

## *Cluster analysis of coffee blends for some sensory properties: a comparative approach to the ABIC's classification criteria*

### *Análise de clusters de blends de café para algumas propriedades sensoriais: uma abordagem comparativa aos critérios de classificação da ABIC*

Daiane de Oliveira Gonçalves<sup>1</sup>; Mariana Resende<sup>2</sup>; Natalia da Silva Martins<sup>3</sup>; Flávio Meira Borém<sup>4</sup>; Marcelo Angelo Cirillo<sup>5</sup>

#### **Abstract**

In Brazil, coffee beverage quality is classified according to technical recommendations of the Associação Brasileira da Indústria de Café (ABIC), which determines cutoff points to discriminate from non-recommended to gourmet coffees. Accordingly, this study aimed to propose the use of cluster analysis to evaluate coffee blends composed of coffees with different qualities and of different varieties regarding a few sensory properties, using continuous and binary scales obtained by a cutoff, which defines whether the coffee is recommendable or not according to the ABIC criteria. It is believed, therefore, that this technique can be used to analyze coffee beverage quality as it is easily accessible and implemented by researchers. In conclusion, a qualitative cluster analysis using the minimum cutoff value of the ABIC had more promising results. This is because blends whose composition contained high and moderate proportions of specialty coffees were more homogeneous in the clustering.

**Keywords:** Grouping. Specialty coffees. Dendrogram. Cutoff score. Taster.

#### **Resumo**

No Brasil, a qualidade da bebida do café é classificada de acordo com recomendações técnicas da Associação Brasileira da Indústria de Café (ABIC), que determina pontos de corte para discriminar entre não recomendados e cafés gourmet. Nesse sentido, o presente estudo teve como objetivo propor o uso da análise de agrupamento para avaliar blends de café compostos por cafés com diferentes qualidades e de diferentes variedades quanto a algumas propriedades sensoriais, utilizando escalas contínuas e binárias obtidas por um ponto de corte, que define se o café é recomendável ou não de acordo com os critérios da ABIC. Acredita-se, portanto, que essa técnica pode ser utilizada para analisar a qualidade da bebida de café por ser de fácil acesso e implementada por pesquisadores. Em conclusão, uma análise qualitativa de cluster usando o valor de corte mínimo da ABIC teve resultados mais promissores. Isso porque os blends cuja composição continha proporções altas e moderadas de cafés especiais foram mais homogêneos no agrupamento.

**Palavras-chave:** Agrupamento. Cafés especiais. Dendrograma. Nota de corte. Proveedor.

<sup>1</sup> PhD student in PPGEE, UFLA, Lavras, Minas Gerais, Brazil, E-mail: daiane.goncalves@estudante.ufla.br

<sup>2</sup> PhD student in PPGEE, UFLA, Lavras, Minas Gerais, Brazil, E-mail: mresende31@gmail.com

<sup>3</sup> Prof. Dr., Dept. of Statistics, UNIFAL, Alfenas, Minas Gerais, Brazil, E-mail: natalia.martins@unifal-mg.edu.br

<sup>4</sup> Prof. Dr., Dept. de Agricultural Engineering, UFLA, Lavras, Minas Gerais, Brazil, E-mail: flavioiborem@ufla.br

<sup>5</sup> Prof. Dr., Dept. of Statistics, UFLA, Lavras, Minas Gerais, Brazil; E-mail: macufla@ufla.br

## Introduction

Given the economic relevance of the coffee sector, forecasts of coffee trade and yield and its by-products must be highlighted. In this context are included the species *Coffea arabica* and *Coffea canephora*, which represent about 68% and 32% of the Brazilian production, respectively (CONAB, 2021).

If compared, arabica and conillon coffees differ considerably in price, beverage quality, and market acceptance. The former has a higher market value since it is most appreciated by consumers given its better quality. However, the higher soluble solid contents and yields of conillon coffee after roasting are notable features in soluble coffee manufacturing. Therefore, this coffee is often used in blends with arabica to produce roasted and ground coffee, soluble, espresso and other modern forms of coffee consumption, as its final product has high market competitiveness, gaining spaces and attracting the attention of the main coffee segments in the world (FERRÃO *et al.*, 2019).

In Brazil, the states of Minas Gerais and Espírito Santo occupy, respectively, the first and second place in terms of national production (COSTA, 2020). The southern region of Minas Gerais has a peculiar relevance due to its production of Arabica coffees certified as specialty coffee, mainly those in the Serra da Mantiqueira region (BORÉM *et al.*, 2019). Given its high sensory quality, this is one of the most important specialty coffee-producing regions in Brazil. Such quality has been consistently proven in prime Brazilian quality contests. Therefore, southern Minas is a region with ease and feasibility for production of specialty coffees (SANTOS; CIRILLO; GUIMARÃES, 2021).

Arabica and Conillon blends have still been criticized for their chemical and sensorial changes (BOTELHO *et al.*, 2016). Coffee evaluation is a complex task due to aroma, flavor, and chemical composition changes after roasting process (ALVARADO; LINNEMANN, 2010; BAQUETA; COQUEIRO; VALDERRAMA, 2019). Nevertheless, constant coffee price fluctuations have justified such blends for reducing costs, as Conillon has lower costs.

Other factors such as coffee processing method have been considered in the roasting industry to improve the quality of blends (RIBEIRO *et al.*, 2017). This is because volatile compounds that impart aroma and flavor to coffee beverages are influenced by that factor, affecting their quality (SUNARHARUM; WILLIAMS; SMYTH, 2014).

Since roasters can perform countless blends between Arabica and Conillon, at different concentrations, numerous new products can be then generated. These have been classified by programs of the Associação Brasileira da Indústria de Café (ABIC, 2021), wherein certifications are issued as part of the Coffee Quality Program (PQC). It has been focused on differentiating blends in terms of consistency of their manufacturing processes, aiming to increase coffee consumption. Thus, the productive sector and the market have established a transparent commercial relationship, based on free choice of consumers, product value addition, and price differentiation, based on the quality thereof.

Based on the above, adding value to the quality and production of blends will need improved and/or new more informative statistical methods. Therefore, this study aimed to classify the blends between Arabica specialty coffees with Conillon coffee and commercial coffees, using cluster analysis in qualitative and quantitative approaches. Their feasibility will be determined by comparison with technical criteria and standards proposed by the ABIC (2021), which are used to classify different beverage quality categories.

## Methodology

The methodology used in this study involved the following steps: blend preparations and experiment planning, cluster analysis of quantitative data, and cluster analysis of qualitative data.

### *Blend preparations and experiment planning*

Four experiments were carried out, in which blends were formed by mixing specialty Arabica (CE), Conillon (CC), and commercial (CT) coffees. Table 1 describes the proportions of each coffee blend, specifying concentrations in each experiment, and differentiating specialty coffee processing into natural (Experiments 1 and 2) and peeled cherry coffees (Experiments 3 and 4).

The experiment was divided into sessions and carried out on a continuous scale. Each sample was evaluated for its sensory properties (e.g., body, flavor, acidity, and bitterness). The final score comprised the sum of all properties, representing the overall quality of a blend (TOLEDO *et al.*, 2016). According to the Cupping Protocol of the SCA (2018), this sensory testing is done to determine the actual sensory differences between samples in terms of taste and establish their preferences by consumers.

**Table 1** – Description of blend compositions in proportions, with specification of concentrations and sample coding

Experiment 1			
Code	Concentration (35g/500 mL)		
Sample (Exp.)	CE	CT	CC
1(1)	1.000	0.000	0.000
2(1)	0.670	0.330	0.000
3(1)	0.340	0.330	0.330
4(1)	0.500	0.500	0.000
5(1)	0.500	0.000	0.500
6(1)	0.340	0.660	0.000
7(1)	0.340	0.000	0.660
8(1)	0.000	1.000	0.000
9(1)	0.000	0.000	1.000
Experiment 2			
Code	Concentration (50g/500 mL)		
Sample (Exp.)	CE	CT	CC
1(2)	0.340	0.330	0.330
2(2)	0.000	0.000	1.000
3(2)	0.340	0.000	0.660
4(2)	0.000	1.000	0.000
5(2)	0.670	0.330	0.000
6(2)	0.340	0.660	0.000
7(2)	1.000	0.000	0.000
8(2)	0.500	0.000	0.500
9(2)	0.500	0.500	0.000
Experiment 3			
Code	Concentration (35g/500 mL)		
Sample (Exp.)	CE	CT	CC
1(3)	0.500	0.000	0.500
2(3)	0.340	0.330	0.330
3(3)	1.000	0.000	0.000
4(3)	0.500	0.500	0.000
5(3)	0.340	0.000	0.660
6(3)	0.340	0.660	0.000
7(3)	0.000	0.000	1.000
8(3)	0.000	1.000	0.000
9(3)	0.670	0.330	0.000
Experiment 4			
Code	Concentration (50g/500 mL)		
Sample (Exp.)	CE	CT	CC
1(4)	0.500	0.000	0.500
2(4)	0.500	0.500	0.000
3(4)	0.340	0.000	0.660
4(4)	0.340	0.330	0.330
5(4)	0.670	0.330	0.000
6(4)	0.000	0.000	1.000
7(4)	1.000	0.000	0.000
8(4)	0.000	1.000	0.000
9(4)	0.340	0.660	0.000

Source: The authors.

The sensory properties evaluated simultaneously constituted a multivariate sample and were analyzed by cluster analysis, as described in the following sections.

*Cluster analysis of quantitative data*

Quantitative data were analyzed by hierarchical clustering represented by dendrograms, summarizing information from similar observation vectors to group them into clusters. To do so, the distance between each pair of observations was used, contextualizing the sensory properties by notations between two assessments, as follows:  $x_1 = (x_{11}, x_{12}, \dots, x_{1p})$  and  $x_2 = (x_{21}, x_{22}, \dots, x_{2p})$ , in which  $p$  represented the number of sensory properties evaluated ( $p = 5$ ). Each element corresponded to a score on the continuous scale, following the sequence body, flavor, acidity, bitterness, and final score. Thus, according to Johnson and Wichern (2014) the Euclidean distance is given by equation (1)

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^p (x_{1j} - x_{2j})^2}. \quad (1)$$

It is noteworthy that since distance increases as two units move apart, so its use in graphical analysis is known as dissimilarity measure and is generalized for all observations, with the dissimilarity matrix  $D$  written by equation (2)

$$D = \begin{bmatrix} 0 & 0 & \dots & 0 \\ d(x_1, x_2) & 0 & \dots & 0 \\ d(x_1, x_3) & d(x_2, x_3) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d(x_1, x_n) & d(x_2, x_n) & \dots & 0 \end{bmatrix}. \quad (2)$$

After obtaining this matrix, a cluster analysis was carried out considering different criteria, as in Table 2, given that  $x_1$  and  $x_2$  are multivariate observations initially belonging to two partitions illustrated in  $A$  and  $B$ , respectively, with the elements  $n_A$  and  $n_B$ .

Regarding the centroid criterion, the two clusters with the shortest distance between centroids are merged in each step. Therefore, assuming that clusters  $A$  and  $B$  are unified, the centroid of the new cluster  $AB$  was obtained by the weighted average, as in equation (3)

$$\bar{x}_{AB} = \frac{n_A x_1 + n_B x_2}{n_A + n_B}, \quad x_1 \in A \quad \text{and} \quad x_2 \in B. \quad (3)$$

**Table 2** – Description of the criteria used in cluster analysis

Criteria	Description
Closest neighbor	$D(x_1, x_2) = \min\{d(x_1, x_2)\}$ , to $x_1 \in A$ and $x_2 \in B$
Furthest neighbor	$D(x_1, x_2) = \max\{d(x_1, x_2)\}$ , to $x_1 \in A$ and $x_2 \in B$
Average linkage	$D(x_1, x_2) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(x_1, x_2)$ , to $x_1 \in A$ and $x_2 \in B$
Centroid	$D(x_1, x_2) = d(\bar{x}_1, \bar{x}_2)$ ; $\bar{x}_1 = \sum_{i=1}^{n_A} (x_{1i}/n_A)$ and $\bar{x}_2 = \sum_{j=1}^{n_B} (x_{2j}/n_B)$

Source: Based on Johnson and Wichern (2014).

If  $n_A > n_B$ , it implies that the centroid given by equation (3) may be closer to  $x_1$  than to  $x_2$ , thus using the median distance presented by equation (4) to calculate new distances regarding other clusters:

$$m_{AB} = \frac{1}{2}(x_1 + x_2). \tag{4}$$

The Ward method is defined as a combination of results from equations (3) and (4), as the two clusters with the shortest distance between centroids are merged in each step. Once a new cluster is formed by combining partitions of  $A$  and  $B$  clusters, and then the weighted average is determined.

The Ward method uses the squared distances within the cluster and the distances between clusters (FERREIRA, 2018), given by

$$SSE_A = \sum_{i=1}^{n_A} (x_{1i} - \bar{x}_A)'(x_{1i} - \bar{x}_A), \tag{5}$$

$$SSE_B = \sum_{i=1}^{n_B} (x_{2i} - \bar{x}_B)'(x_{2i} - \bar{x}_B). \tag{6}$$

If  $AB$  is the cluster obtained by combining clusters  $A$  and  $B$ , the sum of the distances within the cluster is given in equation (7)

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (x_{12i} - \bar{x}_{AB})'(x_{12i} - \bar{x}_{AB}), \tag{7}$$

$\bar{x}_{AB}$  is obtained by expression given in the equation (3) and  $n_{AB} = n_A + n_B$  are the number of observations in cluster  $AB$  obtained by combining  $A$  and  $B$ . Thus, the Ward method joins the two clusters  $A$  and  $B$ , which minimizes increases in the sum of squares within the cluster and is defined by equation (8)

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B). \tag{8}$$

The cophenetic correlation coefficient was used to select the best method, as different methods were used to build the dendrograms (CARVALHO; MUNITA; LAPOLLI, 2019; TOTTI; VENCOVSKY; BATISTA, 2004). It is defined as the Pearson correlation between the matrix elements  $D = [d_{ij}]$  and  $C = [c_{ij}]$ , in which  $C$  is the cophenetic matrix obtained by the Euclidean distances according to the criteria of the clustering, Table 2, and Ward methods.

*Cluster analysis of qualitative data*

Cluster analysis was applied to qualitative data, assuming that data were qualitative ordinal or nominal. Therefore, the criteria of the ABIC (2021) for classifying coffees into different categories of beverage quality were followed, Figure 1.

Given as a reference a minimum quality level, specified in 4.5 points, Figure 1, the responses to sensory properties (body, flavor, acidity, bitterness, and final score) were coded as specified in the standard (9)

$$\begin{aligned} y_{ij} \geq 4.5 &\rightarrow x_{ij} = 1, \\ y_{ij} < 4.5 &\rightarrow x_{ij} = 0, \end{aligned} \tag{9}$$

where  $y_{ij}$  is the continuous scale response of the  $i$ -th taster ( $i = 1, \dots, N$ ) for the  $j$ -th property ( $j = 1, \dots, 5$ ). The coded responses were assigned to  $x_{ij}$ , so that the vector  $X^i = (0, 1; 0, 1; 0, 1; 1, 0; 0, 1)$  was originated for each taster. The application of similarity distances can be understood when comparing the responses of two individuals (I1 and I2) organized in a contingency table, whose frequencies are computed in combinations of responses 1 and 0, see Table 3.

**Table 3** – Layout of the frequency table used to calculate the Jaccard metric

		I1		Totals
		1	0	
I2	1	$a = n_{11}$	$b = n_{10}$	$a + b$
	0	$c = n_{01}$	$d = n_{00}$	$c + d$
Totals		$a + c$	$b + d$	

Source: Based on Johnson and Wichern (2014).

Following this specification, we opted for the Jaccard metric, expressed by Ferreira (2018), as an alternative, which is defined as follows by the equation (10):

$$J = \frac{a}{a + b + c}. \tag{10}$$

**Figure 1** – ABIC technical recommendations



**Source:** Based on figure RECOMENDACOESTECNICASABIC.jpg in ABIC (2021).

Statistical analyses were performed using the vegan packages of the R software, using the vegdist function to obtain distances. The number of partitions to be considered was determined by the NbClust package, as proposed by Charrad *et al.* (2014).

**Results**

*Cluster analysis for continuous data without the minimum cutoff score established by ABIC*

Cluster analysis was performed following the proposed methodology and the criteria established in Table 2. From these, cophenetic correlation coefficients were obtained, Table 4, to determine the best method to justify similarities between the samples evaluated.

**Table 4** – Cophenetic correlation coefficients considering the grouping criteria

Method	Cophenetic Correlation Coefficient
single	0.635
complete	0.767
average	0.780
median	0.715
centroid	0.753
Ward D.	0.588

**Source:** The authors.

Table 4 shows the interpretation of results, with the higher the coefficient, the lesser the distortion between dissimilarity matrices obtained by the Euclidean distance and method used. In this perspective, the average linkage method provided the best fit; therefore, dendrogram construction should proceed to visualize groupings between samples.

The optimal number of clusters was estimated by indices defined by the 30 criteria in the NbClust package (CHARRAD *et al.*, 2014). The results with the highest

agreement among them, as shown in the plot of Figure 2, were considered. Thus, the dendrogram obtained by the average linkage method could be visualized, showing an optimal number of 4 clusters.

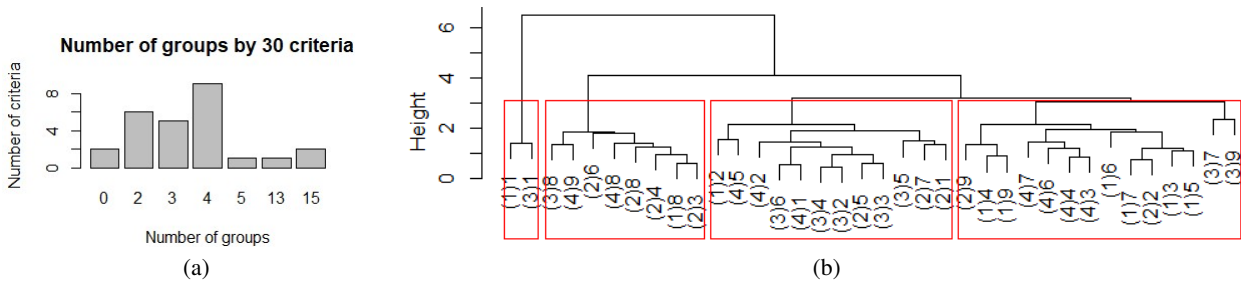
Figure 2 highlights that the first group comprised the samples coded in (1)1 and (3)1. These samples come from experiments with different processing types, but with the same concentration of coffees with a binary mixture, that is, composed of two types of coffee. Thus, there was a similarity between the binary blend and pure specialty Arabica coffee.

In the second grouping, 8 blend samples were similar coffees in terms of sensory properties. However, pure coffees had the most outstanding results, specifically for roasted coffee (CT). Of these, the samples (4)9 and (2)6 differed in their concentrations, despite having the same processing (peeled cherry). The same was observed for the samples of natural coffee (2)4 and (1)(8). Other similarities between blends were detected, but nothing specific to processing and concentration effects.

The third grouping showed greater similarities among samples from experiment 3, that is, blends processed via peeled cherry, and characterized by a concentration of 35g/500ml. Thus, the similarity between blends with binary blends and pure coffees became clear in this cluster.

Quantitative data cluster analysis allowed us to detect patterns to discriminate coffee processing types and blend concentrations based on scores of their main sensory properties. This finding is important for the quality standards defined by the Associação Brasileira da Indústria de Café (ABIC, 2021). Regarding the the breakdown of the categories into which these blends can be classified follows the cluster analysis for qualitative data, the results of which are detailed in the next section.

**Figure 2** – (a) Number of clusters for dendrogram construction; (b) dendrogram generated by the average linkage method.



Source: The authors.

*Cluster analysis of qualitative data without the minimum cutoff score of 4.5 established by the ABIC*

Following the criteria established by the ABIC, the value specified in 4.5 units was considered as a cutoff point, which discriminates whether to recommend a blend. In this context, for each property, in an exploratory way, there is information from the 36 samples, with 9 samples in each of the 4 experiments, and 83.33% of the samples received a value greater than or equal to 4.5 for the property body.

Values greater than or equal to 4.5 were given to 52.78% of the samples for flavor, 44.44% for acidity, 75% for bitterness, 58.33% for final score, that is, 41.67% of the samples did not receive a minimum score for coffee quality, thus not being recommended.

Once this cutoff score is considered to code the samples qualitatively in cluster analyses, the cophenetic correlation coefficient was obtained for each of the methods used, comparing the elements of the Jaccard’s dissimilarity matrix and of the cophenetic matrix, keeping the same grouping criteria, Table 2.

Table 5 shows that the best method is the centroid linkage with a correlation coefficient of 0.895. Notably, the correlation coefficient by the average linkage method had a value close to the centroid linkage method. Thus, the analysis should be carried out considering the average linkage method to compare with the method used in the continuous data with the Jaccard metrics. Figure 3 displays the dendrogram generated.

The qualitative approach analysis showed a better homogeneity between the coffee samples processed by combining the factors processing type and concentration, with emphasis on the third and fourth clustering. In these groups, the presence of experiments 4 and 3 (third cluster), and 2 (fourth cluster) are more pronounced, differing by processing type but with the same concentration (50g/500mL).

**Table 5** – Cophenetic correlation coefficients considering different methods for binary variables.

Method	Cophenetic Correlation Coefficient
single	0.841
complete	0.831
average	0.893
median	0.821
centroid	0.895
Ward D.	0.718

Source: The authors.

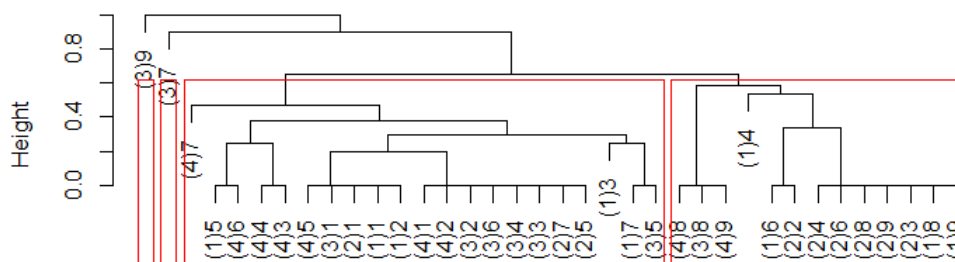
**Discussion**

Studies in the literature that used the same formulations and settings had results that corroborate the groups detected in our quantitative cluster analysis.

Similarities between blend type and score categories are subjectively interpreted since there is no threshold value to measure the magnitude of their distances (COSTA *et al.*, 2020).

Paulino *et al.* (2019) proposed a mixed model to evaluate experimental random effects with the same blend formulation, processing type, but different concentrations. These authors observed that a preference for specialty coffee blends reflects a more discriminated response to conilon and roasted coffee components.

Cirillo *et al.* (2019) proposed a new statistical procedure to analyze coffee blends by sensorial trials. The tool was able to identify outliers and evaluate participants in the sensory test panel when comparing the quality of blends with pure coffees. This was also efficient in discriminating blends containing pure coffees. In this study, concentration and processing type effects did not interfere in any way with statistical evaluations; therefore, the procedure has not been invalidated.

**Figure 3** – Dendrogram considering the average linkage method for binary data

Source: The authors.

Alessandrini *et al.* (2008) suggested that roasting causes chemical, physical, structural, and sensory changes in the coffee beans, which are derived from complex reactions that cause peculiar changes in coffee color, aroma, and flavor, having strong impacts on the quality of final products.

Moura *et al.* (2007) evaluated the physical, chemical, and sensory aspects of blends between Arabica and Robusta coffees. They found that caffeine contents was higher in Robusta than was in Arabica; therefore, when robusta is added to blends, caffeine contents are increased. Caffeine is odorless and has a characteristic bitter taste, contributing significantly to bitter notes in coffee beverages (MONTEIRO; TRUGO, 2005).

Lima Filho *et al.* (2015) commented that some of the blends used in industries may have different proportions between Arabica coffees and certain *Canephora* varieties, such as robusta, and that Arabica coffees are fruitier and more acidic beans, while Robusta ones are more bitter and ensure full-bodied beverages.

## Conclusion

Using cluster analysis considering cutoff score 4.5 as the minimum recommendation criterion (ABIC) was efficient due to corroboration with literature results and due to the addition of specialty coffees in almost all blends, revealing homogeneity in beverage quality. However, it should be used with caution and may vary with the metrics and grouping criteria used. On the other hand, its use for continuous data without the minimum cutoff score of 4.5 (ABIC) promoted more heterogeneous results in terms of coffee quality discrimination.

## Acknowledgments

The authors wish to thank the CAPES and CNPq (N° 140242/2019-8) - Brazil, for financial support.

## References

ABIC - ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. Programa de qualidade do café: norma da qualidade recomendável e boas práticas de fabricação de cafés torrados em grão e cafés torrados e moídos. Rio de Janeiro: ABIC, 2021. Available from: <<http://abic.com.br/>>. Access in: Sept. 01, 2021.

ALESSANDRINI, L.; ROMANI, S.; PINNAVAIA, G.; ROSA, M. D. Near infrared spectroscopy: an analytical tool predicted coffee roasting degree. *Analytica Chimica Acta*, [s. l.], v. 625, n. 1, p. 95-102, 2008. DOI: 10.1016/j.aca.2008.07.013.

ALVARADO, R. A.; LINNEMANN, A. R. The predictive value of a small consumer panel for coffee-cupping judgment. *British Food Journal*, Bingley, v. 112, n. 9, p. 1023–1032, 2010. DOI: <https://doi.org/10.1108/00070701011074372>.

BAQUETA, M. R.; COQUEIRO, A.; VALDERRAMA, P. Brazilian Coffee Blends: a simple and fast method by near-infrared spectroscopy for the determination of the sensory attributes elicited in professional coffee cupping. *Journal of Food Science*, Champaign, v. 84, n. 6, p. 1247-1255, 2019. DOI 10.1111/1750-3841.14617.

BORÉM, F. M.; CIRILLO, M. A.; ALVES, A. P. C.; SANTOS, C. M.; LISKA, G. R.; RAMOS, M. F. LIMA, R. R. Coffee sensory quality study based on spatial distribution in the mantiqueira mountain region of Brazil. *Journal of Sensory Studies*, Westport, v. 35, n. 2, p. e12552, 2019.

BOTELHO, F. M.; CORRÊA, P. C.; BOTELHO, S. C. C.; VARGAS-ELÍAS, G. A.; DINIZ, M. D. M. S.; OLIVEIRA, G. H. H. Propriedades físicas de frutos de café robusta durante a secagem: Determinação e modelagem. *Coffee Science*, Lavras, v. 11, p. 65-75, 2016.

- CARVALHO, P. R.; MUNITA, C. S.; LAPOLLI, A. L. Validity studies among hierarchical methods of cluster analysis using cophenetic correlation coefficient. *Brazilian Journal of Radiation Sciences*, Rio de Janeiro, v. 7, n. 2A, p. 1-14, 2019.
- CHARRAD, M.; GHAZZALI, N.; BOITEAU, V.; NIKNAFS, A. NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, [California], v. 61, n. 6, p. 1-36, 2014.
- CIRILLO, M. A.; RAMOS, M. F.; BORÉM, F. M.; MIRANDA, F. M.; RIBEIRO, D. E.; MENEZES, F.S. Statistical procedure for the composition of a sensory panel of blends of coffee with different qualities using the distribution of the extremes of the highest scores. *Acta Scientiarum. Agronomy*, Maringá, v. 41, p. e39332, 2019. DOI: 10.4025/actasciagron.v41i1.39323.
- CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. *Acompanhamento da safra brasileira: café: segundo levantamento*. Brasília: Conab, 2021. (Safra 2021, v. 8, n. 2).
- COSTA, A. S.; RESENDE, M.; NAKANO, E. Y.; CIRILLO, M. A.; BORÉM, F. M.; RIBEIRO, D. E. Proposal of a metric selection index for correspondence analysis: an application in the sensory evaluation of coffee blends. *Semina: Ciências Agrárias*, Londrina, v. 41, n. 2, p. 479-492, 2020. DOI: 10.5433/1679-0359.2020v41n2p479.
- COSTA, B. R. Brazilian specialty coffee scenario. In: ALMEIDA, Luciana Florêncio; SPERS, Eduardo Eugênio (ed.). *Coffee consumption and industry strategies in Brazil*. Cambridge: Woodhead Publishing, 2020. p. 51-64.
- FERRÃO, R. G.; FONSECA, A. F. A.; FERRÃO, M. A. G.; MUNER, L. H. *Conilon coffee: the coffea canephora produced in Brazil*. 3rd ed. Vitória: Incaper, 2019.
- FERREIRA, D. F. *Estatística multivariada*. Lavras: Ed. UFLA, 2018.
- LIMA FILHO, T.; LUCIA, S. M. D.; SARAIVA, S. H.; LIMA, R. M. Physical and chemical characteristics of espresso coffee beverages prepared from Arabica and Robusta coffee blends. *Revista Ceres*, Viçosa, v. 62, n. 4, p. 333-339, 2015. DOI: 10.1590/0034-737X201562040001.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. London, UK: Pearson, 2014.
- MONTEIRO, M. C.; TRUGO, L. C. Determinação de compostos bioativos em amostras comerciais de café torrado. *Química Nova*, São Paulo, v. 28, n. 4, p. 637-641, 2005.
- MOURA, S. C. S. R.; GERMER, S. P. M.; ANJOS, V. D. A.; Mori, E. E. M.; MATTOSO, L. H. C.; FIRMINO, A.; NASCIMENTO, C. J. F. Physical, chemical and sensorial evaluations of Arabica coffee blends with Canephora (Robusta) coffee blends. *Brazilian Journal of Food Technology*, Campinas, v. 10, n. 4, p. 271-277, 2007.
- PAULINO, A. L. B.; CIRILLO, M.A.; RIBEIRO, D. E.; BORÉM, F. M.; MATIAS, G. C. A mixed model applied to joint analysis in experiments with coffee blends using the least squares method. *Revista Ciência Agronômica*, v. 50, n. 3, p. 345-352, 2019.
- RIBEIRO, M. V. M.; BORALLE, N.; PEZZA, H. R.; PEZZA, L.; TOCI, A. T. Authenticity of roasted coffee using 1H NMR spectroscopy. *Journal of Food Composition and Analysis*, San Diego, v. 57, p. 24-30, 2017.
- SANTOS, P. M.; CIRILLO, M. Â.; GUIMARÃES, E. R. Specialty coffee in Brazil: transition among consumers' constructs using structural equation modeling. *British Food Journal*, Bingley, v. 123, n. 5, p. 1913-1930, 2021.
- SCA – SPECIALTY COFFEE ASSOCIATION. SCA protocols: Cupping specialty coffee. Califórnia: SCA, 2018. Available from: <<http://www.scaa.org/PDF/resources/cupping-protocols.pdf>>. Access in: Aug 25, 2021.
- SUNARHARUM, W. B.; WILLIAMS, D. J.; SMYTH, H. E. Complexity of coffee flavor: a compositional and sensory perspective. *Food Research International*, Ottawa, v. 62, p. 315-325, 2014.
- TOLEDO, P. R. A. B.; PEZZA, L.; PEZZA, H. P.; TOCI, A. T. Relationship between the different aspects related to coffee quality and their volatile compounds. *Comprehensive Reviews in Food Science and Food Safety*, Chicago, v. 15, n. 4, p. 705-719, 2016.
- TOTTI, R.; VENCOVSKY, R.; BATISTA, L. A. R. Utilização de métodos de agrupamentos hierárquicos em acessos de Paspalum (Graminea (Poaceae)). *Semina: Ciências Exatas e Tecnológicas*, Londrina, v. 22, n. 1, p. 25-35, 2001.

Received: Aug. 31, 2021  
Accepted: Oct. 16, 2021  
Published: Oct. 29, 2021