Use of Cluster and Factor Analysis for Grouping Bulgarians Honeys according to Botanical Origin

Kr. Nikolova[1], E. Pisanova[2] and Ir. Ivanova[3]

1. Department of Physics and Biophysics, Medical University, Marin Drinov 55, Varna, Bulgaria

2. Faculty of Physics, University of Plovdiv "P. Hilendarski", 24 Tzar Assen Str., Plovdiv, Bulgaria

3. Department of Mathematics, Computer Science and Physics, Agricultural University, 12, Mendeleev blvd., Plovdiv, Bulgaria

Abstract: Fourteen samples from acacia (Robinia pseudoacacia), lime(Tilia sp.), sunflower (Helianthus annuus), thistle (Onopordum acanthium), forest and honeydew honeys have been characterized according to their parameters of mineral and sugar content, color characteristics and refractive index. The samples have been grouped in two clusters–monofloral lime honey and monofloral acacia honey on the basis of the investigated characteristics by cluster analysis. The honey samples from sunflower and honey from Onopordum acanthium are incorrectly classified according to their botanical origin as lime honey while the honeydew and lime honey are determined correctly. By applying factor analysis the investigated indicators reduced to 6 factors explained 91.55 % from the entire variation. The parameters with the highest factor weight influence on the distribution of honey in clusters.

Key words: Honey, color parameters, cluster analysis.

1. Introduction

The complex estimation of the types of honey according to their botanical origin is based on the values of physicochemical parameters and it is possible by applying a cluster analysis. The conducting and the building-up of a dendrogram allows grouping of the studied samples according to their similarity or distinction on the base of investigated characteristics and gives the opportunity to discover relations between them. The parameters with the strongest influence on the distribution of the samples in the obtained clusters were determined by applying the factor analysis [1]. Similar classifications of honey have been made in Refs. [2-4].

The main objective of this research was to evaluate the similarity and the distance of the different types of Bulgarian honey and their grouping based on physicochemical parameters by using cluster analysis and determination of the strongest influenced

parameters of the distribution of the samples in the clusters, using factor analysis.

2. Materials and Methods

2.1 Samples

Fourteen types of bee honey from acacia (Robinia pseudoacacia), lime (Tilia sp.), sunflower (Helianthus annuus), thistle (Onopordum acanthium), forest and honeydew honey were purchased from the commercial network. The samples were stored at room temperature in dark.

3. Methods

All analysis were made three times according the recommendations of the International Honey Commission [5]. Type of honey, origin and other specifications are given below in Table 1. Each sample was described by 21 measured parameters listed in Table 2.

Determination of colour parameters and pigments The samples were heated to a temperature of 30-35 ℃ to dissolve the sugar crystals. After that they

Corresponding author: Anthony (aka Tony) Spiteri Staines, M.Sc., lecturer, research fields: petri nets, unified modeling language.

Table 1   Specification of the investigated honey samples.

№	Botanical origin	Region	Type
1	Honeydew	Elena	producers
2	Honeydew	Gabrovo	producers
3	Tilia sp.	Plovdiv	producers
4	Robinia pseudoacacia	Samolov	Super-market
5	Onopordum acanthium	Burgas	Super-market
6	Honeydew	Rodopi mountin	producers
7	Honewdew	Gavabovo	producers
8	Tilia sp.	Sopot	Super-market
9	Tilia sp.	Montana	producers
10	Robinia pseudoacacia	Sakar	producers
11	Robinia pseudoacacia	Stara Zagora	producers
12	Helianthus annuus	Plovdiv	supermarket
13	Robinia pseudoacacia	Montana	producers
14	Robinia pseudoacacia	Brezovo	producers

Table 2   Measured variables.

No	Code	Name
1	X1	Potassium K
2	X2	Sodium Na
3	X3	Lithium
4	X4	Copper Cu
5	X5	Iron Fe
6	X6	Manganese Mn
7	X7	Calcium Ca
8	X8	Magnesium Mg
9	X9	Refractive index n
10	X10	Water content W
11	X11	Beta carotene β
12	X12	Color coordinate x
13	X13	Color coordinate y
14	X14	Color parameter a
15	X15	Color parameter b
16	X16	Lightness L
17	X17	Glucose G
18	X18	Fructose F
19	X19	Sacharose S
20	X20	Glucose\fructose ratio F/G
21	X21	Oligosharides

were poured into a 10-mm thick dish. The color parameters in two colorimetric systems XYZ and SIE Lab have been measured by Lovibond PFX 880 (UK) [6].

Determination of water content The refractive index of the honey samples were measured using an Abbe refractometer (Carl Zeiss, Germany) at 20 ℃. The equivalent water content was determined using a table provided in the Official Methods of Analysis [7].

Determination of sugar content After filtration of the honey solution the sugar content was determined by liquid chromatography with Differential Refractometer R401 Detector (Waters, Amsterdam)., and a Column-Aminex HPX-87H. The temperatures of both column and detector were 30 ℃. The volume of the sample injected was 10 μL and the ratio was 0.5 mL/min.

Determination of the mineral content The content of Li, Na, K, Ca, Mg, Cu, Fe, and Mn in the honey samples was determined by flame atomic absorption spectrometry using a Thermo M5 instrument (UK) in an air-acetylene flame under standard conditions. The amount of 3 g of honey was dissolved in 100 mL of distilled water and the obtained solutions were analysed. Aqueous standard solutions of the analytes at the mg l–1 level, prepared from Titrisols (Merck, Germany), were used for calibration. All values of the tested parameters were given after averaging lead after three times repetition of the experience.

4. Mathematical and Statistical Methods

The mathematical processing of the data was performed by cluster analysis and factor analysis using the method of PCA (principal components analysis). The cluster analysis has been applied on the base of all investigated parameters for comparing and grouping the honeys. The grouping of the different honey samples in clusters has been made by hierarchical clustering applying the method of average linkage between groups. Euclidean distance has been used for determination of the similarity between objects [8-10]. In order to avoid differences of the dimensions the data were standardized in advance. The result of the clustering has been presented graphically via dendrogram. A correlation analysis has been made to determine the connection between investigated parameters. The PCA was used on the base of the obtained correlation matrix. Thus the input set of correlating data was transformed into a new set

with a smaller number of uncorrelated artificial variables, so-called "factors" or "principal components", which explain the greatest part of data variation. The reduction in the number of original variables was achieved by grouping the correlated variables in a common factor and the separation of uncorrelated ones in different factors [1, 11]. A stronger distinction of the variables to one or another factor is performed by further factor rotation using Varimax method with Kaiser normalization. The processing was made by using the statistical program SPSS [9].

5. Results and Discussion

The basic optical parameters of different types honey were investigated, the average value and standard deviation are presented in Table 3. Our research also focuses on some trace elements and a database from trace elements contents of the honeys was created (Table 4).

The grouping of the investigated 14 Bulgarian types of honey in the distinct clusters is shown by dendrogram in Fig. 1 and the average linkages between groups have been presented in Table 5. The samples were grouped into two main clusters as a result of the cluster analysis performed. The results are presented in table with the steps of combining the clusters and the intergroup distances.

The first cluster was more homogeneous and united two sub clusters. The first sub cluster involved

Table 3   Average value and standard deviation for optical parameters of honey samples.

№		n	W, %	Carotene, ppm	x	y	a	b	L
1.	X.	1,4961	16.2	65.23	0.5121	0.4597	14.75	84.26	61.98
	SD	$\pm 2 \cdot 10^{-4}$	0.1	0.45	0.0021	0.0097	1.12	1.13	1.16
2	X	1,489	19	27.68	0.4314	0.4453	-1.45	63.63	80.29
	SD	$\pm 1 \cdot 10^{-4}$	0.2	0.32	0.0012	0.0056	0.07	1.14	3.23
3.	X	1,4932	17.2	39.78	0.4616	0.4703	0.25	80.77	77.09
	SD	$\pm 4 \cdot 10^{-4}$	0.1	1.02	0.0013	0.0023	0.05	2.34	2.13
4.	X	1,495	16.6	8.21	0.3528	0.3687	-3.37	24.53	87.29
	SD	$\pm 5 \cdot 10^{-4}$	0.2	0.89	0.0060	0.0012	0.03	1.23	1.76
5.	X	1,494	17	37.33	0.4553	0.4701	1.46	78.23	77.15
	SD	$\pm 7 \cdot 10^{-4}$	0.1	0.23	0.0024	0.0012	0.01	2.35	1.87
6.	X	1,491	18.2	92.53	0.6115	0.3838	29.67	39.17	23.08
	SD	$\pm 4 \cdot 10^{-4}$	0.2	1.16	0.0012	0.0012	1.56	1.16	1.23
7.	X	1,4955	16.4	40.93	0.469	0.4667	3.2	76.25	70.42
	SD	$\pm 3 \cdot 10^{-4}$	0.3	2.43	0.0012	0.0017	0.07	2.54	1.76
8.	X	1,495	16.6	30.16	0.4384	0.4553	-2.35	69.09	80.16
	SD	$\pm 2 \cdot 10^{-4}$	0.1	2.65	0.0045	0.0034	0.05	3.07	3.78
9.	X	1,497	15.8	21.44	0.4105	0.4263	-2.21	50.87	72.25
	SD	$\pm 1 \cdot 10^{-4}$	0.3	1.44	0.0078	0.0067	0.03	2.89	4.12
10.	X	1,4945	16.8	10.31	0.3635	0.3827	-4.47	31.1	87.27
	SD	$\pm 1 \cdot 10^{-4}$	0.2	1.32	0.0023	0.0098	0.02	1.13	2.34
11.	X	1,4933	17.2	16.08	0.4025	0.4296	0.24	1.39	3.47
	SD	$\pm 4 \cdot 10^{-4}$	0.1	0.98	0.0015	0.0076	0.02	0.09	0.21
12	X	1,4935	17.2	17.36	0.3949	0.4221	-6.62	49.72	85.72
	SD	$\pm 3 \cdot 10^{-4}$	0.1	1.17	0.0036	0.0023	0.01	1.07	0.78
13.	X	1,498	15.5	21.8	0.4183	0.423	0.73	29.1	37.38
	SD	$\pm 4 \cdot 10^{-4}$	0.1	2.65	0.0023	0.0023	0.02	0.23	1.23
14	X	1.4923	17.6	18.64	0.403	0.4129	-0.29	29.86	46.56
	SD	$\pm 2 \cdot 10^{-4}$	0.2	2.34	0.0045	0.0056	0.04	0.54	1.45

Table 4　Average value and standard deviation for mineral content of honey samples.

		K, $\mu g\ g^{-1}$	Na, $\mu g\ g^{-1}$	Ca, $\mu g\ g^{-1}$	Mg, $\mu g\ g^{-1}$	Fe, $\mu g\ g^{-1}$	Cu, $\mu g\ g^{-1}$	Mn, $\mu g\ g^{-1}$	Li, $\mu g\ g^{-1}$
1.	X.	1230	34.59	52.2	11.59	3.05	1.34	76.75	36.08
	SD	12	1.45	0.98	0.76	0.09	0.09	2.43	1.34
2	X	1210	12.26	0.1	0.66	1.24	1.79	84.22	25.64
	SD	23	1.12	$\pm 1 10^{-2}$	0.01	0.01	0.04	1.23	2.67
3.	X	290	13.6	0.1	0.42	3.47	0	46.41	11.52
	SD	5.3	1.23	$\pm 1 10^{-2}$	0.02	0.03	0	2.56	3.90
4.	X	250	62.12	0.1	0.34	0.31	0	46.9	13.07
	SD	6.2	2.45	$\pm 1 10^{-2}$	0.01	0.02	0	3.34	2.67
5.	X	360	11.77	0.1	0.09	1.89	0	94.57	17.42
	SD	4.6	1.09	$\pm 1 10^{-2}$	0.01	0.02	0	2.67	1.65
6.	X	1640	22.25	0.57	0.99	5.01	15.58	127.65	109.1
	SD	25	2.57	0.08	0.09	0.12	0.76	7.90	9.56
7.	X	1820	25.4	0.1	0.19	0	0.36	99.32	32.58
	SD	32	2.45	$\pm 1 10^{-2}$	0.01	0	0.04	9.89	1.45
8.	X	400	14.95	0.1	0	0.93	0.07	82.07	21.43
	SD	3.7	1.23	$\pm 1 10^{-2}$	0	0.06	0.02	6.89	1.89
9.	X	380	15.76	0.1	1.29	3.28	0.27	64.68	11.91
	SD	2.9	1.23	$\pm 1 10^{-2}$	0.01	0.05	0.08	4.34	1.56
10.	X	210	9.82	0.1	0.9	2.22	0.01	42.22	6.85
	SD	1.3	1.54	$\pm 1 10^{-2}$	0.01	0.02	0	2.45	0.09
11.	X	280	9.48	0.1	0.76	0.31	0	50.85	9.57
	SD	2.6	1.34	$\pm 1 10^{-2}$	0.01	0.01	0	2.12	0.9
12	X	260	10.2	0.1	1.06	3.33	0	56.79	11.04
	SD	2.3	0.98	$\pm 1 10^{-2}$	0.02	0.23	0	2.45	0.45
13.	X	300	20.61	0.1	2.26	1.05	0	33.07	7.49
	SD	1.7	1.09	$\pm 1 10^{-2}$	0.12	0.06	0	1.34	1.12
14	X	260	15.32	0.1	1.74	0.56	0	22.83	5.19
	SD	1.7	2.12	$\pm 1 10^{-2}$	0.23	0.02	0	2.12	0.76

samples 8, 9, 3, 5, 12 and the second—samples 11, 14, 13, 10. Several types of lime honey, sunflower honey and thistle are involved in the first sub cluster. In the lime honey samples the content of glucose was between 40% and 41%, the fructose content was between 47% and 51%, β-carotene content—between 20 ppm and 40 ppm. The second cluster included acacia honeys which had lower mineral content and β-carotene. The second one involves samples 2, 7, 4, 1 and 6, i.e. honeydew and forests honeys rich in minerals and vitamins, with relatively low values for lightness. Some of the samples have very high water content.

The distribution of the types of honey according to cluster analysis was presented in Table 6. For the classifying parameter "Botanical origin" the samples of sunflower honey and those of thistle honey are classified wrongly as monofloral honey with predominant lime content. The honeydew and forest honey samples were identified fully. One of the honey samples with dominant content of acacia was classified as honeydew, the others were determined correctly.

The results from the factor analysis are presented graphically in Fig. 2. On the base of the obtained results the number of the principal components was determined to be 6 from the total 21 parameters investigated. The percentage of the total variance due
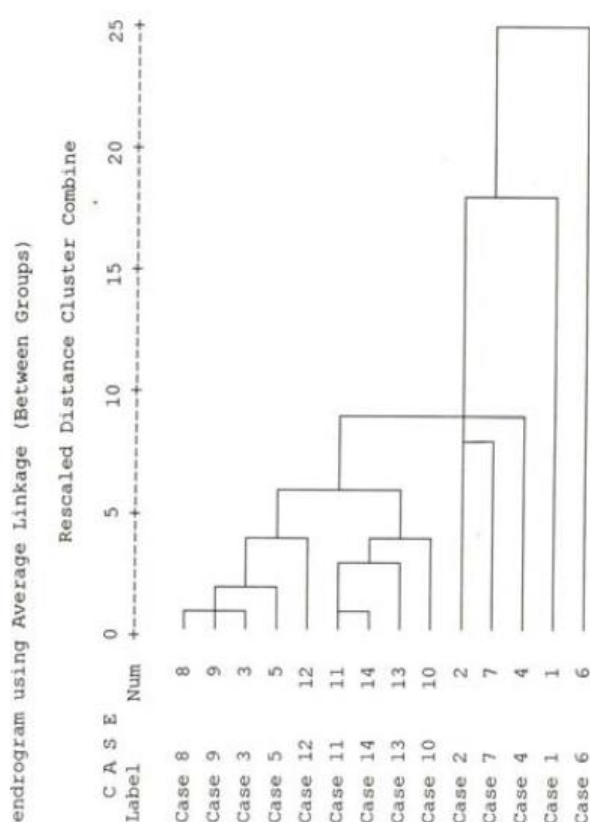
The third one contained oligosaharides, Li and Cu.

These parameters did not differ significantly between groups according to the botanical origin. The index of refraction, water content and the content of fructose included in the fourth factor. The biggest factor weight is the water content, responsible for the development of micro-organisms in the product. These parameters are related to the kinetics of the processes in honey during storage. The fifth and the sixth parameter unite the colorimetric characteristics and the brightness of the samples.

In conclusion it could be said that the classification and grouping of the samples by hierarchical cluster analysis according to their botanical origin allows increasing the objectivity of their assessment. Our results are in accordance with the data found in the literature [12]. According to Refs. [13, 14] it is possible to determine the botanical origin, based on the percentage of sugars.

Fig. 1   Dendrogram samples of honey based on the average intergroup Euclidean distances.

to an appropriate factor is presented in the rotated by Varimax method factor matrix (Table 7). This method allowed more precise interpretation of the factors. The first factor united some of the elements: K, Ca, Na, Mn, the content of the β-carotene and color coordinates a, x. The Mg and β-carotene posses the greatest weight factor.

The second one combined sugars such as sacharosse, glucose, fructose-glucose ratio and the element iron. The F\G ratio had the greatest weight factor which is crucial for the crystallization of honey.

Table 5   Combining of clusters and intergroup distances.

Step #	Combination of clusters		Coefficient
	Cluster 1	Cluster 2	
1	8	9	6.92
2	11	14	7.67
3	3	8	9.24
4	3	5	13.43
5	11	13	16.19
6	3	12	17.32
7	10	11	19.51
8	3	10	24.76
9	2	7	30.82
10	2	3	35.81
11	2	4	37.13
12	1	2	67.58
13	1	6	92.19

Table 6   Recognizability of samples using hierarchical cluster analysis according to the botanical origin.

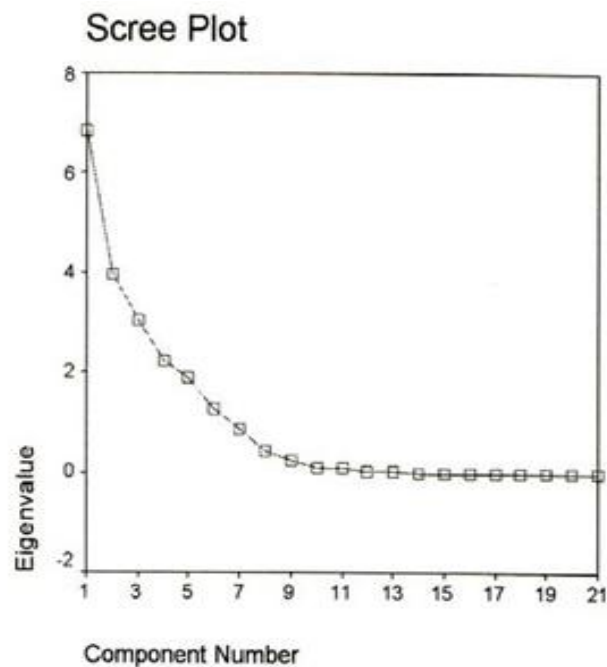|		Using cluster analysis recognized as					
		Acacia	Lime	Honeydew, forest	Sunflower	Thistle	
	Acacia	4		1			5
	Lime		3				3
Studied class	Honeydew, forest			4			4
	Sunflower		1				1
	Thistle		1				1
		4	5	5	0	0	

Fig. 2 Factor analysis by the method of principal components. Values of the eigenvalues of the correlation matrix.

Table 7 Rotated factor matrix determined by varimax transformation of the main components.

#	Parameter	Main Components					
		1	2	3	4	5	6
1	X1	0.80	0.41	0.13	-0.15	0.21	9.5×10^{-2}
2	X2	0.12	0.45	0.38	0.27	-0.49	0.32
3	X3	0.21	-7.30×10^{-3}	0.92	0.16	0.17	7.21×10^{-2}
4	X4	0.15	-1.77×10^{-2}	0.94	0.17	0.10	-4.58×10^{-2}
5	X5	0.53	-0.70	0.17	-4.38×10^{-2}	-7.31×10^{-2}	-7.61×10^{-2}
6	X6	0.88	-0.13	-6.42×10^{-2}	-0.22	-0.33	-0.17
7	X7	0.86	-6.05×10^{-4}	-0.16	-8.4×10^{-2}	0.23	0.24
8	X8	0.97	-4.23×10^{-2}	1.64×10^{-2}	-0.15	-0.17	-5.56×10^{-2}
9	X9	-0.22	3.21×10^{-2}	0.16	0.95	3.82×10^{-2}	-1.91×10^{-3}
10	X10	0.24	-2.36×10^{-2}	-0.16	-0.95	-4.71×10^{-2}	1.46×10^{-2}
11	X11	0.94	-0.11	0.24	-3.88×10^{-2}	0.16	-6.87×10^{-2}
12	X12	0.94	-0.10	0.15	-6.10×10^{-2}	0.23	-0.15
13	X13	8.78×10^{-3}	1.13×10^{-2}	7.81×10^{-2}	4.75×10^{-2}	0.98	4.84×10^{-2}
14	X14	0.92	-1.69×10^{-2}	0.26	-1.93×10^{-2}	-0.11	-0.24
15	X15	0.30	-8.14×10^{-2}	0.19	4.35×10^{-2}	0.72	0.56
16	X16	-0.29	-8.85×10^{-2}	-2.28×10^{-2}	1.13×10^{-2}	0.11	0.93
17	X17	5.57×10^{-2}	-0.93	-0.21	0.14	5.81×10^{-2}	2.51×10^{-3}
18	X18	3.20×10^{-2}	0.49	-0.38	0.57	-0.16	0.12
19	X19	0.12	0.67	7.80×10^{-2}	0.23	8.29×10^{-2}	-0.29
20	X20	-4.25×10^{-2}	0.95	8.67×10^{-3}	6.79×10^{-2}	-0.10	8.11×10^{-2}
21	X21	2×10^{-3}	0.21	0.96	-8.32×10^{-2}	-0.12	-4.81×10^{-3}
	% from total variation	30.17	16.20	15.67	11.67	10.35	7.49

6. Conclusion

The cluster analysis can be used for the quick differentiation of monofloral from the polyfloral and honeydew samples. The factor analysis gives the opportunity to determine the parameters with the greatest weight factor which influence the distribution of the most spread honeys in Bulgaria in distinct clusters.

References

[1]   Gorsuch, R. 1983. *Factor Analysis*, Second Edition. Lawrence Erlbaum Assoc., Inc., Publishers New Jersey.

[2]   Popek, S. 2002. "A Procedure to Identify a Honey Type." *Food Chemistry* 79.

[3]   Anklam, E., and Radovic, B. 2001. "Suitable Analytical Methods for Determining the Origin of European Honey." *American Laboratory*.

[4]   Devillers, J., Morlot, M., and Pham-Delegue, J. C. D. 2004. "Classification of Monofloral Honeys Based on Their Quality Control Data." *Food Chemistry* 86.

[5]   Bogdanov, S., Martin, P., and Lullmann, C. 1997. "Harmonized Methods of the International Honey Commission." *Apidologie*.

[6]   Commission   Internationale   de   I   Elairage   Recommendations on Uniform Color Spaces. Color Difference Equations. Psychometric Color Terms. CIE Publication no 15 (F. 1. 3. 1.) 1971. Supplement 2. Bureau central de la Commission Internationale de I Eclairage. Vienna, 1978.

[7]   Official Methods of Analysis. Association of Official Analytical Chemist, Arlington, VA, 1995.

[8]   Duran, B., and Odelle, P. 1977. *Cluster Analysis*. Moscow, Russia.

[9]   SPSS for Windows. Base System User's Guide.

[10] Ward, J. H. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58.

[11] Kline, P. 1994. *An Easy Guide to Factor Analysis*. London: Routledge.

[12] Mateo, R., and Bosch-Reig, F. 1997. "Sugar Profiles of Spanish Unifloral Honeys." *Food Chemistry* 60.

[13] Terrab, A., Diez, M., and Heredia, F. 2002. "Characterizations of Maroccan Unifloral Honeys by Their Physicochemical Characteristics." *Food Chemistry* 79.

[14] Terrab, A., Vega-Perez, J. M., Diez, M., and Heredia, F. 2002. "Characterization of Northwest Maroccan Honeys by Gas Chromatographic-Mass Spectrometric Analysis of Their Sugar Components." *Journal of the Science Food and Agriculture* 82.