

Review of the contributions of some statistical methodologies related to the quality of specialty coffee

Marcelo Angelo Cirillo¹ , Flávio Meira Borém² , Mariana Resende¹ , Paulo César Ossani³ ,
Gilberto Rodrigues Liska⁴ , Fortunato Silva Menezes⁵ 

¹Universidade Federal de Lavras/UFLA, Departamento de Estatística, Lavras, MG, Brasil

²Universidade Federal de Lavras/UFLA, Departamento de Engenharia Agrícola, Lavras, MG, Brasil

³Universidade Estadual de Maringá/UEM, Departamento de Estatística, Maringá, PR, Brasil

⁴Universidade Federal de São Carlos/UFSCAR, Departamento de Tecnologia Agroindustrial e Socioeconomia Rural, Araras, SP, Brasil

⁵Universidade Federal de Lavras/UFLA, Departamento de Física, Lavras, MG, Brasil

Contact authors: macufla@ufla.br; flavioborem@ufla.br; mresende31@gmail.com; pccossani2@uem.br; gilbertoliska@ufscar.br; fmenezes@ufla.br

Received in July 19, 2024 and approved in August 26, 2024

ABSTRACT

Sensory perception involves psychological issues and cognitive skills. The greatest individual differences are observed in sensory experiments, since tasters' senses assign positive or negative attributes to a product, which are numerically represented by means of a hedonic scale. Thus, when considering statistical methodologies to evaluate the results, the question becomes a challenge to answer because its application requires presumptions that are not always met. In addition, answering a single objective limits the search for the state of the art, contextualized in the exploration of other results that arises from new applied methodologies. Based on this argument, this study aims to present a methodological review that presents the results and contributions of different statistical techniques, addressed by mixing distributions, resampling techniques to build indices robust to outliers, machine learning to improve multivariate classifiers and visualization techniques of the data. These are derived from a single database, which references analyses of the quality of specialty coffees with denomination of known origins that are evaluated by heterogeneous groups of tasters. Based on the knowledge base of the researchers and a review of the literature, the aforementioned methodological procedures are applied, and their advantages and aggregate results are compared to the understanding of the behavioral attitudes that lead to the distinction of the quality of the coffees evaluated.

Key words: Taster; attributes; data mining; models; machine learning.

1 INTRODUCTION

Based on a scientific approach, a sensory experiment is planned with the aim of generating reliable responses. In addition to the preparation of the samples and necessary installations that minimize external influences in relation to the acquisition of sensory responses, it is recommended that the approach considers several statistical methodologies that include variations in measurements and robustness for outliers.

With the technological demand, from the data sciences perspective, the feasibility of proposing or applying machine learning and data visualization techniques also arises, aiming at the improvement in classificatory and predictive models with the use of computational resources. The amount of information to be combined and processed is gigantic, especially if we consider the heterogeneity of the sensory perceptions of the individuals who form a sensory panel for analyzing the quality of a particular product.

In the sensory analysis of coffees, the challenge in assigning an evaluation that differentiates the quality of the beverage is highlighted by Pereira et al. (2017) and Bhumiratana, Adhikari and Chambers (2011), who state that the nuances of flavor and aroma during the roasting process can generate more than 800 aromatic compounds in the beverage.

Given this complexity, the taster is limited in expressing their opinion when using the hedonic scale, with a range up

to 9 points, given a neutral category, and a bipolar structure that represents four positive categories, which are labeled in affective levels of tastes. Negative categories are divided into affective levels of disgust (Peryam; Pilgrim, 1957; Moskowitz; Sidel, 1971). It is common practice to assign a continuous scale in a range of 0-9 points, which favors the use of a wide class of statistical methodologies that can be applied.

Lim, Wood and Green, (2009) and 2011 criticize the construction of this scale, suggesting that the results be considered in an ordered or interval metric, and the use of usual statistical methodologies, such as analysis of variance, regression models or multiple comparisons, may provide inconsistent results because they violate assumptions of normality, independence and homogeneity of variances.

Regarding the reduction of the categories of the hedonic scale, Schutz and Cardello (2001) mention that the effect of this reduction makes the scale more vulnerable since there is a lower incidence of individuals marking the extreme categories. Lim and Fujimaru (2010) and Villanueva and Da Silva (2009) complement these statements with studies that emphasize the limitation and ability of individuals to differentiate between highly appreciated and unappreciated stimuli.

Regardless of the use of the scale, the measurements are subjective in nature, and in the composition of a sensory panel, Pereira et al. (2017) note that the selection of a specialist is

related not only to experience but also to the degree of sensory sensitivity particular to each individual.

When these characteristics, which are portrayed sensitively through the experience, are not associated, there are repercussions regarding a lack of agreement between opinions, causing contradictory or discrepant results in the evaluations. To resolve this, it is pertinent to use a more refined statistical methodology for analyzing the data in the breakdown of products.

In experiments involving consumers, this fact is more challenging because Costel (2002) warns of a potential fallacy that the opinion of experts represents the preference of consumers. This statement is scientifically supported by theoretical studies that connect the attitudes and psychological behavior of consumers to their expectations. (Sorensen; Andersen; Andersen, 2014; Dai et al., 2015; Brookes, 2014). The main ones are highlighted below.

The assimilation theory reported by Lankton and McKnight (2012) offers the hypothesis that consumers are reluctant to accept discrepancies with respect to previously assumed positions. In the case of the sensory analysis of coffees, it is plausible to suppose that the evaluations given in the first repetition in relation to a particular beverage generate some feeling considered to be the initial stimulus.

Thus, in the next repetition, when tasting the same beverage, the consumer tends to maintain the same score, based on their sensory memory, which results from the initial stimulus acquired in the previous evaluation. By reporting opinions in numerical terms, the results will have little variation.

Makarem and Jae (2016) state that the emotions perceived in the consumer while sampling a coffee with a high degree of concentration, acidity or flavor are identified by a set of hedonic responses that originate in the initial experiment, such as facial expressions. In such a way that the stimuli of the other consumers are influenced in the sense of provoking reactions such as pleasure/displeasure, relaxation/action or calmness/excitement.

The attribution theory reported by Oliver (2015) is explained by inferences of causality in a kind of confrontation between internal and external causes. Aiming for a better understanding, we consider a group of consumers accustomed to coffees of a particular brand or region when tasting coffees of other brands or different regions, who tend to confront their expectations and attribute a negative evaluation to the new drink.

The aspects raised in the synthesis of these theories make the analysis of subjectivity a key issue in a sensory experiment, and its measurement should be addressed to correlate with the measurements obtained by the sensory panel.

With this motivation, Singham, Birval and Yadav (2015) recommended the use of different statistical tools and

mathematical modeling in conjunction with psychophysical models. In this context, recommendations given by Wagenmakers et al. (2021) can be performed through seven steps of statistical analyses with greater transparency. Among them, the authors mention data visualization procedures and other techniques aimed at quantifying uncertainties, such as the use of different inference approaches and multiple response analysis.

The importance of questioning statistical methodologies is that it can aid in achieving reliable results. Onofri et al. (2009) describe a systematic review conducted by Johnson and Berger (1982) in biological and agricultural journals, in which 70% of the articles use or report incorrectly applied statistics. In journals in the media area, Clayton (2007) presents a discussion of the misuse of statistics for not considering the rigorous description of the experimental design involved in the research.

In view of the above, reporting the aforementioned discussions regarding behavioral and psychological attitudes of consumption, variability and data treatment in the continuous hedonic scale, the following objective is stated.

The main objective of this article is to review the contributions of different statistical methodologies to approximate the search for the state of the art in the understanding of the behavioral and psychological attitudes of tasters in relation to the discrimination of the quality of specialty coffees produced in a mountainous region of the state of Minas Gerais in Brazil.

For this purpose, this review references several scientific publications over 10 years with different statistical approaches applied to a single database used as a research source in several projects with funding from CNPq - National Council for Scientific and Technological Development regarding the understanding of the skills of consumers in the evaluation of specialty coffees.

2 METHODOLOGY

2.1 Description of the experiment and sample preparation

Coffea arabica L. coffee samples were collected during three consecutive harvests (2010/11, 2011/12 and 2012/13) in commercial crops of properties in the municipality of Carmo de Minas, Minas Gerais, Brazil. The selected study area is delimited by the geographical coordinates 22°07'21" south latitude and 45°07'45" west longitude (Instituto Brasileiro de Geografia e Estatística - IBGE, 2009), with a territorial extension of 32,332 ha.

With these specifications, the coding assigned to each type of coffee and its main characteristics related to the differentiation of the samples are given in Table 1.

Table 1: Description of the specialty coffees evaluated in the sensory analysis.

Type	Genotype	Altitude	Processing
A	Bourbon Amarelo	above 1,200 m	Natural
B	Acaia	below 1,100 m	Peeled cherry
C	Acaia	below 1,100 m	Natural
D	Bourbon Amarelo	above 1,200 m	Peeled cherry

The selection of samples of 100% Arabica coffee was performed by removing the defective beans and roasting them within a period of 24 hours for tasting, following the SCA protocols described by Lingle (2011). The roasting point was determined visually using the color classification system using standardized discs (Agtron, 2010). Regarding the preparation of the beverage, the concentration of 7%w/v was maintained using filtered water ready for consumption, which was free of any contaminant and without added sugar.

2.2 Sensory analysis

The sensory panel was constructed considering two profiles of tasters in two groups. The first, named G_1 , was composed of 43 inexperienced individuals; however, all individuals underwent basic training in coffee tasting. The second group, named G_2 , was formed by 52 individuals who acted as technicians or researchers and were considered experienced in the sensory analysis of coffees.

One of the objectives of the sensory analysis performed in this experiment is to verify the compatibility of the sensory responses of the attributes flavor, acidity, and body. We then note the analysis of the coffees, described in Table 1, relative to the acceptance, discrimination and preference for specialty coffees among the two groups G_1 (beginners) and G_2 (experts), respectively.

A form was used as an evaluation instrument, as shown in the layout of Figure 1, so that each attribute was evaluated for each type of coffee on a continuous scale of 0-10 points. In the cup tests in both groups, each individual received a single sample for each type of coffee, for which the origin, variety and type of processing were omitted.

IDENTIFICATION

City /State: _____ University graduate: _____ Age: _____ Gender: M() W()

Using a scale, mark a score from zero to ten for the attributes of the coffees below.

A	Overall Score:	0 ----- 10	NA()
	Body:	0 ----- 10	NA()
	Acidity:	0 ----- 10	NA()
	Sweetness:	0 ----- 10	NA()
B	Overall Score:	0 ----- 10	NA()
	Body:	0 ----- 10	NA()
	Acidity:	0 ----- 10	NA()
	Sweetness:	0 ----- 10	NA()
C	Overall Score:	0 ----- 10	NA()
	Body:	0 ----- 10	NA()
	Acidity:	0 ----- 10	NA()
	Sweetness:	0 ----- 10	NA()
D	Overall Score:	0 ----- 10	NA()
	Body:	0 ----- 10	NA()
	Acidity:	0 ----- 10	NA()
	Sweetness:	0 ----- 10	NA()

Figure 1: Layout of the sensory sheet used in the sensory evaluations of the specialty coffees described in Table 1.

To seek the state of the art in understanding the differences and similarities between the two groups of consumers, different statistical methodologies were implemented for the same data set.

The contribution of each is verified in the confirmation of the results, as well as the aggregation of new conclusions particular to each technique. The purpose of this technique is to gain a better understanding of the sensorial perceptions used, since there is intrinsic subjectivity in every individual participating in the process.

2.3 Expected results

The effect of altitude is associated with coffee quality (Worku et al., 2018; Bertrand et al., 2012). In Colombia, at higher altitudes between 1,450 and 1,650 m, the sensory quality of coffee was higher compared to coffees produced at lower altitudes, showing that altitude has a determining effect on coffee characteristics (Buenaventura-Serrano; Castaño-Castrillón, 2002). In Honduras, coffees of superior quality were found in places with high altitudes and annual rainfall below 1,500 mm (Decazy et al., 2003).

In considering the effect of altitude, the common characteristics of coffees A and D (Table 1) are limited by the fact that, regardless of the type of processing, samples of Bourbon Amarelo coffee produced above 1,200 m altitude are characterized by a high sensory quality. In general, this variety produces coffees with citrus acidity, high sweetness, creamy bodies and floral and fruity notes in higher environments (Borém et al. 2020, Ramos et al., 2016).

In the case of coffees B and C (Table 1), the most relevant characteristics are explained by the fact that these varieties come from red fruits, and when in regions with altitudes below 1,000m, they result in beverages with less complex acidity, sweetness and flavor, with a predominance of caramel flavors and medium to low body (Borém et al. 2020).

With a certain frequency, these varieties can also produce coffees with slight astringency. With these characteristics, the sensory profile of the beverage is closer to the standard known by most Brazilian consumers, except for cases where defective coffees are consumed.

3 CONTRIBUTIONS OF STATISTICAL METHODOLOGIES APPLIED TO THE SENSORY ANALYSIS OF THE QUALITY OF SPECIALTY COFFEE

3.1 Mixing of distributions

Given the continuous scale in the sensory evaluation form (Figure 1), it became possible to determine the parametric values for the proportions of individuals belonging to groups

G_1 and G_2 while maintaining the specification of the two groups of consumers; thus, the exact proportions were $\phi_1 = 0.43$ and $\phi_2 = 0.57$, respectively.

When conjecturing in everyday situations, it is impossible to distinguish the ability of sensory perception in a sample of consumers to classify them as experienced or inexperienced, as the exact proportions of sensory perception cannot be estimated. Thus, they are considered latent variables. One way to predict them is to use the EM algorithm – hope and maximization (Dempster; Lair; Rubin, 1977; Li et al., 2019).

Based on the above, the plausibility of using this algorithm in the sensory experiment to discriminate the quality of specialty coffees is verified. Thus, Oliveira et al. (2023), on the assumption that the grades assigned on a continuous scale are influenced by the other sensory attributes, body, acidity and sweetness, justified consideration of the bivariate normal distribution. For a better understanding of the parametric values, referring to the mean and covariance structure that formalize this distribution, the following notation is used.

Given the sensory attributes, body, acidity, sweetness and final note, x_h was assumed ($h=1,...,4$). The identification of the type of coffee was given following the description in Table 1, which was given by ($k=1,...,4$). Based on the above, the vector of bivariate responses $\mathbf{x}_k = (x_h, x_l)$ ($h=2,...,4$) was considered for each consumer group G_j ($j = 1$ and 2), referring to beginners (G_1) and experts (G_2), respectively. Thus, the parameters were defined by means of the bivariate correlations Σ_{jk} with the respective mean vectors μ_{jk} .

Finally, the parametric vector to be estimated by the EM algorithm was represented by $\theta = (\mu_{1k}, \mu_{2k}, \Sigma_{1k}, \Sigma_{2k}, \phi_{1k}, \phi_{2k})$, with ϕ_j ($j=1,2$) being the latent variable to be evaluated as the main parameter of interest to be estimated and compared with the exact proportion. The results are described in Table 2.

Table 2: Proportions of consumer groups G_1 and G_2 estimated by the EM algorithm compared with the exact proportions $\phi_1 = 0.43$, $\phi_2 = 0.57$ considering the covariances between the score and the other attributes.

Coffee	Score e Body		Score e Acidity		Score e Sweetness	
	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$
A	0.437	0.563	0.387	0.613	0.471	0.529
B	0.429	0.571	0.369	0.630	0.496	0.504
C	0.398	0.602	0.414	0.586	0.422	0.577
D	0.407	0.593	0.446	0.554	0.351	0.649

The estimates of the proportions, described in Table 2, approximated the exact proportions, with small Monte Carlo oscillations and, of course, the correlation structure between each variable related to the grade, as described in Table 3.

Table 3: Estimates of the correlation matrix considering responses of sensory attributes correlated with grade for each type of coffee.

Group (G ₁) beginners individuals			
Coffee	Score e Body	Score e Acidity	Score e Sweetness
A	$\begin{bmatrix} 1 & 0.603 \\ 0.603 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.500 \\ 0.500 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.651 \\ 0.651 & 1 \end{bmatrix}$
B	$\begin{bmatrix} 1 & 0.625 \\ 0.625 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.596 \\ 0.596 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.775 \\ 0.775 & 1 \end{bmatrix}$
C	$\begin{bmatrix} 1 & 0.520 \\ 0.520 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.284 \\ 0.284 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.600 \\ 0.600 & 1 \end{bmatrix}$
D	$\begin{bmatrix} 1 & 0.729 \\ 0.729 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.331 \\ 0.331 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.680 \\ 0.680 & 1 \end{bmatrix}$
Group (G ₂) experts individuals			
Coffee	Score e Body	Score e Acidity	Score e Sweetness
A	$\begin{bmatrix} 1 & 0.381 \\ 0.381 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.156 \\ 0.156 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.545 \\ 0.545 & 1 \end{bmatrix}$
B	$\begin{bmatrix} 1 & 0.504 \\ 0.504 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.267 \\ 0.267 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.662 \\ 0.662 & 1 \end{bmatrix}$
C	$\begin{bmatrix} 1 & 0.132 \\ 0.132 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.455 \\ 0.455 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.545 \\ 0.545 & 1 \end{bmatrix}$
D	$\begin{bmatrix} 1 & 0.476 \\ 0.476 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.205 \\ 0.205 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.542 \\ 0.542 & 1 \end{bmatrix}$

Based on the estimates (Table 3), the correlations align with the expected results, as the highest correlations were influenced by the sweetness attribute. For the G₂ group, which was characterized by more experienced individuals, the type C coffee had characteristics closer to those of conventional coffees, which resulted in lower correlations to the other attributes. The evaluations intrinsic to this group differentiated the coffees with higher quality relative to the expected results, with cases A and D reaching expectation.

Finally, regarding the discrimination of the groups, based on the bimodal behavior of the densities (Figures 2, 3 and 4), the assumption of bivariate normality imposed on sensory evaluations discriminates the presence of two distinct groups, that were previously considered in the experiment.

Given this application, the contribution of the EM algorithm in sensory analysis is marked as a relevant point.

3.2 Construction of statistical indices

Numerous factors are related to individuality in sensory perception. This makes it challenging to adjust a model that

explains the behavior of numerous consumers involved in the experiment, including their origin and specific characteristics as the samples were prepared. For example, genotype and type of processing may require more complex models that treat the data in a categorized way.

The desirable thing would be to synthesize the information into a single index that allows the researcher to discriminate the quality of the coffees and allows for interpretation of the degree of consumer preference. With this as motivation, Resende, Cirillo and Borém (2020) proposed an index based on the combination of several statistical techniques, which are described below sequentially in five steps to formalize the index.

The first stage consisted of processing the data to identify occurrences of outlier observations. To do so, using a multivariate approach, the construction of this index considers a rescheduling procedure proposed by Filzmoser, Maronna and Werner (2008), which is applied to a data matrix and contextualized in the sensory analysis, so that the lines correspond to the responses of individuals and the columns correspond to the sensory attributes.

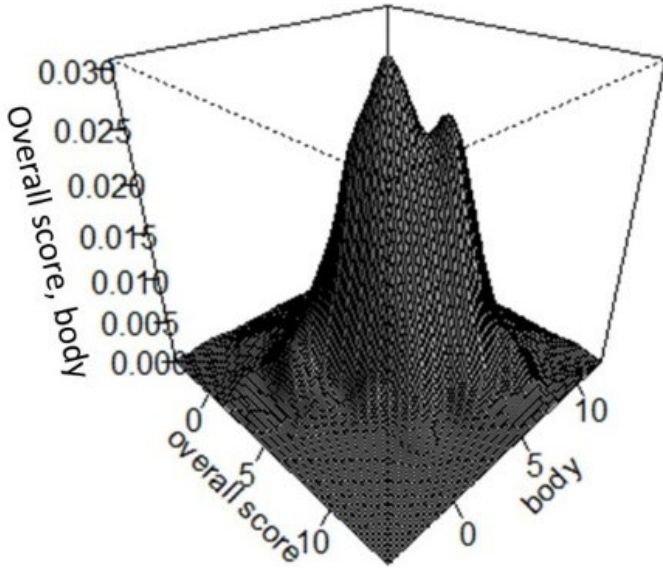


Figure 2: Density of the mixture of normals considering the covariance given by the final grade and body.

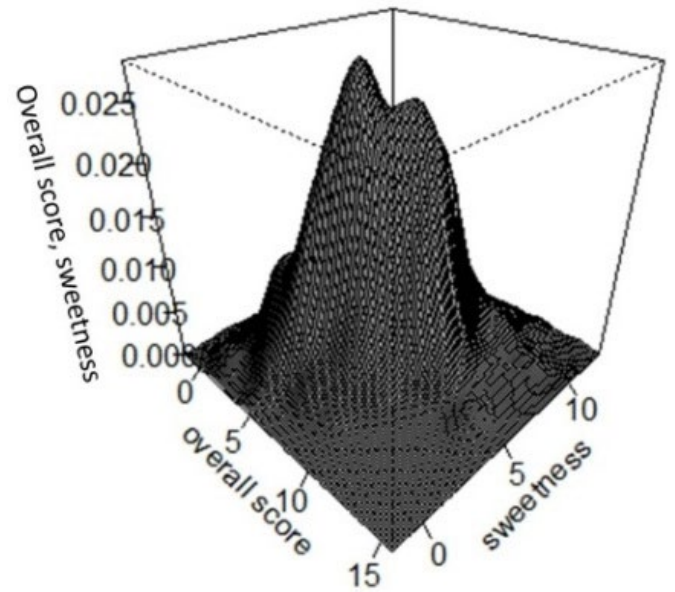


Figure 4: Density of the mixture of normals considering the covariance given by the final grade and sweetness.

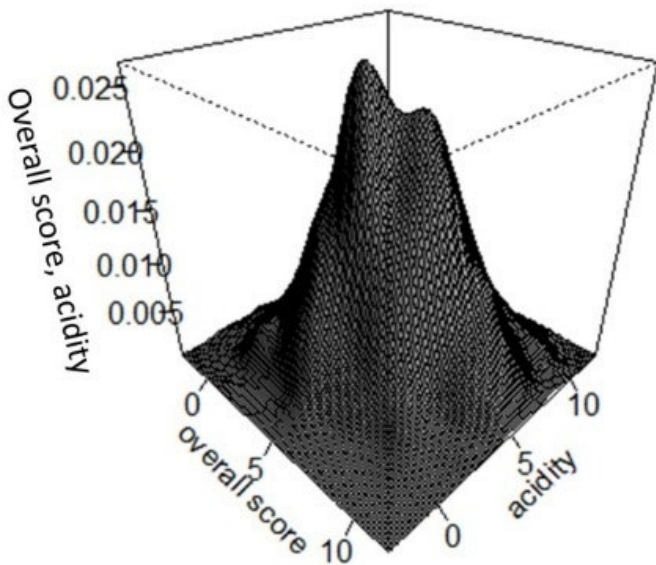


Figure 3: Density of the mixture of normals considering the covariance given by the final grade and acidity.

The second stage involves adjusting linear and quadratic regression models, considering a sensory attribute as the dependent variable; in this case, the most appropriate would be the final grade as a function of the scores obtained by the principal components, which better discriminate the outliers that are still present following the rescheduling procedure performed in the previous step. In this case, the components that best discriminate the outliers are identified using the kurtosis coefficient proposed by Peña and Prieto (2001).

The fit of the models, as well as an understanding of the factors involved in the experiment, is better understood by the following notation.

Considering a sample of N consumers and specification of the dependent variable, Y_{kic} represents the final grades assigned by the i -th taster ($i=1, \dots, N$), classified in the k -th group ($k=1,2$) and c -th type of coffee ($c=1,2,3$ and 4), in the sequence described in Table 1. The other attributes are assumed to be a data set, in which X_{kric} and X_{ksic} represent the components ($r \neq s$) that best discriminate the outliers and ζ_i indicates the random error associated with each consumer in the models fitted according to equations (1) and (2).

$$Y_{kic} = \beta_0 + \beta_r X_{kric} + \beta_s X_{ksic} + \zeta_{ic} \quad (i=1, \dots, N) \quad e \quad (k=1,2) \quad (1)$$

$$Y_{kic} = \beta_0 + \beta_r X_{kric} + \beta_s X_{ksic} + \beta_{rr} X_{kric}^2 + \beta_{ss} X_{ksic}^2 + \zeta_{ic} \quad (i=1, \dots, N) \quad e \quad (k=1,2) \quad (2)$$

The third stage consisted of assigning a binary variable, defined by $Z_{kic}(X_{kric}, X_{ksic})$, having as a criterion the expected value, in this case, predicted for the final grade $E(Y_{kic})$, after adjusting the models for each type of coffee, according to equation (3).

$$\begin{aligned} Z_{kic}(X_{kric}, X_{ksic}) &= 1 \text{ se } \hat{Y}_{kic} \geq E(Y_{kic}) ; \\ Z_{kic}(X_{kric}, X_{ksic}) &= 0 \text{ se } \hat{Y}_{kic} < E(Y_{kic}). \end{aligned} \quad (3)$$

Through expression (3), it is clear that the assignment of binary values depends on the selection of the model, whether linear (1) or quadratic (2).

Thus, the fourth stage is represented by the validation of the model, using the like-of-fit test, using the F test to judge the suitability of a higher-order model, in this case, quadratic compared to linear.

To avoid the fact that the residuals of the fitted models do not meet the assumptions of the F test, the use of bootstrap

tests is recommended to compute the p values; however, the weight of each observation to be resampled should preferably account for information of the diagnostic measures of the models, found in the literature of Dffits, Covratio and Cook's distance. More details are found in Andrade, Cirillo and Beijo (2014) and Davison and Hinkley (1997).

Finally, in the fifth stage, the index, called the preference index, is stated for a particular coffee genotype and each consumer group, following equation (4).

$$I_{\text{pref}}(X_{\text{ric}}, X_{\text{sic}}) = 100 \times \left[\frac{\sum_{i=1}^{N_{\text{kc}}} Z_{\text{kic}}(X_{\text{kric}}, X_{\text{ksic}})}{N_{\text{kc}}} \right] \text{ para } k = 1, 2; c = 1, 2, 3, 4, \quad (4)$$

where N_{kc} is the total number of consumers of the k th group who tasted the c -th coffee.

With the completion of this methodological procedure, the results obtained by Resende, Cirillo and Borém (2020) are illustrated in Figure 5, computing the p values of the like-of-fit statistic, obtained by the bootstrap test with weights given by the influence measures Dffits (fits); covratio (cov) and cook distance (cook) and compared with the significance levels, 0.05; 0.1 and 0.2 for each consumer group.

Considering the measure of influence given by the cook distance, the p values for both groups were greater than 0.20 for the results shown in Figure (5). Therefore, the

quadratic model is used in the estimation of the preference indices I_{pref} (4).

With the estimates of preference indices (Table 4), there is a confirmation of the expected results, with emphasis on the agreement between the two groups in the preference of type D coffee, grown at an altitude equal to 1,200 m. In the case of coffee type A, which differed only in how it was processed, when compared to coffee types B and C, the results were compatible with coffee D, as identified in the G_2 group formed by more experienced individuals.

Given the innovation in the formalization of indices that summarize information on the discrimination of coffee quality, we emphasize the relevant point that marks the contribution of the proposed methodology to be implemented in future sensory experiments, with emphasis on advantages compared to other methodologies.

3.3 Improvement of classificatory models via machine learning

A classificatory model in which a new evaluation is introduced involves the correct rates and errors in classifying an observation from a group of tasters. Liska et al. (2015) proposed the use of the boosting algorithm (Schapire, 1990; Freund; Shapire, 2012) as a way to improve classifiers of the Fisher discriminant model.

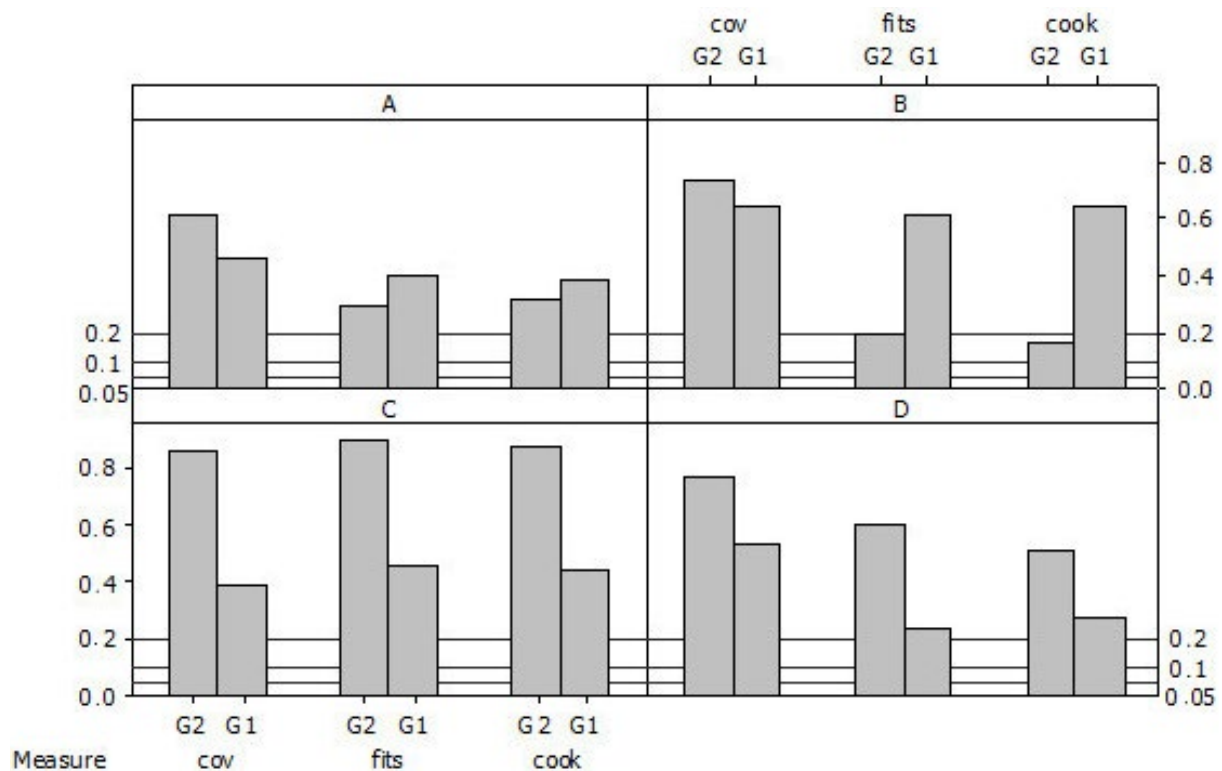


Figure 5: Significance probabilities obtained from the F statistical considering the covRatio (cov), Dffits (fits) and Cook distance (Cook) influence measures applied to the samples of specialty coffees tested by the consumer group G_1 and G_2 .

Table 4: Preference index (Ipref) computed with the selection of the quadratic model given the final grade as a function of the scores of the principal components that best discriminate the outliers.

Group	Coffee types			
	A	B	C	D
G ₁	60.15	65.78	67.92	73.88
G ₂	65.67	50.62	41.32	69.81

This methodology requires a new organization of the data set, adapted to the use of the mentioned techniques. That said, for better understanding, follow the layout of the data and the notation assigned to each variable.

The notation used to organize the data consists of indexing an index r for each attribute, considering the coding ($r=1, \dots, 4$) indexed sequentially, following the order of the attributes described in Figure 1 and combined with the coffee genotypes ($k=1, \dots, 4$), sequentially described in Table 1. Thus, a total of 16 variables were obtained, named X_j ($j=1, \dots, r_k=16$) with the response of the panelists, denoted by irk , where $i=1, 2, \dots, n_1$, the total number of panelists of the G group 1 e; $n_1 + 1, \dots, n_2$, with n_2 equal to the number of panelists in group G₂. With these specifications, the layout of the data panel, described in Table 5, follows.

Table 5: Layout of the data panel used to compute the classification rates of the discriminant analysis procedure and incorporation of the boosting method.

Group	Panelist	X_1	X_2	X_3	X_4	...	X_{16}
Beginner individuals (G ₁)	1	a_{11}	a_{12}	a_{13}	a_{14}	...	a_{113}
	2	a_{21}	a_{22}	a_{23}	a_{24}	...	a_{213}
	.	.	.	.	.	.	.
	.	.	.	.	.	...	.
	.	.	.	.	.	.	.
	n_1	a_{n11}	a_{n12}	a_{n13}	a_{n14}	...	a_{n113}
Experts individuals (G ₂)	1	$a_{(n1+1)1}$	$a_{(n1+1)2}$	$a_{(n1+1)3}$	$a_{(n1+1)4}$	...	$a_{(n1+1)13}$
	.	.	.	.	.	.	.
	.	.	.	.	.	...	.
	.	.	.	.	.	.	.
	n_2	a_{n21}	a_{n22}	a_{n23}	a_{n24}	...	a_{n213}

In addition to the possible estimates regarding the interpretations, the specificity and sensitivity rate, it becomes more interesting to evaluate the occurrence of false negatives, i.e., the panelist belongs to the beginners group (G₁), but the model classifies him as belonging to the group of experts (G₂). Otherwise, a false positive may occur,

To evaluate the performance of the classifiers obtained by the two methods, the data were separated into two parts, resulting in the training set that consisted of 70% of the original sample and the complementary set, with 30% constituting the test set used to calculate the probabilities.

The false positive and negative rates, when using the classifier resulting from Fisher's discriminant analysis, were estimated at 41.18% and 46.15%, respectively. As a confirmatory analysis, cross-validation was applied; however, the results (Table 6) did not show significant improvements. Therefore, as motivation, there is an appeal to incorporate machine learning techniques as a way to improve classification rates.

The results relative to the percentage of observations classified in both groups are described in Table 6, with the most prominent results being the reduction of false negative rates, i.e., the proposed classifier improved by the boosting technique has a greater ability to correctly classify the observations from the group (G₁).

For the formation of an expert panel, this result serves as a basis for selecting nonexperienced individuals. After training, the reliability of the evaluations can be judged by this classifier with high accuracy rates.

Notably, the application of the boosting procedure, reported in the previous results (Table 6), only considered the discrimination between two groups.

For the same experimental situation, Oliveira et al. (2019) proposed an extension that considered a multiclass approach, maintaining the same theoretical foundation and number of variables but considering the formation of three groups defined by T₁ trained panelists aged between 19 and 50 years; T₂ untrained panelists aged between 19 and 50 years; and T₃ untrained panelists aged over 50 years.

Another modification performed in this approach was the inclusion of the classifier generated by the analysis of the quadratic discriminant model, and as a confirmatory study, the bagging method (Breiman, 1996), known as aggregate bootstrap, was used, which consists of generating bootstrap samples of the data as new training sets, yielding new predictors.

The results of the adaptations of the overall error rates are described in Table 7.

According to the results reported in Table 7, the machine learning procedures together with the bootstrap techniques were efficient in reducing the overall error rates compared to the conventional classifiers.

In summary, the use of machine learning as a mechanism for improving classifiers provided the gain of information for the understanding of perceptions, with advantages highlighted below.

Another point is relevant when considering age groups as criteria for age formation. There is statistical evidence that the machine learning approach will provide lower classification error rates.

Table 6: Count of false positives and negatives referring to the classification of the panelists into trained and untrained groups using the classifier generated by Fisher's discriminant analysis, cross-validation procedures and the boosting method.

Classification Observation	Classification by model by discriminant analysis	
	Group 1 (Trained) (%)	Group 2 (Untrained) (%)
Group G ₁ (Beginners)	53.85	46.15
Group G ₂ (Experts)	41.17	58.82
Classification Observation	Classification by the discriminant model after the cross-validation procedure	
	Group 1 (Trained)	Group 2 (Untrained)
Group G ₁ (Beginners)	64.25	34.28
Group G ₂ (Experts)	42.21	57.79
Classification Observation	Classification by the model using the classifier generated by the boosting method	
	Group 1 (Trained)	Group 2 (Untrained)
Group G ₁ (Beginners)	85.71	14.29
Group G ₂ (Experts)	68.75	31.25
Classification Observation	Classification by the model using the classifier generated by the boosting method after the cross-validation procedure	
	Group 1 (Trained)	Group 2 (Untrained)
Group G ₁ (Beginners)	80.63	19.37
Group G ₂ (Experts)	71.45	28.55

Table 7: Overall error rates obtained by the classification methods: linear and quadratic discriminant analysis and adaptation of the boosting and bagging methods.

Classification Methods	Error rates (%)
Boosting	19.56
Bagging	28.26
Quadratic Discriminant Analysis	36.95
Linear Discriminant Analysis	45.65

3.4 Pursuit projection in the analysis of unsupervised classifiers

Pursuit projection consists of an exploratory data visualization technique in reduced dimensions generated by linear projections obtained according to objective function maximization, specified as projection indices (Friedman; Tukey, 1974; Ossani; Ramos; Cirillo, 2020).

Naturally, its applicability is based on a data set with a greater number of variables; however, as this number is increased, the more equidistant the observations become. This fact is known in the literature as the "curse of dimensionality" (Aggarwal, 2015).

The contextualization and use of this projection for the sensory experiment evaluated in this study is motivated by the number of variables generated by combining the sensory attributes (Figure 1) with information from the coffee genotypes (Table 1), groups of tasters (G₁ and G₂), sex (male and female) and age range of the individuals. For such a

situation, the use of classification models may still present imprecise classification rates.

With this argument, Ossani et al. (2020) specified the Holes projection index (Cook; Buja; Cabrera, 1993; Cook; Swayne, 2007), reducing the data set by three dimensions. Subsequently, they adapted several unsupervised classifiers in relation to the discrimination of specialty coffees. The results are illustrated in Figure 6.

Considering the data set reduced in three dimensions, Figure 6 illustrates that it was possible to detect the discrimination of the special coffees, which was expected, based on the very characterization of each type of coffee (Table 1). This demonstrates that the original evaluations were not considered, but linear projections were.

The classification error rates and the main information of the classifiers used are described in Table 8.

Based on the results described in Table 8, most classifiers presented percentages of classification errors lower than 30%, which are considered reasonable in some cases, with estimates lower than 2%, due to the distance differences used in combination with other clustering procedures.

Following the statistical approaches mentioned in this review, Table 9 provides a summary of the main advantages that can be analyzed in future applications involving continuous, categorical, and classificatory variables, contextualized in sensory analysis and coffee quality.

The use of computational resources is appropriate in situations where numerous experiments would need to

be conducted to better validate the results. In this context, Monte Carlo simulation procedures, considering different statistical distributions combined with the importance sampling method, were evaluated by Ferreira et al. (2021) for the same data assessed in previous procedures. Thus, the main contribution was identified in recognizing the coffee with the greatest variability in scores across the simulation of 1,000 experiments. Another application of Monte Carlo simulation was conducted by Calcagnoto and Cirillo (2024), who considered scenarios with different levels of spatial dependence in the analysis of coffee

quality, following the ABIC criteria. It is worth noting that numerous data analysis techniques can be applied and compared with the techniques presented in this review. When it comes to machine learning techniques, assuming that the classification rule is not linear, such as Fisher's discriminant analysis, the Support Vector Machine (SVM) technique can be used. With this approach, Velásquez et al. (2021), compared to the Random Forest (RF) technique, achieved better accuracy and hit rates in classifying the maturity stages of coffees processed via pulped natural, considering physicochemical and spectral variables.

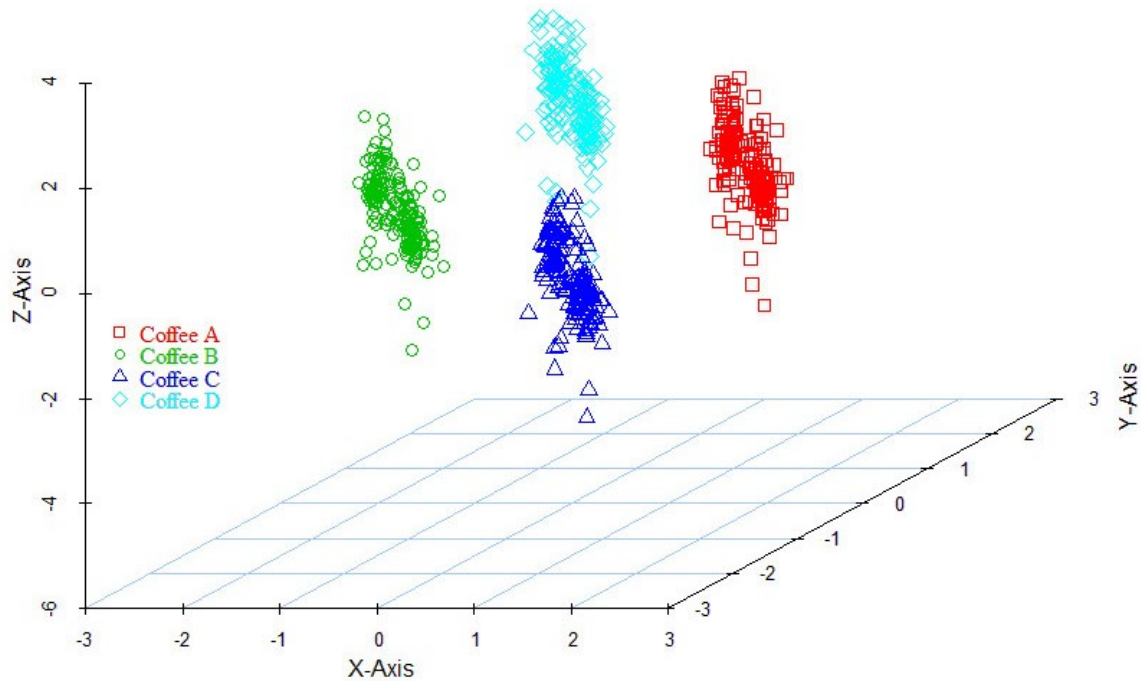


Figure 6: Graph of the results of the pursuit projection in data using the Holes index in three dimensions.

Table 8: Results of the classification techniques, with better results in the unsupervised classification in the data with reduced dimensions.

Nº	Classifier	Distance	Linkage	Validation error rate %
1	Farthest First	-	-	50.43
2	K-means	Euclidean	-	36.49
3	Hierarchical Clustering	Euclidean	Single	24.86*
4	Hierarchical Clustering	Euclidean	Average	25.57*
5	Hierarchical Clustering	Maximum	Ward	1.44*
6	Hierarchical Clustering	Maximum	Average	25.57*
7	Hierarchical Clustering	Maximum	McQuitty	32.04
8	Hierarchical Clustering	Manhattan	Single	24.86*
9	Hierarchical Clustering	Manhattan	Average	25.29*
10	Hierarchical Clustering	Canberra	Complete	24.28*
11	Hierarchical Clustering	Minkowski ³	Ward	1.44*
12	Hierarchical Clustering	Minkowski ³	Average	25.57*

* Results with a rate of less than 30%.

Table 9: Summary of the main contributions of the reviewed methodologies in the analysis of sensory data for the discrimination of specialty coffee quality.

Methodology	Advantages	Fundamentals	References
EM algorithm	The method allows for classifying individuals in relation to sensory perceptions, which was previously considered a primary classification; thus, it has indications of providing confirmatory results.	Mixture the distributions with continuous and correlated responses, given the assumption of multivariate normality.	Oliveira et al. (2023)
Construction of statistical indices	The combination of statistical methodologies that are robust to outliers arising in numerical evaluations provides reliable and discriminatory results regarding the quality of specialty coffees, related to their processing particularities, altitudes and genotypes, without imposing assumptions on statistical models that are more complex.	Response surface models, outlier treatment.	Resende, Cirillo and Borém (2020)
Improvement of classificatory models via machine learning	The machine learning procedure in classificatory models generated lower probabilities to favor reliability in sensory evaluations when a group of individuals with little experience, who may be consumers when receiving training, may act as panelists in the absence or difficulty of finding expert tasters	Refinement of hit and error rates generated in discriminant analysis.	Liska et al. (2015) Oliveira, et al. (2019)
Pursuit projection in the analysis of unsupervised classifiers	The linear projections obtained by the pursuit projection preserved the discrimination of specialty coffees in reduced dimensions.	Multidimensional data visualization.	Ossani et al. (2020)

Regarding the modeling of the covariance structure, considering the coffees produced in the Serra da Mantiqueira region, which is similar to the coffees evaluated in this work, Resende, Borém and Cirillo (2023) proposed a covariance structure by fitting structural equation models that include some environmental variables. Thus, given the structure adjusted by the structural model, when applying principal component analysis, the performance, in terms of the percentage of explained sample variability, was better than that of the sample covariance.

4 CONCLUSIONS

Given the different statistical approaches applied, the construction of knowledge regarding the sensory aptitudes of groups with different skills was noticeable for each technique applied. However, each technique requires arrangement of the data, in the panel format, as presented in the layouts and a rigorous scientific notation, so that they are understandable for their interpretation and implementation in computational algorithms.

After highlighting the advantages of each approach (Tabela 9), compared to the expected results, the aggregated information is validated against the previous expectation, as described in the expected results with supporting scientific references.

5 ACKNOWLEDGEMENTS

The authors would like to thank the Research Supporting Foundation of Minas Gerais State (Fundação

de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG), for the financial support through research in the project n° BPD 00448/22 and National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico–CNPq) for the financial support through research in project 304939/2021-8.

6 AUTHORS' CONTRIBUTION

Conceptual Idea: Cirillo, M.A., Resende, M., Ossani, P.C., Liska, G.R., ; **Methodology design:** Cirillo, M.A., Resende, M., Ossani, P.C., Liska, G.R., Menezes, F.S.; **Data collection:** Borém, F.M., **Data analysis and interpretation:** Cirillo, M.A., Borém, F.M., Resende, M., Ossani, P.C., Liska, G.R., Menezes, F.S. **and Writing and editing:** Cirillo, M.A., Resende, M., Menezes, F.S.

7 REFERENCES

- AGGARWAL, C. C. **Data mining:** The textbook. New York: Springer, 2015. 763p.
- AGTRON. Specialty Coffee Association. Roast color classification system. **Long Beach: SCAA**, 2010. 9p.
- ANDRADE, L. R.; CIRILLO, M. A.; BEIJO, L. A. Proposal of a bootstrap procedure using measures of influence in non-linear regression models with outliers. **Acta Scientiarum Technology**, 36(1):93-99, 2014.

- BERTRAND, B. et al. Climatic factors directly impact the volatile organic compound Finger print in green *Arabica* coffee bean as well as coffee beverage quality. **Food Chemistry**, 135(4):2575-2583, 2012.
- BHUMIRATANA, N.; ADHIKARI, K.; CHAMBERS, E. Evolution of sensory aroma attributes from coffee beans to brewed coffee. **LWT Food Science and Technology**, 44(10):2185-2192, 2011.
- BORÉM F. M. et al. Coffee sensory quality study based on spatial distribution in the mantiqueira mountain region of Brazil. **Journal of Sensory Studies**, 35(2):e12552, 2020.
- BREIMAN, L. Bagging predictors. **Machine Learning**, 24(2):123-140, 1996.
- BROOKES, C. Consumer satisfaction: Advancements in theory, modeling, and empirical findings. **International Journal of Marketing Research**, 56(4):123-145, 2014.
- BUENAVENTURA-SERRANO, C. E.; CASTAÑOCASTRILLÓN, J. J. Influencia de la altitud en la calidad de la bebida de muestras de café procedente del ecotopo 206B en Colombia. **Cenicafé**, 53(2):119-131, 2002.
- CALCAGNOTO, L. R.; CIRILLO, M. A. Simulação da qualidade de cafês utilizando critérios da ABIC na discriminação de regiões com diferentes estruturas de dependência espacial. **Revista Brasileira de Computação Aplicada**, 16(2):31-44, 2024.
- CLAYTON, M. K. How should we achieve high-quality reporting of statistics in scientific journals? A commentary on "Guidelines for reporting statistics in journals published by the American Physiological Society". **Advances in Physiology Education**, 31(4):302-304, 2007.
- COOK, D.; SWAYNE, D. F. **Interactive and dynamic graphics for data analysis: With R and GGobi**. New York: Springer, 2007. 300p.
- COOK, D.; BUJA, A.; CABRERA, J. Projection pursuit indexes based on orthonormal function expansions. **Journal of Computational and Graphical Statistics**, 2(3):225-250, 1993.
- COSTEL, E. A. Comparison of sensory methods in quality control. **Food Quality and Preference**, 13(6):341-353, 2002.
- DAI, H. et al. Explaining consumer satisfaction of services: The role of innovativeness and emotion in an electronic mediated environment. **Decision Support Systems**, 70:97-106, 2015.
- DAVISON, A. C.; HINKLEY, D. V. **Bootstrap methods and their application**. Cambridge: Cambridge University Press, 1997. 582p.
- DECAZY, F. et al. Quality 568 of different Honduran coffees in relation to several environments. **Journal of Food Science**, 68(7):2356-2361, 2003.
- DEMPSTER, A. P.; LAIRD, N. M. E.; RUBIN, D. B. Maximum likelihood from in-complete data via the EM algorithm. **Journal of the Royal Statistical Society**, 39(1):1-22, 1977.
- FERREIRA, H. A. et al. Monte Carlo simulation and importance sampling applied to sensorial analysis validation of special coffee quality. **Revista Ciência Agronômica**, 52(2):1-6, 2021.
- FILZMOSER, P.; MARONNA, R.; WERNER, M. Outlier identification in high dimensions. **Computational Statistics and Data Analysis**, 52(3):1694-1711, 2008.
- FREUND, Y.; SCHAPIRE, R. E. **Boosting: Foundations and algorithms**. Cambridge, MA: MIT Press, 2012. 544p.
- FRIEDMAN, J. H.; TUKEY, J. W. A projection pursuit algorithm for exploratory data analysis. **IEEE Transaction on Computers**, 23(9):881-890, 1974.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA- IBGE. **Cidades**. Rio de Janeiro: IBGE, 2009. Available in: <http://geoftp.ibge.gov.br/MME2007/MG/CARMO%20DE%20MINAS.pdf>. Access in: September 19, 2024.
- JOHNSON, S. B.; BERGER, R. D. On the status of statistics in phytopathology. **Phytopathology**, 72(8):1014-1015, 1982.
- LANKTON, N. K.; MCKNIGHT, H. D. Examining two expectation disconfirmation theory models: Assimilation and asymmetry effects. **Journal of the association for information systems**, 13(2):88-115, 2012.
- LI, Y. et al. Algorithm for gaussian mixture model based wireless channel multipath clustering. **International Journal of Antennas and Propagation**, 4:1-10, 2019.
- LIM, J.; WOOD, A.; GREEN, B. G. Derivation and evaluation of a labeled hedonic scale. **Chemical Senses**, 34(9):739-751, 2009.
- LIM, J. Hedonic scaling: A review of methods and theory. **Food Quality and Preference**, 22(8):733-747, 2011.

- LIM, J.; FUJIMARU, T. Evaluation of the labeled hedonic scale under different experimental conditions. **Food Quality and Preference**, 21(5):521-530, 2010.
- LINGLE, T. R. **The coffee cupper's handbook**: Systematic guide to the sensory evaluation of coffee's flavor. Long Beach: Specialty Coffee Association of America. 2011. 66p.
- LISKA, G. R. et al. Avaliação de painéis sensoriais com consumidores de bebidas de cafés especiais utilizando o método boosting na análise discriminante. **Semina - Ciências Agrárias**, 36(6):3671-3679, 2015.
- MAKAREM, S. C.; JAE, H. Consumer boycott behavior: An exploratory analysis of twitter feeds. **Journal of Consumer Affairs**, 50(1):193-223, 2016.
- MOSKOWITZ, H. R.; SIDEL, J. L. Magnitude and hedonic scales of food acceptability. **Journal of Food Science**, 36(4):677-680, 1971.
- OLIVEIRA, L. K. et al. Feasibility of the expectation-maximization algorithm for assessing individuals with different sensory perceptions in discrimination of specialty coffees. **Acta Scientiarum Technology**, 45:e60184, 2023.
- OLIVEIRA, L. M. et al. Machine Learning techniques in multiclass problems with application in sensorial analysis. **Concurrency and Computation**, 32(2):1-12, 2019.
- OLIVER, R. **Satisfaction**: A behavioral perspective on the consumer. New York, Routledge: Taylor and Francis Group, 2015. 544p.
- ONOFRI, U. M. et al. Current statistical issues in weed research. **European Weed Research Society Weed Research**, 50(1):5-24, 2009.
- OSSANI, P. C. et al. Unsupervised classification of specialty coffees in homogeneous sensory attributes through machine learning. **Coffee Science**, 15:e151780, 2020.
- OSSANI, P. C.; RAMOS, M. F.; CIRILLO, M. A. Proposition of a new index for projection pursuit in the multiple factor analysis. **Computational and Mathematical Methods**, 3(4):1-19, 2020.
- PEÑA, D.; PRIETO, F. Multivariate outlier detection and robust covariance matrix estimation. **Technometrics**, 43(3):286-310, 2001.
- PEREIRA, P. P. et al. The consistency in the sensory analysis of coffees using Q-graders. **European Food Research and Technology**, 243:1545-1554, 2017.
- PERYAM, D. R.; PILGRIM, F. J. Hedonic scale method of measuring food preference. **Food Technology**, 11:9-14, 1957.
- RAMOS, M. F. et al. Discrimination of the sensory quality of the *Coffea arabica* L. (cv. Bourbon Amarelo) produced in different altitudes using decision trees obtained by the CHAID method. **Journal of the Science of Food and Agriculture**, 96(10):3543-3551, 2016.
- RESENDE, M.; CIRILLO, M. A.; BORÉM, F. M. An index to evaluate the acceptance of specialty coffees in consumer groups. **Engenharia Agrícola**, 40(5):624-630, 2020.
- RESENDE, M.; BORÉM, F. M.; CIRILLO, M. A. Case study of modeling covariance between external factors and sensory perception of coffee, **Coffee Science**, 18:e182112, 2023.
- SCHAPIRE, R. E. The strength of weak learnability. **Machine learning**, 5:197-227, 1990.
- SCHUTZ, H. G.; CARDELLO, A. V. A labeled affective magnitude (LAM) scale for assessing food liking/disliking. **Journal of Sensory Studies**, 16(2):117-159, 2001.
- SINGHAM, P.; BIRWAL, P.; YADAV, B. K. Importance of objective and subjective measurement of food quality and their inter-relationship. **Food Processing and Technology**, 6(9):2-7, 2015.
- SORENSEN, B.; ANDERSEN, C.; ANDERSEN, M. P. Consumer satisfaction and confirmation of habits of comprehension: the effects of inductive print advertisements – Peircean comments. **Semiotica**, 199:159-174, 2014.
- VELÁSQUEZ, S. et al. Classification of the Maturity stage of coffee cherries using comparative feature and machine learning, **Coffee Science**, 16:e161710, 2021.
- VILLANUEVA, N. D. M.; DA SILVA, M. A. A. P. Comparative performance of the nine-point hedonic, hybrid and self-adjusting scales in the generation of internal preference maps. **Food Quality and Preference**, 20(1):1-12, 2009.
- WAGENMAKERS, E. J. et al. Seven steps toward more transparency in statistical practice. **Nature Human Behaviour**, 5:1473-1480, 2021.
- WORKU, M. et al. Effect of altitude on biochemical composition and quality of green arabica coffee beans can be affected by shade and postharvest processing method. **Food Research International**, 105:278-285, 2018.