0308-8146(95)00080-1

Food Chemistry, Vol. 55, No. 2, pp. 189–191, 1996
Copyright © 1996 Elsevier Science Limited
Printed in Great Britain. All rights reserved
0308-8146/96 $15.00+0.00

Analytical, Nutritional and Clinical Methods Section

Application of discriminant analysis to physico-chemical variables for characterizing Spanish cheeses

R. Millán,* P. Saavedra, E. Sanjuán & M. Castelo

Nutrición y Bromatología, Facultad de Veterinaria de la ULPGC, C/Francisco Inglott Artiles, 12A, 35016 Las Palmas, Spain

(Received 22 September 1993; revised version received 31 March 1995; accepted 31 March 1995)

Spanish cheeses have been traditionally grouped into varieties according to their origin, acceptance and production peculiarities. Although pronounced sensorial differences can be detected among them, the variations are subtle and hard to evaluate. The variables which discriminate are now determined optimally among 10 classes of cheeses.

INTRODUCTION

Cheeses with many different shapes, odours and flavours are produced in Spain. They are generally of high quality and the criterion for their official classification has been their Spanish origin. They are genuine products in their respective origin regions with appreciation at least in the regional markets.

Among the principal varieties of Spanish cheeses that we have studied (Millán, 1981) we include the Galician' cheeses, Tetilla and San Simón; Cabrales, a blue vein cheese from Asturias; Burgos and Villalón, fresh cheese varieties, manufactured in Castilla; Manchego, elaborated with sheep's milk in La Mancha; Idiazábal from the Basque Country; Roncal, elaborated in Navarra with sheep's milk, and fresh varieties from Cádiz and the Málaga Mountains.

Although the origin guarantees and appearance of cheeses were sufficient for their identification, it is also true that among some varieties, the differences are often subtle and hard to evaluate; this is complicated when some distinctive characteristics are missing.

In this respect, lineal discriminant analysis has been used in different cheese varieties (Bevilacqua & Califano, 1992; Martín Hernández *et al.*, 1992; Lucisano *et al.*, 1991; Smith & Nakai, 1990; Alonso *et al.*, 1987*a*,*b*; Aishima & Nakai, 1987; Santa Maria *et al.*, 1986; Pham & Nakai, 1984; Godwin *et al.*, 1977; Cepeda *et al.*, 1990).

We have selected from the analysed variables, in the aforementioned Spanish cheese varieties, the ones that let us obtain the corresponding discriminant functions and to calculate the best distinguishable matrix of distances among classes to achieve a correct identification. Then to single these out mathematically in order to typify them from the analytical data.

MATERIALS AND METHODS

Eight commercial cheese samples as diversified as possible from each of 10 representative Spanish varieties, were purchased from retail markets in several areas of the country.

Recommended methods for chemical analysis of cheeses were used to determine non-protein nitrogen (Mogensen), ammonia nitrogen (microdiffusion), moisture (air oven method) and salt (modified Volhard test). The pH was measured potentiometrically with a pH meter on 10% suspensions of cheese (w/v) in distilled water. All these methods have been described by Millán (1981).

First, a subset of variables (NNP, NNH_3, pH, NaCl and moisture), optimum for the discrimination from the variables considered from the begining (NS, NNP, NNH_3, a_W, pH, NaCl, moisture and fat), was selected for the statistical analysis. It is denoted by C_i the i-th class of cheese, with $i=1,...,10$. The feature vector of variables X on a cheese is assumed to be distributed $N(\mu_i,\sum_i)$ in C_i $(i=1,...,10)$, where $\mu_1,...,\mu_{10}$ denote the group means, and $\sum_1,..,\sum_{10}$ the group covariance matrices. We used the Bartlett test to contrast the covariance matrices homogeneity. Since this contrast is significant, it uses the Fatti & Hawkins (1986) method to select the variables in the heteroscedastic case. They consider three different methods for the

*To whom correspondence should be addressed.

Table 1. Group means (and standard deviations)

Class	NPN	NNH_3	pH	ClNa	Moisture
Burgos	2.31	0.560	5.78	0.223	57.60
	(0.768)	(0.056)	(0.126)	(0.178)	(3.37)
Cabrales	37.57	10.45	5.80	1.18	38.54
	(3.63)	(2.38)	(0.381)	(0.192)	(3.11)
Cádiz	4.73	0.456	5.63	0.93	46.92
	(0.799)	(0.145)	(0.324)	(0.175)	(3.58)
Idiazábal	13.72	2.23	5.41	1.34	34.85
	(2.52)	(1.07)	(0.218)	(0.210)	(5.58)
Málaga	3.31	0.371	5.56	0.801	49.34
	(0.947)	(0.113)	(0.269)	(0.169)	(3.52)
Manchego	14.05	2.47	5.64	1.22	34.0
	(5.35)	(3.57)	(0.286)	(0.215)	(4.39)
Roncal	15.82	1.90	5.71	1.33	29.77
	(2.31)	(0.208)	(0.0556)	(0.154)	(2.743)
San Simón	15.25	1.54	6.06	0.901	41.86
	(4.38)	(0.753)	(0.289)	(0.222)	(3.75)
Tetilla	6.76	0.779	5.71	0.686	44.09
	(1.84)	(0.135)	(0.186)	(0.162)	(2.88)
Villalón	1.87	0.54	6.30	0.485	55.19
	(0.252)	(0.148)	(0.361)	(0.162)	(2.24)

Table 2. Generalized square distances between classes

	Burgos	Cabrales	Cádiz	Idiazábal	Málaga	Manchego	Roncal	San Simón	Tetilla	Villalón
Burgos	−14.66	861.11	14.48	105.52	10.57	50.31	303.25	24.77	26.63	10.66
Cabrales	160475	−2.25	10.142	382.5	20.835	42.12	40.16	201.12	10829	96250
Cádiz	146.72	429.65	−10.61	16.15	−8.35	57.89	120.24	37.53	28.88	454.78
Idiazábal	2830	186.00	359.83	−5.82	836.55	47.13	164.75	76.61	156.75	9402
Málaga	85.30	533.18	−4.68	30.39	−11.29	79.51	192.19	36.18	32.07	97.88
Manchego	3470	208.33	418.21	1.93	1056	−2.08	6.26	66.01	255.72	10347
Roncal	765.98	170.85	402.27	46.83	605.91	11.53	−12.42	65.94	115.69	13143
San Simón	631.41	206.28	588.89	184.34	271.07	48.06	297.24	−4.93	58.07	11601
Tetilla	121.23	528.36	15.63	23.45	65.16	15.35	102.44	8.44	−9.30	1517
Vilallón	72.86	635.27	21.25	115.93	11.41	57.12	1715	19.62	13.88	−12.94

It should be noted that this distance is not symmetrical. Thus, for instance, the distance from class 1 to class 2 is 861.11 while from 2 to 1 is 160475. This is due to a longer spread of class 2. This is to say, all cheeses of class 1 are near class 2, while some cheeses of class 2 are quite far from class 1. The smaller the distance between classes, the closer is the similarity between the samples. This is shown in the sensorial characteristics of each variety. This is the case of goat's cheeses from Cádiz and Málaga, produced in a nearby area. Idiazabal, Roncal and Manchego sheep's cheeses are sometimes difficult to distinguish, and can be identified by the distance. This is the same in Galician' cheeses, Tetilla and San Simon, and in the Castilian' cheeses, Burgos and Villalón with a very short distance between them.
The most clear reference is constituted by Cabrales, a fungi-matured cheese, with very different sensorial characteristics, that cannot be identified with the other classes because of a long distance.

combined statistics, and we have used the Fisher method.

Once the optimal set of discriminant variables were selected, the Bartlett test was again used to contrast the covariance matrices homogeneity of selected variables. Since this contrast was again significant, the discrimination in the generalized square distance was found: generalized square distance from an arbitrary cheese with feature vector X to class C_i is defined by:

$$d_i^2(X) + \log|S_i| \qquad (1)$$

where $d_i^2\ (X) = (X - \overline{X_i})' S_i^{-1}\ (X - \overline{X_i})$ is the Mahalanobis distance from X to i-th class, and $\overline{X_i}$ and S_i are the estimate means vector and covariance matrices, respectively, in class C_i. Then, an arbitrary cheese with feature vector X is assigned to class C_i such that:

$$d_i^2(X) + \log|S_i| = \min_{j=1,\ldots,10} d_j^2(X) + \log|S_j|. \qquad (2)$$

Then, the generalized square distance from class C_i to class C_j is defined as the generalized square distance from $\overline{X_i}$ to class C_j. The generalized square distance from class C_j to itself is $\log|S_j|$.

Assuming that the prior probabilities of membership groups are equal, we compute the posterior probabilities $P(C_j|X)$ for all cheeses to belong to each class considered, as

$$P(C_j|X) = \frac{\exp(-\frac{1}{2}(d_j^2(X) + \log|S_j|))}{\sum_k \exp(-\frac{1}{2}(d_k^2(X) + \log|S_k|))} \qquad (3)$$

and X, the feature vector of variables. Then, each cheese is reclassified in the class it is most likely (posterior probability) to belong to.

We wrote a Pascal program to develop the Fatti & Hawkins (1986) method for the selection of variables in the heteroscedastic case. The generalized square distances between classes and posterior probabilities were computed by means of the SAS statistical package.

RESULTS

Table 1 gives the means and standard deviations for each selected variable using the Fatti & Hawkins (1986) method in each class.

For the selected variables, we computed the generalized square distance from each class C_i to each class C_j. Table 2 gives the generalized square distances between the classes considered.

By application of discriminant analysis, from the selected variables, all cheeses are classified in the same class as they were in the begining (100%), and this confirms the suitability of this method.

REFERENCES

Aishima, T. & Nakai, S. (1987). Pattern recognition of GC profiles for classification of cheese variety. *J. Food Sci.*, **52**(4), 939–42.

Alonso, L., Juárez, M., Ramos, M. & Martin, Alvarez, P. J. (1987*a*). Overall composition, nitrogen fractions and fat characteristics of Cabrales cheese during ripening. *Zt. Lebensmitt. Unters. Forsch.*, **185**(6), 481–6.

Alonso, L., Ramos, M., Martín Alvarez, P. J. & Juárez, M. (1987*b*). Application of stepwise discriminant analysis to parameters for characterizing frozen Cabrales cheese. *J. Dairy Sci.*, **70**(5), 905–8.

Bevilacqua, A. E. & Califano, A. N. (1992). Changes in organic acids during ripening of Port Salut Argentino cheese. *Food Chem.*, **43**(5), 345–9.

Cepeda, A., Paseiro, P., López de Alda, M. L. & Rodriguez Otero, J. L. (1990). Intento de tipificación de distintos tipos de quesos gallegos comerciales mediante el análisis discriminante. *Alimentaria*, **213**, 25–7.

Fatti, L. P. & Hawkins, D. M. (1986). Variable selection in heteroscedastic discriminant analysis. *J. Am. Stat. Ass.*, **81**, 494–500.

Godwin, D. R, Washam, C. J. & Powers, J. J. (1977). Selection of Blue cheese quality descriptors by sensory evaluation. *J. Dairy Sci.*, **60**, Suppl. 1, 42.

Lucisano, M., Casiraghi, E., Pompei, C. & Dellea, C. (1991). Characterization of Provolone cheese. II. Proposal for a quality control procedure. *Ital. J. Food Sci.*, **3**(2), 113–25.

Martín Hernández, C., Amigo, L., Martín Alvarez, P. J. & Juárez, M. (1992). Differentiation of milks and cheeses according to species based on the mineral content. *Zt. Lebensmitt. Unters. Forsch.*, **194**(6), 541–4.

Millán, R. (1981). Tesis Doctoral. Facultad de Veterinaria. Universidad de Córdoba, Córdoba.

Pham, A. M. & Nakai, S. (1984). Application of stepwise discriminant analysis to high pressure liquid chromatography profiles of water extract for judging ripening of Cheddar cheese. *J. Dairy Sci.*, **67**(7), 1390–6.

Santa Maria, G., Ramos, M. & Ordóñez, J. A. (1986). Application of linear discriminant analysis to different proteolysis parameters for assessing the ripening of Manchego cheese. *Food Chem.*, **19**(3), 225–34.

Smith, A. M. & Nakai, S. (1990). Classification of cheese varieties by multivariate analysis of HPLC profiles. *Can. Inst. Food Technol. J.*, **23**(1), 53–8.