.7*	35.2*	7.5*	48.88*	53.9*	1.18*	3.56*	29.1*
Tec x Ag	0.21	2.69*	3.17*	1.49*	2.85*	2.15*	1.08*	1.52*	4.88*	5.74*

Ap = appearande; AM = amount of mold; Tx = texture; CO = characteristic odor; Ar = aroma; CF = characteristic flavor; ST = salty taste; AT = acid taste; BT = bitter taste; RF = residual flavor.
* Correspondent p-value < 0.05. SV = Sources of Variation.

TABLE 2 – SUMMARY OF MANOVA: P-VALUES ACCORDING TO WILK, LAWLEY-HOTELLING AND PILLAI CRITERIA

Source of Variation	Wilk	Lawley-Hotelling	Pillai
Assessors	0.240 ns	0.087 ns	0.421 ns
Ages	0.000***	0.000***	0.000***
Technologies	0.000***	0.000***	0.000***
Ages x Technologies	0.000***	0.000***	0.000***

*** p-value < 0,1%.

Principal axes analysis performed after GPA revealed that the first two principal axes could explain more than 86% of the total variance. Therefore, the profile graphics were considered in a plain. Figure 1 illustrates the behavior of variance explained by the principal axes.

After GPA, all transformed scores (panel agreement) were plotted in Figure 2(a) highlighting the sets of scores for a same cheese. It is worth to note that all assessors seemed to distinguish the cheeses in the same way.

FIGURE 1 – VARIANCE EXPLAINED BY THE PRINCIPAL AXES

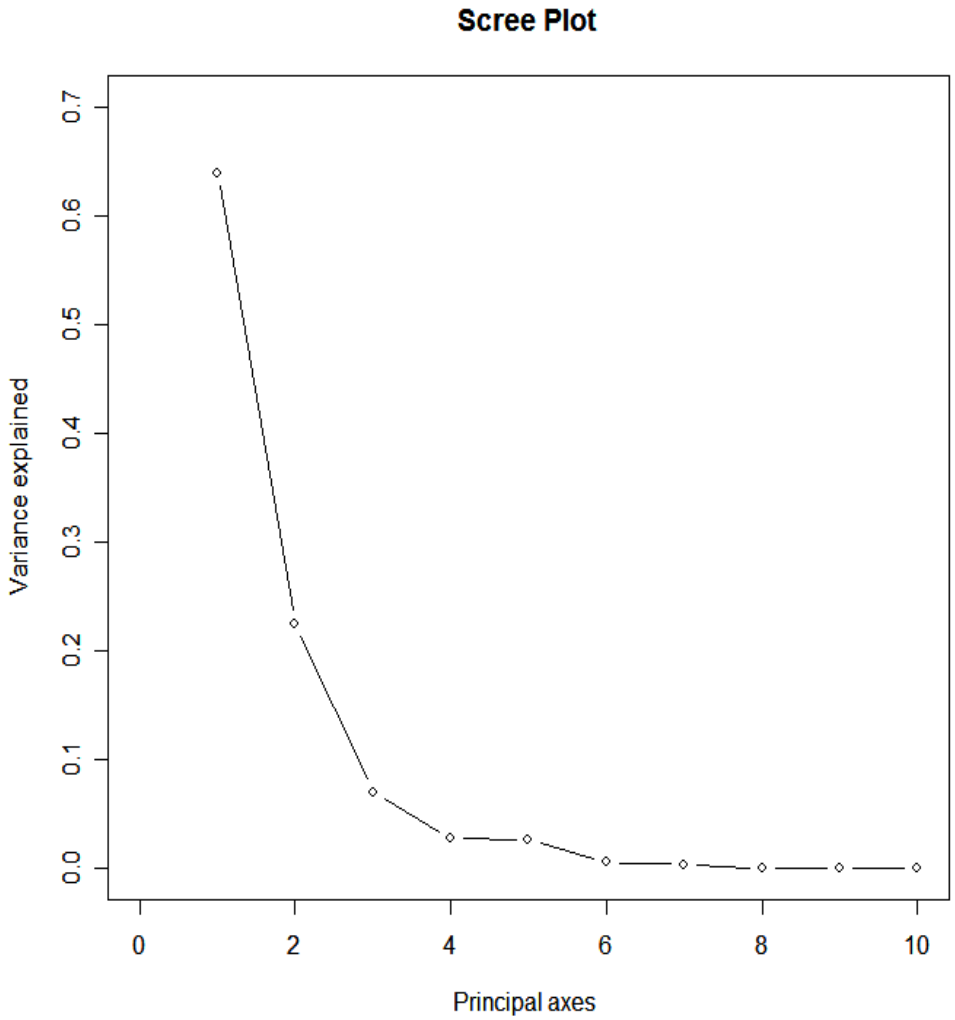
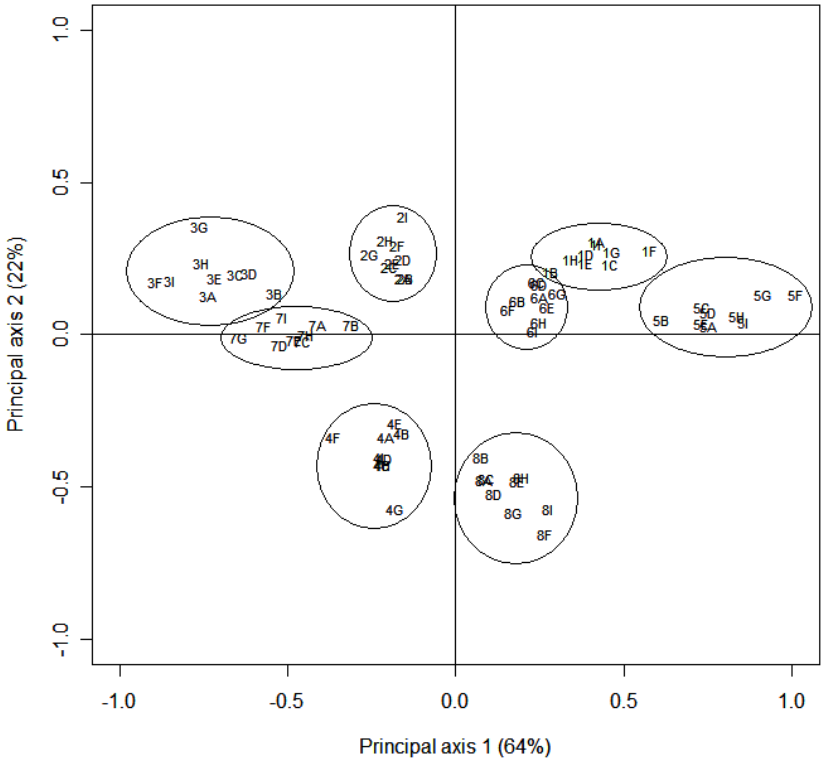


Figure 2(b) plots only the centroids of the assessor's configurations, i.e., the agreement average for the location of each product in that space. Technology 1 (cheeses 1, 2, 3 and 4) and technology 2 (cheeses 5, 6, 7 and 8) seem to have the same behavior, i.e., they seem to run unclockwisely along the time, since the cheeses are ordered by age (1 and 5 are 30 days old, 2 and 6 are 45 days old, and so on). That is a very interesting pattern, a *sensory run along the ripening* according to those attributes.

Though the two technologies describe the same trajectory along ripening, they can be distinguished by analyzing what the new axes mean. A correlation circle in Figure 3 suggests that the majority of desirable attributes is concentrated in quadrants II and III. Consequently, the cheeses around are considered better by that panel. Hence, the more left situated, the better the cheese. That suggests that technology 1 is considered better than technology 2 in all studied ages, what agrees with MAGALHÃES (2002).

FIGURE 2 - (a) ASSESSOR'S COFIGURATIONS. SCORES FOR EIGHT CHEESES (1 TO 8) FROM NINE ASSESSORS (A TO I). (b) CONCENSUS' AVERAGE HIGHLIGHTING THE SENSORY UNCLOCKWISE RUN ALONG THE RIPENING FOR TECHNOLOGY 1 (——) AND 2 (- - -)

(a)



(b)

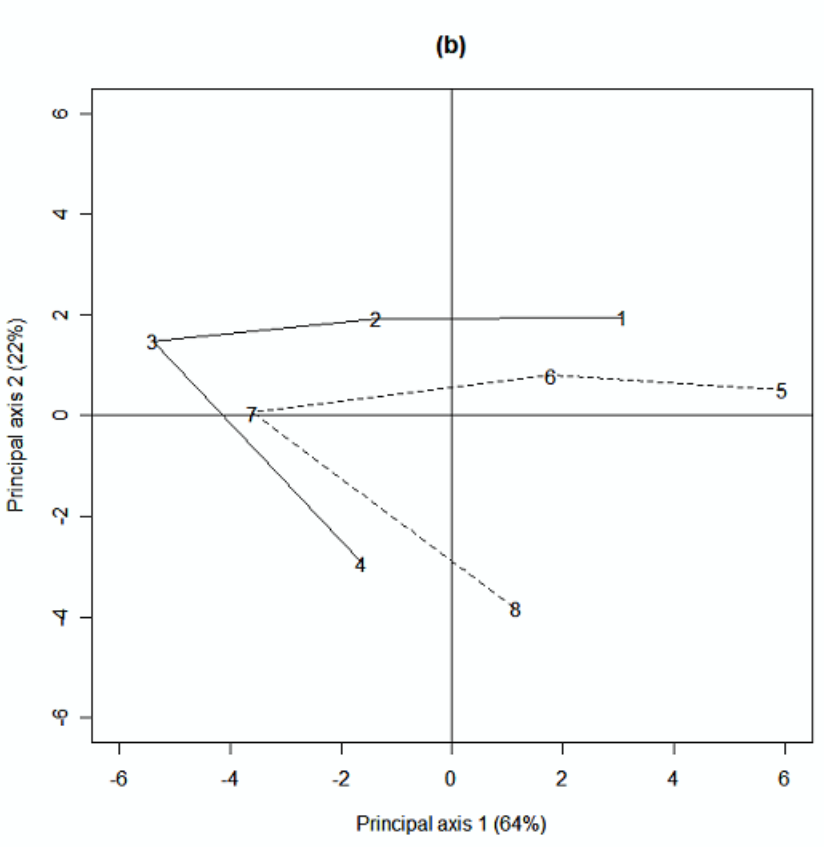
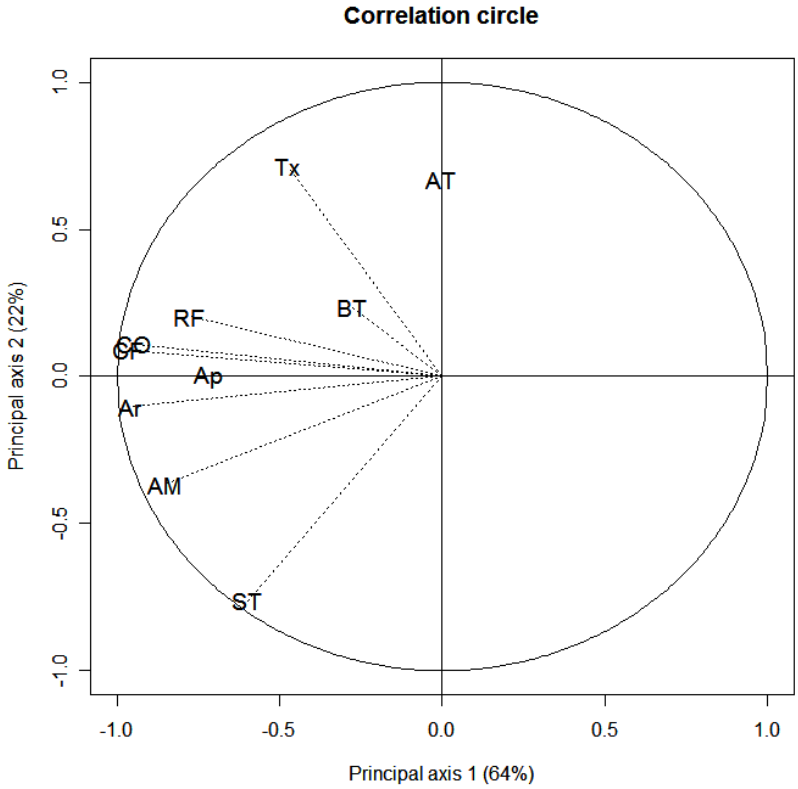


FIGURE 3 – CORRELATION CIRCLE BETWEEN THE ATTIBUTES AND THE NEW AXES



MAGALHÃES (2002) emphasizes that Gorgonzola cheese reaches its maximum of desired characteristics around 60 days of ripening. Relating Figures 2(b) and 3, one can note that cheeses 3 and 7 are preferred by the panel and they are both 60 days old. Whole panel seems to agree that chesses too young (1, 5, and 6) or too old (8) don't have the desired characteristics yet or have lost them already. It is worth noting that technology 1 maintain itself with high acceptation rates from 45 to 75 days, what happens with technology 2 just around 60 days of ripening.

Technology 1 produced cheeses with higher fat contents. That fact might be crucial in its acceptance and developed good characteristics like flavor and aroma (MAGALHÃES, 2002).

4 CONCLUSION

Generalized Procrustes analysis enabled a visual confirmation of the preference for cheeses with higher fat content (technology 1). That technology provided cheeses with superior sensory profile along the whole ripening.

Univariate and multivariate analysis of variance were fundamental in detection of significant number of the factors and interactions. Therefore, ANOVA, MANOVA and GPA are complementary rather than equivalent, or competitor.

Performing GPA applied to sensory context in a free statistical software (say, R) is possible and quite easy using one ready function and few additional implementations.

Using GPA with more assurance to analyze sensory data (perhaps, without complementary tools) depends on the development of the knowledge about its performance in statistical inference, what means a wide field to be explored by future works.

RESUMO

PERFIL SENSORIAL DO QUEIJO GORGONZOLA VIA ANÁLISE GENERALIZADA DE PROCRUSTES USANDO O R

Dados provenientes de análise sensorial de alimentos têm sido avaliados, normalmente, no Brasil por meio de análises de variância (ANAVA) univariadas. Dessa maneira, perdem-se as informações de correlação intravariável fornecidas pelas ferramentas da Estatística Multivariada. A Análise Generalizada de Procrustes (AGP), que geralmente demanda software caro, constitui análise exploratória de dados multivariada que tem por objetivos: (i) retirar os viéses que podem estar contidos nas notas; (ii) resumir informações numa configuração de consenso plotada, geralmente, no plano; e (iii) ressaltar as diferenças relativas entre os objetos analisados. Para ilustrar sua aplicação, dados provindos da análise sensorial de queijos tipo Gorgonzola foram submetidos a ANAVA uni e multivariadas e GPA, utilizando-se o software livre R. Detectou-se a preferência por queijos com maiores teores de gordura e a AGP propiciou a visualização dessas diferenças e a evolução sensorial dos queijos. A AGP pode ser satisfatoriamente executada por software estatístico grátis (R) e deve ser utilizada como complemento às análises de variância.

PALAVRAS-CHAVES: ANÁLISE GENERALIZADA DE PROCRUSTES; ANÁLISE SENSORIAL; ESTATÍSTICA MULTIVARIADA; GORGONZOLA.

REFERENCES

1 AKCA, M. D. **Generalized Procrustes Analysis and its applications in Photogrammetry.** Zuerich: Swiss Federal Institute of Technology, Institute of Geodesy and Photogrammetry, ETH-Hoenggerberg, 2003. 23 p. Available on: <http://e-collection. ethbib. ethz. ch/ecol-pool/bericht/bericht_363. pdf>. Accessed on: 28 out. 2004.

2 BENASSI, M. T.; DAMÁSIO, M. H.; CECCHI, H. M. Avaliação sensorial de vinhos Riesling Itálico nacionais utilizando perfil de livre escolha. **Ciência e Tecnologia de Alimentos**, Campinas, v. 18, n. 3, p. 265-270, ago./out. 1998.

3 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. Available on: <http://www. perception-sensorielle. com/Congress2003/Dairou_SAE. pdf>. Accessed on: 2 ago. 2004.

4 DE JONG, S.; HEIDEMA, J.; VAR DER KNAAP, H. C. M. Generalized Procrustes Analysis of coffee brands tested by five European sensory panels. **Food Quality and Preference**, Oxford, v. 9, n. 3, p. 111-114, May 1998.

5 DELAHUNTY, C. M.; MCCORD, A.; O'NEILL, E. E.; MORRISSEY, P. A. Sensory characterization of cooked hams by untrained consumers using Free-Choice Profiling. **Food Quality and Preference**, Oxford, v. 8, n. 5/6, p. 381-388, Sept./Nov. 1997.

6 DIJKSTERHUIS, G. B. Procrustes analysis in studying sensory-instrumental relations. **Food Quality and Preference**, Oxford, v. 5, n. 2, p. 115-120, May 1994.

7 FERREIRA, D. F. Análises estatísticas por meio do Sisvar para Windows versão 4. 0. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 45., 2000, São Carlos. **Programas e Resumos**... São Carlos, SP: UFSCar, 2000. p. 255-258.

8 FERREIRA, E.B. **Análise generalizada de procrustes via R:** uma aplicação em laticínios. Lavras, 2004. 115 p. (Mestrado em Estatística e Experimentação Agropecuária) – Departamento de Ciências Exatas, Universidade Federal de Lavras.

9 GOWER, J. C.; DIJKSTERHUIS, G. B. **Procrustes problems**. New York: Oxford University Press, 2004. 256 p.

10 GOWER, J.C. Generalized Procrustes Analysis. **Psychometrika**, Williamsburg, v. 40, n. 1, p. 33-51, 1975.

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

12 MIRONOV, A.; PEVZNER, P.; GELFAND, M. Procrustes: gene recognition via spliced alignment. Available on: <http:// www-hto.usc.edu/software/procrustes/ index.html#salign>. Accessed on: 29 feb. 2004.

13 McEWAN, J. A. Review of statistical software for sensory evaluation. **Food Quality and Preference**, Oxford, v. 1, p. 179-181, 1989.

14 MORAES, M. A. C. **Métodos para avaliação sensorial dos alimentos**. 6. ed. experimental. Campinas: Editora da UNICAMP, 1988. 93 p.

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

16 PASTOR, M. V.; COSTELL, E.; IZQUIERDO, L.; DURAN, L. Sensory profile of peach nectars: evaluation of attributes and assessors by generalized Procrustes analysis. **Food Quality and Preference,** Oxford, v. 7, p. 347 – 357, 1996.

17 R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available on: <http://www. R-project. org>. Accessed on: 2004.

18 RANGARAJAN, A.; CHUI, H.; BOOKSTEIN, F. L. **The softassign matching algorithm**. Available on: <http://citeseer.nj.nec.com/rangarajan97softassign. html>. Accessed on: 18 dec. 2003.

19 RODRIGUES, L. A.; SANTOS, V. F. **Morfometria geométrica – aplicações em paleobiologia de dinossauros**. Disponível em: <http://www.mnhn.ul. pt/dinos/public_html/Luis_A_Rodrigues/ABstractCONGGEo.htm>. Acesso em: 17 feb. 2004.

20 SINESIO, F.; MONETA E. Sensory evaluation of walnut fruit. **Food Quality and Preference,** Oxford, v. 8, n. 1, p. 35-43, Jan. 1997.

21 TEIXEIRA, E.; MEINERT, E. M.; BARBETTA, P. A. **Análise sensorial de alimentos**. Florianópolis: Ed da UFSC, 1987. 180 p.