Cluster Analysis of the Properties of Wines Using Various Parameters

Krastena Nikolova[1], Ivan Maslinkov[2], Irena Ivanova[3]

[1]Faculty of Economics at University of Food technologies
Plovdiv, Bulgaria
kr.nikolova@abv.bg

[2]Technical Faculty at University of Food technologies
Plovdiv, Bulgaria

[3]Faculty of Iconomics at Agricultural university
Plovdiv 4000, Bulgaria
irenaiv@au-plovdiv.bg

ABSTRACT: *In this work we basically intended to do a few assessments of wines that generate results with the help of statistical processing. We have studied the – color coordinates, lightness, Hue angle, chroma and transmission coefficient parameters. We have used cluster analysis for the study of the parameters and variables. We have clustered the samples into two principal groups with the use of dendrogram. It contains both the samples of white wines, and the combination of red wines and separates them by kind of grape, production method and region of growing the grapevine. The made classification allows increasing of the objectivity of the evaluation and uses optical properties in various selection programs. A factor analysis using PCA was applied on the basis of the found correlation matrix. Thus, the number of the studied indices was reduced to 2 factors, which explained 93.32% from the entire variation. The results address the color parameters a and b, lightness L, Hue angle H in SIELab and transmission coefficient at four different wavelength. We conclude that these properties are significant to understand the nature of wines.*

Keywords: Wine, Optical Properties , Cluster Analysis, Factor Analysis, PCA, Dendrogram

Received: 26 August 2021, Revised 10 November 2021, Accepted 13 December 2021

DOI: 10.6025/jio/2021/11/4/102-108

1. Introduction

Wine is a fermented alcoholic beverage containing various compounds of different type, with polyphenols which are one of the most important components influencing their quality parameters [1]. Polyphenols contribute to the organoleptic characteristics, such as color, astringency and bitterness and also exert antimicrobial and antioxidant properties [2].

Wine is a fermented alcoholic beverage containing various compounds of different type, with polyphenols which are one of the most important components influencing their quality parameters [1]. Polyphenols contribute to the organoleptic characteristics, such as color, astringency and bitterness and also exert antimicrobial and antioxidant properties [2].

Authenticity and commercial value of wines is often linked to its geographical origin and certain countries or regions are known for producing excellent wines of high commercial value [3]. Due to its composition and worldwide availability, the controlled denomination of origin (CDO) is usually required to demonstrate the provenance of wines. In this work the optical properties of wines are investigated and the following goals are presented: using mathematical and statistical methods to group different types white and red wines on the base of investigated physico- chemical parameters and to identify those affecting most heavily on the clustering.

The same technics are used for estimation the European kinds of wine by factor and cluster analysis [4-5].

2. Matherials and Methods

The study involves 17 types of wines commercially available or obtained from a winery in the region of Starosel village. The specification of the tested wines is presented in Table 1.

	Place	**Type**
1	Wine-cellar	Cabernet Sauvignon
2	Wine-cellar	Rkaciteli
3	Wine-cellar	Misket
4	Wine-cellar	Merlot
5	Wine-cellar	White wormwood wine
6	Wine-cellar	Red wormwood wine
7	Wine-cellar	Rosé
8	Wine-cellar	Cabernet Sauvignon
9	supermarket	Gamza
11	Supermarket	Red wine
12	Supermarket	Vechernij zvon
13	Supermarket	Mehandjijsko
14	Supermarket	Mavrud
15	Supermarket	Trakia
16	Supermarket	Manastirsko shushukane
17	Supermarket	Assenitza

Table 1. Investigated Samples

The following 14 physical parameters have been investigated: color coordinates (1) X , (2) Y, (3) Z, (4) x, (5) y, lightness (6) L, color parameters (7) a, (8) b, (9) Cab, Hue angle (10) H, transmission coefficients at 445 nm (11), 495 nm (12), 550 nm (13), 625 nm (14).

Using a software package VISIONlite ColorCalc for spectrophotometer Helios Omega with a cuvette of a 10 mm length, the color parameters in CIELab colorimetric system have been obtained. All measurements have been carried out at room temperature.

Color coordinates, color parameters *a, b* and brightness *L* of tested samples have been measured.

Parameters such as chroma Cab and hue angle *H* were defined as follows:

$$C_{ab} = \sqrt{a^2 + b^2}\ , \quad h_{ab} = arctg\left(b/a\right) \tag{1}$$

To compare and group the types of wines on the base of all investigated physical parameters a hierarchical cluster analysis witha measure of similarity Euclidean distance [6, 7] have been used. The data have been preliminary standardized in order to avoid the influence of different dimensions.

A factor analysis was conducted. 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 the 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. A stronger distinction of the variables to one or another factor is performed by further factor rotation using Varimax method.

All data analyses have been performed using the SPSS software program.

3. Results and Discussion

17 samples of wine have been analyzed in regard to 14 physical parameters. Using cluster analysis, the studied wines have been divided according to their similarity and distance depending on the studied physical parameters.The grouping of the samples in clusters is presented with the dendrogram on Figure 1 and the values of the distancesbetween groups is shown in Table 2.

Steps No	Combined clusters		Coefficients
	Cluster 1	Cluster 2	
1	14	15	0,117
2	1	4	0,227
3	10	16	0,284
4	10	13	0,356
5	2	3	0,637
6	10	12	0,783
7	6	9	0,929
8	11	17	1,191
9	1	7	1,236
10	2	5	1,249
11	6	14	2,568
12	6	10	3,490
13	6	11	6,948
14	2	8	7,877
15	1	6	9,875
16	1	2	58,220

Table 2. Combining of Clusters and Distances between Group

The results from the investigation show that the wines are divided into two main clusters. The first of them includes 13 samples and is divided into some sub-clusters. The closest are the samples 14 and 15- red wormwood wine and gamza. Samples 10, 16, 13 and 12 form another sub-cluster. They are combined from the fact, that all are produced from the same variety of grape – Cabernet Sauvignon. The samples 1,4 and 7- red and pink wines, united in other sub cluster obtained from the wine cellar. The samples 11 and 17 are semi-dry wines and this is the reason of their similarity.

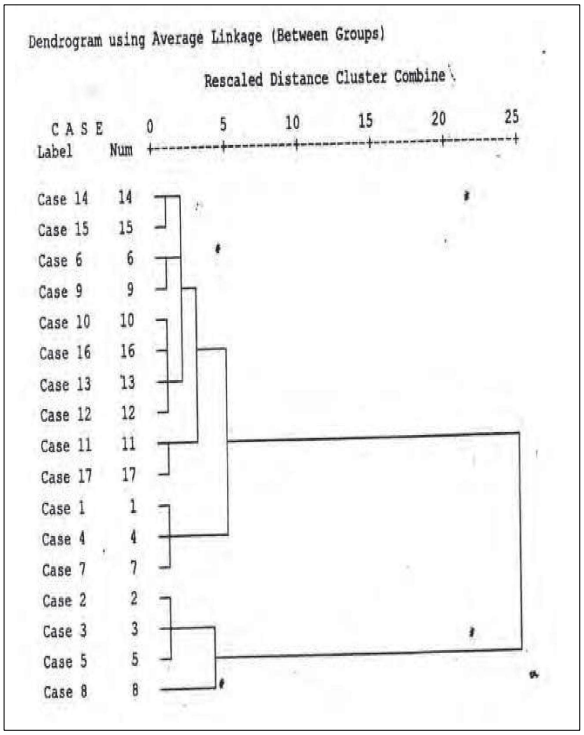

Figure 1. Dendrogram based on the hierarchical cluster analysis

The second cluster combines samples of the white wines 2, 3, 5 and 8, which have been obtained directly from the winecaller. The distance between 2 and 8 is the greatest in this cluster - 7,88.

The percentage distribution of the wine in the distinct subcluster is presented in table 3.

No	Samples in clusters	Number of samples	Distribution, %
1.1	14, 15	2	11,8
1.2	6, 9	2	11,8
1.3	10,16, 13,12	4	23,5
1.4	11, 17	2	11,8
1.5	1, 4, 7	3	17,6
2	2, 3, 5,	8 4	23,5
Total		17	100 %

Table 3. Distribition of the Wine Samples in Clusters

The samples from white wine and the second sub-cluster of the first one have the highest percentage of the samples - 23.5%. The wines from the first group are characterized with high lightness between 93-98 relative units, Hue angle is in the range between 80.0 and 87.0, these samples have high transmission coefficient at wavelengths 445 nm and 625 nm, which are 66 % - 87 % and 83% - 93% respectively. The yellow color predominates over the blue, but it is very weakly expressed – the color parameter "b" is between 5 and 12 . The wines which are from obtained from the wine-cellar possess the smallest Hue angle – between 16.0 and 27.0, the lowest lightness 5-13 relative units and their transmission coefficient is less than 1% at 445 nm, it is from 1 % to 8 % at 625 nm. The red component prevails over the yellow ($a>b$). The red Wormwood wine and the Gamza are characterized with almost the same lightness L and chroma Cab. The values of the red and yellow components are similar and they are almost equal $a \sim b$. The sub-cluster with the wines obtained from the supermarket has lightness in the range from 15 to 23, Hue angle from 41.0 to 49.0. The lightness for the Mavrud and Trakia wines is close to this of the red wormwood wine and Gamza - at about 30 a.u., Hue angle is between 53.0 and 54.0. However, the yellow component dominates the red ($b>a$) for them.

It is known that the investigated parameters are involved with different weight. To determine the influence of parameters in the split into clusters factor analysis with the method of principal components (PCA) has been performed. In principle the component analysis helps to identify variables and their relative importance in explaining the observed variation and types of wine. Data was found suitable for PCA and the factor analysis gives the correct results in our investigation, because the value of KMO-test is 0,705 (> 0, 5) and Bartlett's test of sphericity has a significance level lower than 0.001. The results of PCA are presented at Figure 2. Two principle components out of the fourteen have eigenvalued greater than one and explained 93,32 % of total variation. That is why we choice two factors. The first factor (principle component one - PC 1) explains 81, 57 % of total variation and strongly correlated with color parameters in color system CIELab and transmission coefficients in visible area. The color coordinates *x, y* in XYZ colorimetric system and chroma Cab in SIELab colorimetric system mainly contributed towards the second factor (principle component two - PC2), which explains 11,75 % of total variation. This factor integrates color coordinates in the color space for a small color differences SIELab and spectral characteristics in the visible area PC 2 gives the relation between some parameters from different color systems XYZ and SIELab. Color coordinates *x, y* determines one point in the plane which is connected to the characteristics of the colorbrightness, dominant wavelength and purity. These characteristics correlated with the Chroma – parameter in SIELab color system. The rotated component matrix, obtained with varimax transformation of the main components is presented in table 4. The results of the factor analysis are shown on Figure 2.

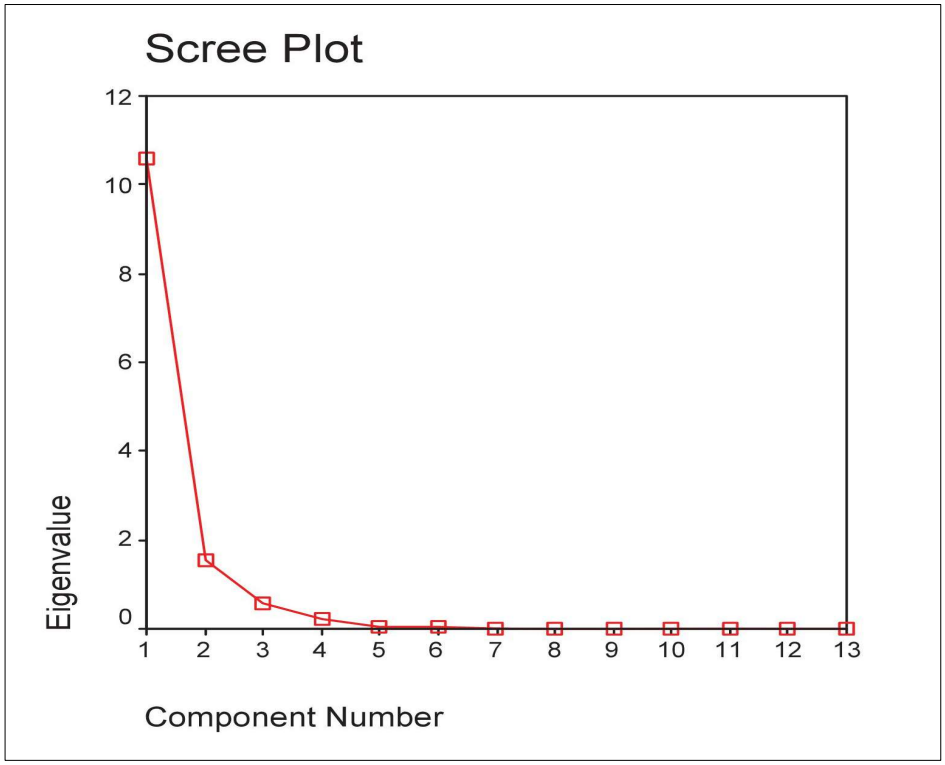

Figure 2. Factor analysis

No	Parameters	Principal Components	
		1	2
1	X1	0.711	0.699
2	X2	0.759	0.640
3	X3	-0,657	-0,729
4	X4	0,044	0,828
5	X5	0,549	0,823
6	X6	-0,887	-0,418
7	X7	-0,970	0,053
8	X8	-0,967	-0,185
9	X9	0,158	0,876
10	X10	0.765	0.623
11	X11	0.760	0.633
12	X12	0,761	0,632
13	X13	0,760	0,590

Table 4. The rotated component matrix with varimax transformation of principal components

4. Conclussions

The research allows the following conclusions: Knowing the spectral and color characteristics of the wines allow the separation of the samples according to the predominant grape variety or the place of purchase - from a winery or from a supermarket.

The most distinguishing samples according to the angle of Hue - with the smallest values are samples obtained from the wine cellars, followed by those obtained from the supermarkets and which are not semi-dry wines.

References

[1] Fernondez-Pachyn, M. S., Villaco, D., GarcHa-Parrilla, M. C., Troncoso, A. M. (2004). |Antioxidant activity of wines and relation with their polyphenolic composition, *Analytica Chimica Acta*, 513, 113–118.

[2] Godoy-Navajas, J., Aguilar-Caballos, M. P., Gymez-Hens, A. (2015). Automatic determination of polyphenols in wines using laccase and terbium oxide nanoparticles, *Food Chemistry*, 166, 29–34.

[3] Selih, V. S., Sala, M., Drgan, V. (2014). Multi-element analysis of wines by ICP-MS and ICP-OES and their classification according to geographical origin in Slovenia, *Food Chemistry*, 153, 414–423.

[4] Pges Escofier. (1994). Multiply factor analysis, *Computational Statistics and data analysis*, 18, 121-140.

[5] Katsota Arvanitoyannis. (1999). Aplication of quality control methods for assessing wine authensity, *Food science technology*, 10, 321-336.

[6] Duran, B., Odelle, P. (1977). *Cluster analysis*, Moscow, Statistics.

[7] Ward, J H. (1963). Hierarchical grouping to optimize all objective functions, *Journal of the American Statistical Association*, 58, 236-244.