ELSEVIER

Food Chemistry 50 (1994) 389–392

0308-8146/94/$7.00

Differentiation of legumes through elemental chemical composition using factor analysis

E. Barrado, R. Pardo, B. Camarero

Departamento de Química Analítica, Facultad de Ciencias, Universidad de Valladolid, 47005 Valladolid, Spain

A. Tesedo & H. Romero

Departamento de Nutrición y Dietética, Hospital Clínico Universitario, Valladolid, Spain

(Received 9 June 1993; revised version received and accepted 23 September 1993)

The proximate composition (water, carbohydrates, proteins, fats and ash and also sodium, potassium and chloride ions) of some legumes in the diet of a clinical hospital, over a 10-year period, has been determined. From the average values, or their graphical representations, large differences between several kinds of legume can be observed, but these differences are clearer and better explained by chemometrics methods (factor analysis and varimax rotation) which produce groupings of the samples into three types (fresh, canned and frozen) according to two factors: nutrients and salt.

INTRODUCTION

The composition of food is determined by many factors: order, family, species, climate, age, soil composition, etc. According to Forina and Lanteri (1984) these parameters constitute the 'cause space' and determine the quality, price, taste, odour, etc, which constitute the 'space effects'. The chemical composition or 'chemical space' can be used for the interpretation of cause–effect relationships. This can be done by analysing all components of food, or by measuring a few components and using chemometric methods to obtain the maximum information from the data. There are many examples, such as those of Castino (1975) who uses linear statistical discriminant analysis (LSDA) to classify Italian wine, and of Forina and Tiscornia (1982) who apply some methods (SIMCA, etc.) to the geological classification of olive oils.

Modern methods of analysis, such as gas chromatography, HPLC and voltammetry, provide numerous parameters in only one determination. This has greatly contributed to knowledge of chemical composition, but requires the use of computerized multivariate methods (Martens & Harries, 1983).

On the other hand, the nutritional value of food products depends upon their composition in essential nutrients. In general, fresh legumes are important in alimentation because they offer a rather balanced composition of carbohydrates, proteins and fat (Mataix & Salido, 1985). Other products on the market (canned and frozen legumes) have been industrially treated and their composition and nutritive value might have been modified.

In this paper a chemometric elemental analysis study (determined over 10 years, 1981–1991) of a legume group included in the diet of a clinical hospital has been carried out. When these data are treated by 'principal component analysis', this shows the relationships and differences between three kinds of legume (fresh, canned and frozen) which are badly perceived or difficult to see by simple observation of the data or their graphic representation.

MATERIALS AND METHODS

The elemental analysis includes: % water (wind furnace method), % total protein (Kjeldahl method, conversion factor 6·25), total ash (%), carbohydrates (%) and % fat (AOAC, 1975), Na^+ and K^+ (ppm) by flame photometry and Cl^- (ISE).

All calculations were done on a PC computer using a home-made BASIC program available upon request.

RESULTS AND DISCUSSION

Table 1 shows the average results of 30 analyses carried out between 1981 and 1991, three each year, of different kinds of legumes:

- natural: chick-pea (cp), lentil (l), white bean (wb) and coloured bean (cb);
- frozen legumes: green bean (fgb), broad bean (fbb) and pea (fp);
- canned legumes: pea (cp) and green bean (cgb).

Table 1. Average results of 30 analyses

Legume	Water (%)	Carbh (%)	Prot (%)	Fat (%)	Ash (%)	Na^+ (ppm)	K^+ (ppm)	Cl^- (ppm)
cp	12·00	58·9	19·8	4·86	2·64	69·0	881	58·5
l	10·96	58·7	23·2	1·62	1·40	27·4	689	64·6
wb	14·90	60·5	20·3	1·87	1·49	54·1	1 160	25·9
cb	13·47	58·4	19·8	1·95	1·95	69·6	972	1·23
fgb	89·96	5·70	2·28	0·11	0·71	5·50	269	26·1
fbb	70·36	20·3	7·29	0·88	0·51	21·9	386	11·6
fp	83·69	9·84	4·50	0·35	0·17	1·63	341	33·1
cp	82·90	10·8	3·67	0·38	0·18	256	198	312
cgb	94·60	3·79	1·17	0·04	0·04	232	121	9·30

In each analysis water, carbohydrates (carbh), proteins (prot), fat and ash are in percent and sodium, potassium and chloride ion are in ppm.

By presenting the first five rows of Table 1 as a pie chart, some differences between one type of legume and the others can be observed. Figure 1 shows this for chick-pea and frozen broad bean. The sectors corresponding to water, carbohydrates and proteins differ considerably.

Another graphical representation groups together fresh legumes, on the one hand, and frozen and canned on the other, as shown in Fig. 2, in which the percentage of water for each legume is shown opposite the type of legume. However, in this Figure no differences were evident between frozen and canned legumes.

These differences are more clearly seen by chemometric treatment of the data, i.e. factor analysis and varimax rotation. In fact, if we take the legumes as objects and chemical compositions as variables, Table 1 can be treated by multivariate statistical methods. By factor analysis (FA) we can find an internal structure of the measurements not easily accessible from the original analysis, and explain the original results with a series of 'latent' factors, of fewer number than the original variables.

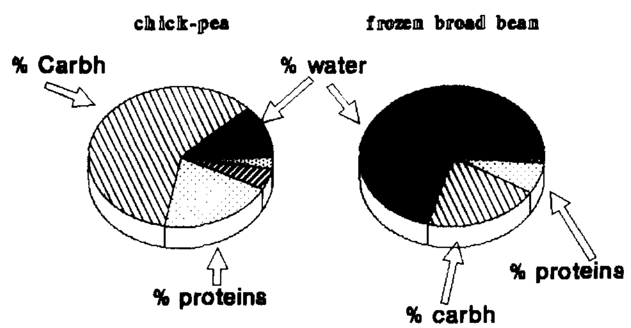

Fig. 1. Pie chart showing the differences in composition between a fresh and frozen legume.

First, we have applied FA to all the data sets shown in Table 1 that include the analyses of legumes. The numerical values of the variables are very different, since they express the compositions and because the values are expressed in different units of concentration (percentages and ppm). In this way the first step is a normalization of variables by autoscaling to unit variance. After that, the correlation matrix of the autoscaled variables can be obtained (Table 2).

The critical value of r for $P = 0{\cdot}05$ and seven degrees of freedom is 0·666, so all variables are strongly correlated, as can be seen in the table, except for sodium which only correlates with chloride ions.

The utility of carrying out an FA of the data set can be ascertained by means of Bartlett's sphericity test, based on the calculation of the statistics:

$$X^2_{\text{calc}} = -(\text{NOBJ} - 1 - (2\text{VA} + 5)/6) \ln |R|$$

where NOBJ and VA are the number of objects and variables, respectively, and R is the determinant of the correlation matrix and its comparison with the X^2_{crit} value for VA(VA − 1)/2 degrees of freedom and the required significance level. In our case, X^2_{calc} was 138·9 and $X^2_{\text{crit}} = 28{\cdot}9$ (28 degrees of freedom, $P = 0{\cdot}05$), so the null hypothesis of spherical distribution of the original variables can be rejected and the FA will provide a reduction in the dimensionality of the data set.

Table 3 shows the results of FA, which are based on extraction of the 'eigenvalues' and 'eigenvectors' of the

Table 2. Correlation matrix

	Water	Carbh	Prot	Fat	Ash	Na	K	Cl
water	1·000	−0·999	−0·995	−0·790	−0·898	0·325	−0·931	0·488
carbh		1·000	0·993	0·785	0·892	−0·318	0·943	−0·487
prot			1·000	0·749	0·862	−0·355	0·912	−0·497
fat				1·000	0·921	−0·211	0·723	−0·352
ash					1·000	−0·342	0·852	−0·515
Na						1·000	−0·388	0·950
K							1·000	−0·586
Cl								1·000

$r_{\text{critical}} = 0{\cdot}666$ ($P = 0{\cdot}05$, $\nu = 7$).

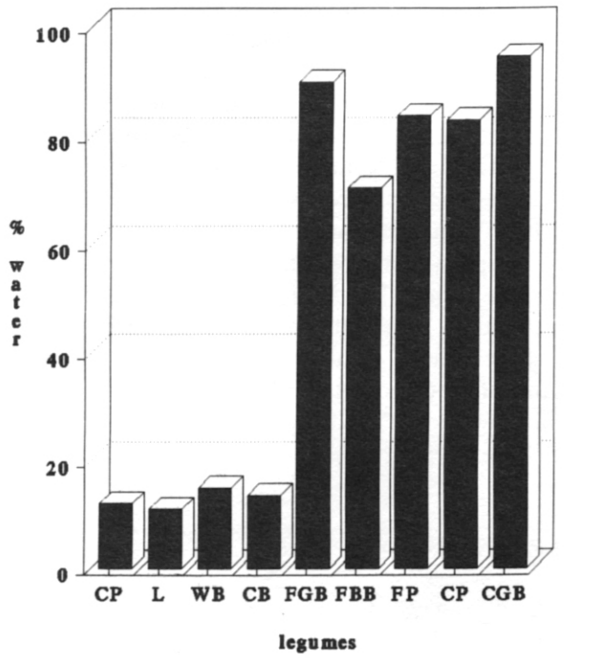

Fig. 2. Bar diagram showing the different percentages of water contained in several legumes.

correlation matrix. As can be observed, there are two significant factors (eigenvalues are greater than unity) which explains 92·71% of the variance and, therefore, most of the information contained in the original data set.

The new 'latent' factors are obtained by a linear combination of the original variables and their corresponding factor loadings (Table 4).

As can be observed, water, carbohydrates, proteins, fat, ash, and potassium mainly appear in the first factor, which explains 74·21% of variance. Sodium and chloride are found in the second factor (18·50% of variance).

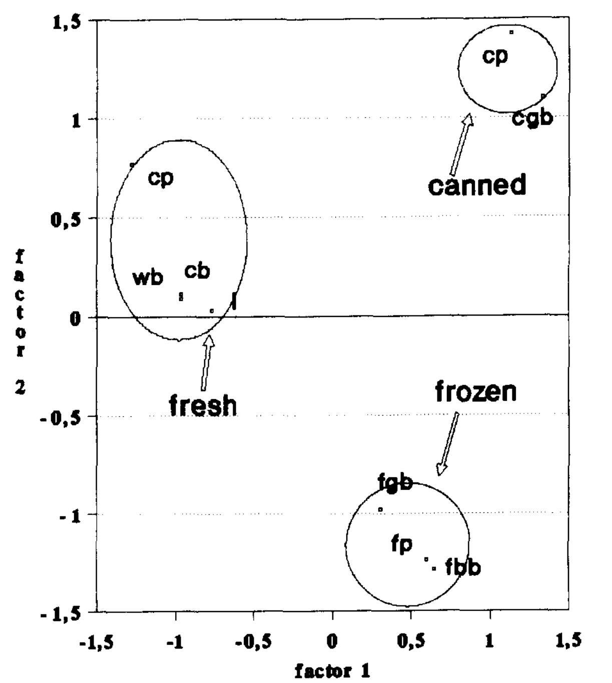

Fig. 3. Representation of the first two factors of the FA: Factor 1: 'nutrients', Factor 2: 'saline'.

Table 3. Factor analysis: eigenvalues

Value	Percent variance	Percent accumulated variance
5·937	74·21	74·21
1·489	18·50	92·71
0·407	5·05	97·76
0·372	1·46	99·22
0·097	0·60	99·82
0·046	0·18	100·00
0·001	0·00	100·00
0·000	0·00	100·00

Table 4. New latent factors

	Factors			
	1	2	3	4
H_2O	0·969	−0·180	0·149	0·079
Carbh	−0·968	0·184	−0·164	−0·045
Prot	−0·957	0·145	−0·204	−0·148
Fat	−0·836	0·272	0·462	−0·024
Ash	−0·940	0·150	0·253	0·041
Na	0·495	0·857	−0·051	0·109
K	−0·945	0·057	−0·180	0·255
Cl	0·653	0·747	−0·030	−0·085
Percent variance explained	74·21	18·50		

As the new factors have a greater amount of variance than the original values, a plot of these will correspondingly contain a greater amount of information. Figure 3 shows a graphic representation of the two first 'latent' factors. In the same way, the figure also contains 92·71% of the global information.

The appearance of three well-defined groups, fresh, canned and frozen, can be clearly observed. Moreover, these three groups appear to be classified according to 'factor 1' which is composed of the nutrients, but with a negative sign. This means that the more a legume is situated to the left, the more nutrients it contains.

A varimax rotation can clarify the above picture since it decreases the contribution of variables with

Table 5. Varifactor loadings

	Varifactor	
	1	2
H_2O	0·963	0·211
Carbh	−0·963	−0·207
Prot	−0·938	−0·238
Fat	−0·875	−0·075
Ash	−0·924	−0·228
Na	0·123	0·982
K	−0·893	−0·315
Cl	0·311	0·942
Percent variance explained	65·78%	26·92%
Total	92·70%	

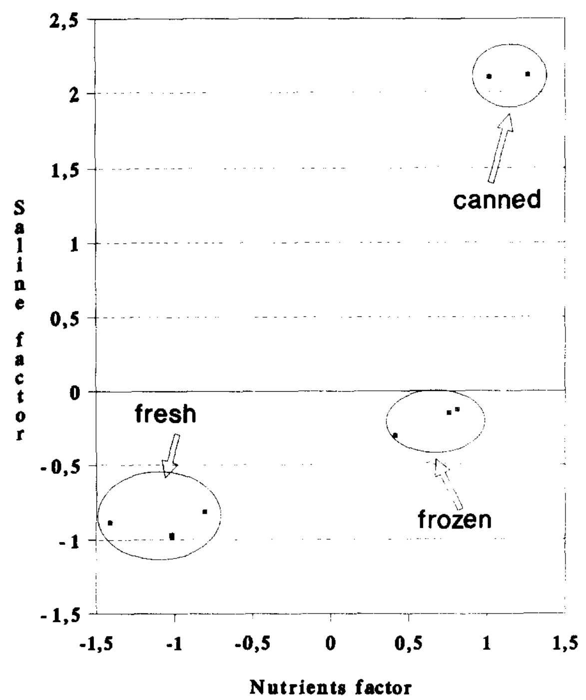

Fig. 4. Representation of the first two varifactors.

minor contributions and increases the more significant ones.

Table 5 shows the corresponding factor values and also the percentage of variance explained by each one. It can be observed again that the first two factors contain 92·70% of the original variance. The first factor is comprised of the nutrients (again with negative sign) and the second one is of sodium and chloride (saline factor).

In the graphical representation of two variables (Figure 4), the three well-defined groups mentioned above can be clearly observed.

Moreover, and confirming our initial supposition about nutritional differences between the three kinds of legume, fresh legumes appear to the left of the nutritional factors, because they contain the highest quantity of nutrients and the least water. Also, fresh legumes appear very low on the saline axis.

On the other hand, frozen legumes contain less salt and more nutrients than canned ones. The latter appear very high on the saline axis due to the extra amount of NaCl added as a preservative.

The nutritional and dietetic importance of these results is evident.

REFERENCES

AOAC (1975). *Official Methods of Analysis*, Washington DC, USA.

Castino, M. (1975). Use of discriminant analysis in the evaluation of wine type. *Atti, Accad. Ital. Vite Vino, Siena*, **27**, 109–21.

Forina, M. & Lanteri, S. (1984). Data analysis in food chemistry. In *Chemometrics. Mathematics and Statistics in Chemistry*, ed. B. R. Kowalski. D. Reidel Publishing Company, Dordrecht, The Netherlands, pp. 305–49.

Forina, M. & Tiscornia, E. (1982). Pattern recognition methods in the prediction of Italian olive oil origin by their fatty acid content. *Ann. Chim.*, **72**(3,4), 143–55.

Martens, H. & Harries, J. M. (1983). *Food Research and Data Analysis*. Applied Science Publ., Barking, UK.

Mataix, F. J. & Salido, G. M. (1985). *Importancia de las legumbres en la nutrición humana*. Fundación española de la Nutrición, Granada, Spain.