Brazilian Journal of Biometrics, **41**, 30–43 (2023)
doi:10.28951/bjb.v41i1.590

BRAZILIAN JOURNAL OF BIOMΣTRICS
ISSN:2764-5290

ARTICLE

Sensorial analysis of categorized data of special coffee to identify similar crop seasons pairs using Kappa

Jackelya Araujo da Silva,*[,1] Marcelo Angelo Cirillo,[2] Flávio Meira Borém,[3] Diego Egídio Ribeiro,[4] and Lourenço Manuel[5]

[1]Department of Statistics, Federal University of Piaui, Teresina-PI, Brazil.
[2]Department of Statistics, Federal University of Lavras, Lavras-MG, Brazil.
[3]Department of Agricultural Engineering, Federal University of Lavras, Lavras-MG, Brazil.
[4]PhD in Agricultural Engineering, Department of Agricultural Engineering, Federal University of Lavras, Lavras-MG, Brazil.
[5]Center of Excellence in Agri-Food Systems and Nutrition, Eduardo Mondlane University, Maputo, Mozambique.
*Corresponding author. Email: jackelya@ufpi.edu.br

(Received: November 13,2021; Revised: March 21,2022; Accepted: May 09,2022; Published: March 1,2023)

Abstract

This paper presents the proposal of a statistical method to analyse dependent agreement data with categorical ordinal responses for a longitudinal study in sensorial analysis of special coffee. The assessment of sensory attributes of special coffees were carried out by certified raters using a continuous scale of grades. The approach aimed to applying data categorization methods commonly used in machine learning which generated not only a concise summary of continuous attributes to describe the data but also allowed to maximize the agreement grades in a longitudinal study. A previous analysis was carried out to identify the similarity of grades in all sample unities. The categorization allowed the construction of marginal models for all distinct pairs time observed in the longitudinal study for modeling the concordance correlations kappa. It also enabled to conclude that samples of harvests related to yellow grain fruits have similar sensorial characteristics. Higher altitudes are significantly favorable to obtain samples with similar sensorial characteristics identifying the set of covariates which contributed either in positive or negative way while estimating kappa.

Keywords: Sensory attribute; Coffee blends; GEE; Kappa Coefficient; Ordered Categorical data.

1. Introduction

Generalized Estimating Equations (GEE) proposed by Liang and Zeger (1986) are class of models useful to analyse both categorical and continuous correlated response variables (Zeger & Liang, 1986). Under the GEE approach it is not necessary to specify the joint distribution of the repeated responses. However, the GEE methodology requires the specification of the first and second moments of the joint distribution(Zeger & Liang, 1986).For ordinal data in GEE the odds ratio are used to measure the association between two observations within the same group of response.

Williamson *et al.* (1995) studying some eyes characteristics in a group of subjects they noticed that observations of interested responses could be collected in each eye of each individual as well as at the individual level. In their study, the authors aimed to evaluate the severity of diabetic retinopathy was originally graded on a 10-point scale based on the identification of risk factors associated with the disease. The classifications on the original scale were then used to define four ordinal categories as follows: none, mild, moderate and proliferative.

Kappa coefficient (κ) is the common method used to assess the degree of agreement or reproducibility between two set of data. Cohen (1960) applied Kappa coefficient to determine the agreement of binary responses between raters while Fleiss (1971) also used Kappa for categorical data to measure the degree of concordance between raters.

A classification statistic such as κ can be used to assess the goodness of fit in GEE regression models for categorical responses. In this case kappa is a measure used to evaluate how the categorical responses are well predicted through the GEE fitted model (Akanda *et al.,* 2005).

Another application of kappa modelling was proposed by Williamson *et al.* (2000). The authors developed a method to analyse dependent agreement data with categorical responses using covariates. In their paper, Williamson *et al.* (2000) have presented two sets of estimating equations in which the second set aimed to modelling the paired associations of classifications with the kappa coefficient as a metric.

Gonin *et al.* (2000) have proposed to consider two set of estimating equations. The first set is used to modelling the marginal distributions of categorical responses while the second set of estimating equation is introduced to estimate κ by modelling binary variables in order to depict the agreement between responses.

In order to test the equality of two dependent κ statistics when there are two raters and a binary response, Donner *et al.* (2000) have considered that each sample of subjects was classified in two distinct context and the goal was to compare the observed agreement level between raters and within each group of raters.

Klar *et al.* (2000) proposed the inclusion of covariates under GEE approach to model pairs of associations using the κ coefficient with binary responses obtained from a random number of raters by individual rated. Therefore, the approach also allows identifying predictive covariates for kappa.

Following the same approach previously described, we do propose the application of kappa modelling in GEE for ordinary categorical data.

Although the sensorial responses are continuous grades, in our approach we propose to study them as categories.There are several categorization methods for continuous variables and most of them have their genesis in machine learning (Kerber, 1992; Kurban & Cios, 2004; Tsai *et al.,* 2008). The discretization algorithms have shown an important role in data mining and to produce new knowledge. They produce a concise summary of continuous attributes helping specialist to easily understand the data. Furthermore, the algorithms allow the machine learning to be more quick and precise.

The discretization method CAIM (Class – Attribute Interdependency Maximization) proposed by Kurgan & Cios (2004), consists in transforming the values of a continuous variable in a finite number of intervals in which each interval is associated to a discrete numeric value. The method automatically selects a set of discrete intervals and at the same time the range of those intervals is estimated based on the interdependency between classes and the values of the variable.

The well known K-means method, aims to divide in clusters a random sample with size N such that the sum square within each cluster is minimized. Thus, the length and dimension of all clusters are built such that they are asymptotically similar among them.

Coffee production is the result of the relationship between local environmental conditions and coffee cultivars that grow in this place(Scholz *et al.,* 2018). Specialty coffees are recognized for their differentiated quality, and they are valued the more rare and exotic their sensory profiles are. The quality of the coffee drink is influenced by various factors, the most notable being those related to the environment, the processing and the genetic constitution of the plants (species/cultivating) (Giomo & Borém, 2011).

The studies of specialty coffees are associated with many variables of interest that explain the quality of the drink. Avelino *et al* (2005), assessed the effects of slope exposure, altitude and yield on several cup quality criteria of coffees from two altitude terroirs of Costa Rica, Orosi and Santa Maria de Dota (between 1,020 and 1,250 m above sea level). The effects of altitude in association with temperature have been described in several studies as the main factors affecting coffee quality(Borém *et al.,* 2019; Joet *et al.,* 2010; Decazy *et al.,* 2018). The variability in environmental conditions is sufficient to modify the green and roasted coffee bean characteristics and sensory attributes (Scholz *et al.,* 2018).

Other aspects of interest in describing the quality of coffees are studies related to harvesting, post-harvesting, grain processing and the biochemical composition of specialty coffees(Knopp *et al.,* 2006; Duarte *et al.,* 2010; Borém & Shuler, 2014; Scholz *et al.,* 2018; Tolessa *et al.,* 2017).

Here we are concerned with modeling dependent categorical agreement for all distinct pairs of time, considering the repeated responses for the pooled samples assessed by raters.

The proposal for the differentiated use of κ modeling in the construction of several marginal models focuses on the analysis of the identification of pairs of crop seasons that have more similar longitudinal sensory characteristics.

2. Materials and Methods

Consider a longitudinal study with n subjects, J_i agreement observers, and T_i assessments to be based on a categorical rating scale with C categories: $r_1, r_2, \ldots, r_C$. In addition let Y_{ijtc} be a binary outcome as described by Heagerty & Zeger (1996), that is, denoting $\boldsymbol{O}_{ij} = \{\boldsymbol{O}^t_{ijt}, \ldots, \boldsymbol{O}^t_{ijt}\}^t$ a vector of ordinal measures for the jth observers(judges) at the ith subjects, O_{ijt}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, J$ observers and $t = 1, 2, \ldots, T$ observation times. The ordinal measure $O_{ijt} = c$, $c = r_1, r_2, \ldots, r_C$, with C categories of responses correspond to a vector of accumulated indicator variables $Y_{ijtc} = I_{(O_{ijt} > c)}$, $c \in \{r_1, r_2, \ldots, r_{c-1}\}$ such that

$$Y_{ijtc} = \begin{cases} 1, & \text{if } O_{ijt} > c \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

In the above, $y_{ijt} = (y_{ijt1}, y_{ijt2}, \ldots, y_{ijtC})^t$ represents the ratings of the jth observer on the ith subject at time t. Thus, $y_{it} = (y^t_{i1t}, y^t_{i2t}, \ldots, y^t_{iJt})^t$ represents the rating on the ith subject for all the J judges and assessments.

Generalized estimating equation methodology is useful for analysing correlated response data and there is no need to specify a joint distribution for the responses (Liang & Zeger, 1986; Zeger & Liang, 1986). In the following method, the response of interest is a categorical outcome with C categories and the response vector $\boldsymbol{Y}_i$ with $J_i T_i (C-1) \times 1$ elements, defined in (1) consist of the binary random variables Y_{ijtc}. Let $\boldsymbol{X}_{it} = \{\boldsymbol{x}^t_{i1t}, \ldots, \boldsymbol{x}^t_{iJt}\}$ represent the $J_i \times p$ matrix of covariates for the ith subjects for all the J judges on the tth time.

The marginal distribution of $\boldsymbol{Y}_{it}$ is Bernoulli with $\pi_{itc} = P(\boldsymbol{O}_{ijt} = c) = P(\boldsymbol{Y}_{ijtc} = 1 | \boldsymbol{x}^t_{ijt}, \boldsymbol{\beta})$ such that

$$ln\left(\frac{\pi_{itc}}{1 - \pi_{itc}}\right) = \boldsymbol{x}^t_{ijt}\boldsymbol{\beta} \tag{2}$$

and where $\boldsymbol{\beta}$ is $p \times 1$ vector of parameters. Thus, the first set of estimating equations for the marginal distribution of the response given by

$$\nu_1(\beta) = \sum_{i=1}^{n} \boldsymbol{D}_i^t \boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \pi_i(\beta)) = \boldsymbol{0} \tag{3}$$

where $\boldsymbol{D}_i = \boldsymbol{D}_i(\beta) = d\pi_i(\beta)/d\beta$ and $\boldsymbol{V}_i = Var(\boldsymbol{Y}_i)$ is a working covariance matrix of $\boldsymbol{Y}_i$ (Liang & Zeger, 1986; Zeger & Liang, 1986).

Under the inference for Kappa in longitudinal and clustered data reported by Ma *et al.* (2008) which introduces a new class of estimates for kappa indexes and their generalizations, we consider an analysis to detect changes in responses based on observation times. The κ index will be used to measure similarity between evaluations from raters based on a categorical classification scale to all distinct pairs of time.

Therefore, if two independent raters agree they generate responses that do not identify changes at the same evaluated subjects in distinct time. Then, the proportions reported in the diagonal of Table 1 are interpreted as common responses of subjects on observation times.

Table 1. Proportion of agreement based on combination of pairs of times from a categorical variable with $C = (r_1, r_2, \ldots, r_C)$ categories; $\hat{\pi}_{.ist}$ denote the cell proportion, and $\hat{\phi}_{.Cs}(\hat{\phi}_{.Ct})$ the marginal cell corresponding proportion

	Time $t(Y_{i2t})$			
Time $s(Y_{i1s})$	r_1	$\ldots$	r_C	Marginal
r_1	$\hat{\pi}_{11ist}$	$\ldots$	$\hat{\pi}_{1Cist}$	$\hat{\phi}_{11s}$
$\ldots$	$\ldots$	$\ddots$	$\ldots$	$\ldots$
r_C	$\hat{\pi}_{C1ist}$	$\ldots$	$\hat{\pi}_{CCist}$	$\hat{\phi}_{1Cs}$
Marginal	$\hat{\phi}_{21t}$	$\ldots$	$\hat{\phi}_{2Ct}$	1

Within the above context, the Kappa coefficient has been used to determine the agreement of binary and categorical outcomes between pairs of responses of the ith subject in time. The kappa coefficient is an agreement measure limited in $(-1, 1)$. Its interpretation is as follows: When Kappa is 0 indicates no agreement beyond chance. Values of kappa near 1 suggest non-randomness of responses. Negative κ values indicate that the observed agreement was less than that expected by chance and therefore there is disagreement between responses. However, in this situation the κ value is not interpreted as a disagreement intensity measure.

Kappa is based on the number of agreed responses, that is, the number of cases in which the results are the same between times of all raters and it measures the degree of agreement beyond what would be expected only by random effect.

The general expression for the kappa statistic is:

$$\kappa_{ist} = \frac{P_{oist} - P_{eist}}{1 - P_{eist}}, \quad 1 \leq s < t \leq T \tag{4}$$

where P_{eist} is the probability that the pair of categorical variables are the same when the independence assumption holds and P_{oist} is the joint probability that the pair of responses are the same for the combinations of times s and t.

In the classic approach, κ_{ist} is estimated by substituting the sample proportions in place of the respective parameters in (4), so that $P_{eist} = \sum_{c=1}^{C} P(Y_{isc} = 1)P(Y_{itc} = 1)$ and $P_{oist} = \sum_{c=1}^{C} P(Y_{isc} = 1, Y_{itc} = 1)$, where $P(Y_{isc} = 1)$ and $P(Y_{itc} = 1)$ are the marginal probabilities of subject i fall in the sth e tth time and $P(Y_{isc} = 1, Y_{itc} = 1)$ is the probability that both responses for subject i fall in the cth category.

The inference for Kappas for longitudinal Study proposed by Ma *et al.* (2008) which focus on inference about pairwise comparisons of multiple observers' ratings, which in a cross-sectional study with J judges, they can assess overall observer agreement by averaging over all between rater kappas: $k_{ave} = \binom{J}{2}^{-1} \sum_{(m,l)\in A} \kappa_{ml}$, where κ_{ml} is the kappa between judges m and l and $A = \{(m,l) : 1 \leq m < l \leq J\}$. We can estimate κ_{ave} by substituting estimates $\hat{\kappa}_{ml}$ of κ_{ml}. Following the similar approach, in the study of preference identification of time combination pairs (s,t), $s < t$ by judges, we propose estimating the concordance correlations to all distinct pairs of time considering the concordance between judges in each time.

Thus, for the agreed responses we will assume P_{oist}, $s \leq t$, such that the difference $P_{oist} - P_{eist}$ is the excess of the agreement over that expected by chance. Therefore, the marginal fitted models coefficients represent changes of kappas related to covariates.

Note that $P(Y_{ijsc} = 1)$ is the mean of all samples from the jth rater at the jth time classified on c category of grades. The κ in the case of the joint distribution is then sometimes referred to as the common correlation model since κ can also be derived as the correlation between responses the special case of this joint distribution which arises when there are two raters per subject and when there is marginal homogeneity: $P(Y_{ijs}) = P(Y_{ih_s})$ (Donner & Klar, 1996).

From the marginal model proposed by Heagerty & Zeger (1996) in the study of associations between responses in an odds ration regression model, we will use the following link function

$$g(\kappa_{ist}) = ln\left(\frac{1+\kappa_{ist}}{1-\kappa_{ist}}\right) = \boldsymbol{z}'_{ist}\gamma \tag{5}$$

to modelling κ in terms of covariates, where $\boldsymbol{z}'_{ist}$ is a vector of covariates.

From the suggestion of Liang *et al.* (1992) and Williamson *et al.* (2000), the multiplication of indicators variables describes the concordance between responses at the sth and tth times of all raters, i.e, $U_{ist} = \sum_c^C Y_{isc} Y_{itc}$, with $E(U_{ist}) = P_{oist}$.

For the J_i responses of subject i, there are $J_i(J_i-1)/2$ distinct pairs of responses for each combination of two pairs of time, see Table 1. Fixing distinct pairs of times s and t, with $(s < t)$, κ is then estimated by solving the following set of estimating equations:

$$\boldsymbol{\nu}_2(\beta,\gamma) = \sum_{i=1}^{n} \left[\frac{\partial \boldsymbol{P}_{oist}}{\partial \alpha^*}\right]^t \boldsymbol{W}_{ist}^{-1}(\boldsymbol{U}_{ist} - \boldsymbol{P}_{oist}(\beta,\gamma)) = \boldsymbol{0} \tag{6}$$

where $\boldsymbol{\gamma}$ are parameters estimates of the marginal model κ_{st} for each of $T_i(T_i-1)/2$ distinct pairs of time combinations and α^* is a vector of association parameters estimates between responses of the jth rater at the ith subject for pair of times. Thus, $\boldsymbol{W}_{ist}$ is the working covariance matrix of $\boldsymbol{U}_{ist}$.

The modified Kappa modeling in GEE proposed in this paper was applied to a dataset from a sensory analysis in a longitudinal experiment with repeated multiple responses.

A data set was obtained with the project approved in the public notice CNPq/MAPA 064/2007 (Borém, 2007). The experiment was conducted using specialty coffees from four varieties (Yellow Bourbon, Yellow Catuai, Acaia and Mundo Novo) harvested in four agricultural seasons (2010/11, 2011/12, 2012/13 and 2013/14) in commercial farming of properties located in municipality of Carmo de Minas, Minas Gerais Brazil.

The experimental design was based in the study of genetics environment interaction including processing type. There were 288 tastings for each of the harvests, from three altitudes (lower than $1000m$, between $1000m - 1200m$ and higher than $1200m$) in two different processing ways (dry and wet) and formed by two slope groups(sun and shade).

3. Results and discussion

3.1 Application

The sample sensorial assessments were carried out in each harvest time by four well trained raters qualified as certified judges of special coffee by SCA based on the methodology proposed by Lingle (2011). The raters provided grades for the following attributes: fragance, body, aftertaste, sweetness, acidity, flavor, clean cup, balance, smoothness and overall impression. The final results of sensorial assessment were obtained using SCA grades scale.

Sensory analyses were carried out by four independent raters. However, the results of their sensory perceptions are correlated in each rated sample. For notation simplification we denote the assessed harvesting years by 1, 2, 3 and 4.

The distribution of final grades of sensorial assessments is presented in Figures 1 and 2.

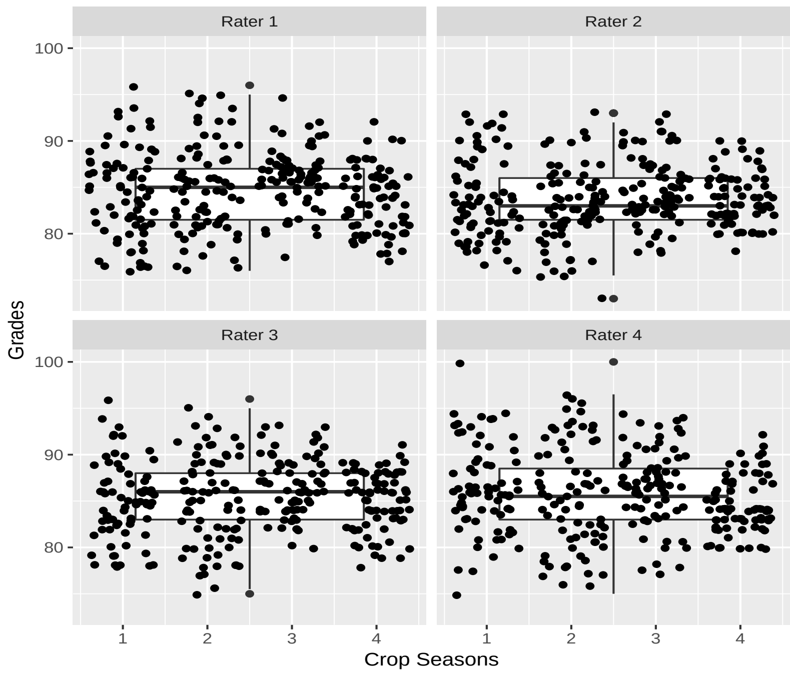

Figure 1. Distribution of grades for each rater across the four crop seasons

Figure 1 reports the grades of each rater in each of the harvesting years. It is noticed that there is a high variability in the grades across the harvesting years. For instance, in the harvesting year 4 (2013/14) the cloud of points indicates more concentration of sensorial grades in lower scales when compared to other harvesting years.

Figure 2 shows the distribution of sensorial grades across the harvesting years for each category of covariates used in the study. The low variability on the sensorial grades is noticed in harvesting year 4 due to the high concentration of the cloud points in this season. Moreover, the highest sensorial grades are observed in higher altitudes where the greatest variability is verified in harvesting year 2 (2011/12). Analyzing the distribution of grades taking into account the genotype, figure 2 shows a similarity in the cloud of points across the harvesting years for the two genotypes analyzed (Yellow Borbon and Acaia).

Described the behavior of sensorial grades as previously reported in Figures 1 and 2, the K-means method was applied considering univariate case in order to categorize the sensorial grades.

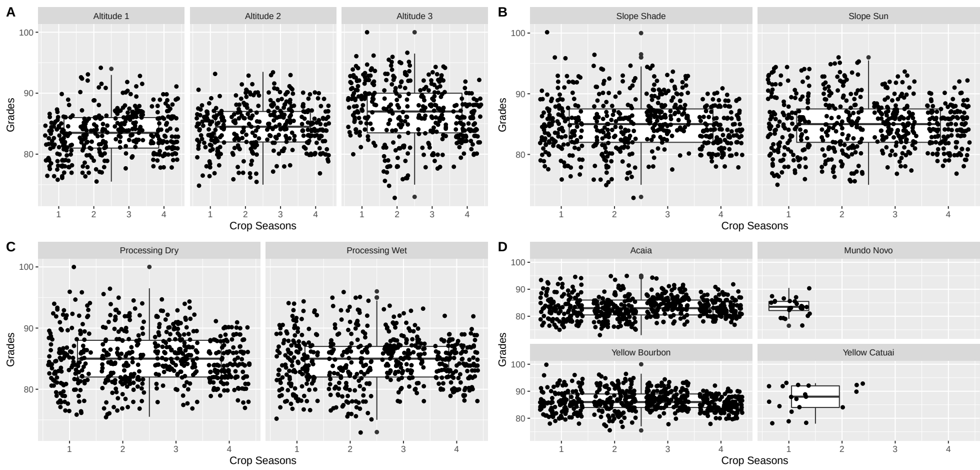

Figure 2. Distribution of sensorial grades according of covariates:**A** – Altitude, **B** – Slope, **C** – Processing type and **D** – Genotype.

For a sample size N, the algorithm divides the sample in k groups minimizing the sum square within each group such that they are asymptotically similar between them.

The construction of clusters asymptotically with the same size will contribute significantly in the identification of comparisons between harvesting years and the characterization of similarities between those harvesting years using GEE methodology. Analyzing the data and applying the Elbow method it was possible to reach an optimal number of clusters equals four ($k = 4$) as reported in Figure 3.

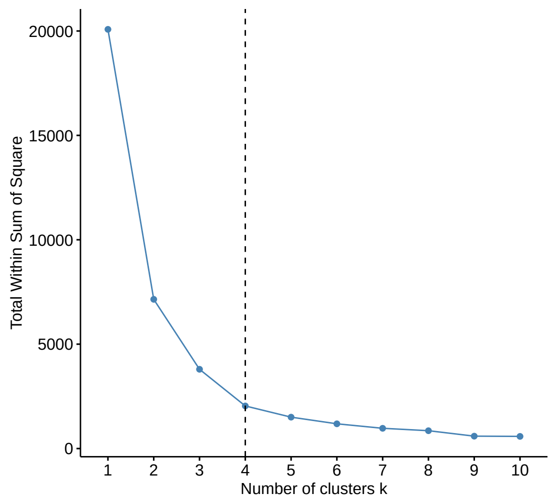

Figure 3. Optimal numbers of clusters

According to Lingle (2011), the SCA approach that coffee quality is quantified by a scale of grades varying from zero to 100 points. Coffee graded between 85 – 89 and 80 – 84 are respectively classified as special (excellent) and special (very good); coffee with grades between 75 – 79, receive regular classification, although rated as good quality coffee; Finally, coffee rated in 70 to 75 interval are classified as average (weak).

The data categorization through K means method in each combination of pairs of harvesting years will generate different groups of grades which contribute to identify the covariates that maximize κ estimated through the fitted model. Therefore, the data categorization will help to summarize the grades behavior considering the agreement proportion between raters.

Figure 4 reports the midpoints of classes or the grades categories reached while applying K-means method. The number of categories obtained in each combination of pairs of raters or harvesting years varies.

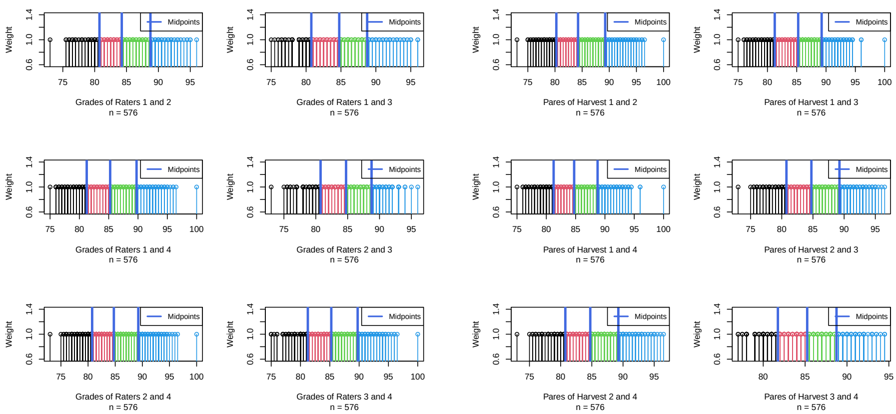

Figure 4. Midpoints of categories for the combination of pairs of raters and pairs of crop seasons

Table 2 describes the common agreement proportions for each combination pair of raters across the four harvesting years as depicted in Figure 4. The results are fair given that sensorial analysis aims reaching individual specific perceptions related to sensorial attributes that contributes to final grades of tasted samples.

However, it is expected a slight variation in the classification of grades between the raters. Note that when the number of clusters is high, the agreement proportions were diluted in grades categories and therefore it would be difficult to achieve higher agreement proportions.

Table 2. Proportion of agreement between all pairs distinct of raters to each harvest

Pairs of raters	Harvest 1	2	3	4
1 - 2	0.5555	0.4583	0.4583	0.3750
2 - 3	0.4305	0.4305	0.3333	0.4444
3 - 4	0.4166	0.4166	0.4028	0.4861
1 - 3	0.4722	0.4722	0.5972	0.4583
2 - 4	0.4583	0.3750	0.5000	0.4444
1 - 4	0.5277	0.5139	0.5972	0.4167

Table 2 shows that the categorization enabled to notice that rater 1 has 60% of sensorial grades agreeing with raters 3 and 4 for harvesting year 3 (2012/13). However, when rater 1 is compared to rater 2, the agreement proportion drops slightly to 56% when the deemed harvesting year is 1 (2010/11).

Furthermore, the data categorization enabled estimating the following agreement correlations for all pairs of harvesting years $\hat{k}_{st} = (\hat{k}_{12}, \hat{k}_{13}, \hat{k}_{14}, \hat{k}_{23}, \hat{k}_{24}, \hat{k}_{34}) = (0.1305, 0.1311, 0.0819, 0.0248, 0.0866, -0.0061)$.

It is noticed that a negative agreement correlation was observed for harvesting pairs 3 and 4 indicating that samples from those harvests have differences in sensorial characteristics for all pairs of raters. Nonetheless, the $\hat{k}$ estimates are very low in general and they were reached without taking in account the effect of covariates depicted in Figure 2.

Thus, the grades were evaluated to verify if the sensory attributes the sensory attributes of special coffee have changed significantly along the time or if the sensory analysis of samples from pairs of crop seasons are more similar between them due to some variables such as altitude, processing ways and slope groups.

Thus, six distinct pairs of crop seasons were generated ($T_i(T_i-1)/2 = 6$). Solving the equation (6) considering that adjustment for covariates associated with kappa can be accomplished using the model

$$g(\kappa_{ist}) = ln\left(\frac{1+\kappa_{ist}}{1-\kappa_{ist}}\right) = \gamma_1 Altitude + \gamma_2 Slope + \gamma_3 Processing + \gamma_4 Genotype. \quad (7)$$

where represent the marginal model for tasting notes, $i = 1, 2, \ldots, 288$, evaluated at times $t = 1, 2, 3, 4$, with the covariates altitude, slope, processing, and genotype.

Under the null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$ at 5% significance level, we carried out model selection procedure in order to identify covariates affecting significantly the models. Therefore, our analyzes are based on those covariates. For instance, for the pair of harvesting years 1 and 2 (2010/11 and 2011/12) the models identified all covariates as significant except the processing type where kappa estimate was $\hat{k} = 0.342$.

However, when the analysis takes into account only higher altitudes and shade slope kappa estimates for genotypes Yellow Bourbon and Catuai were 0.802 and 0.501, respectively.

This information indicates that both harvesting years have similarities in sensorial characteristics mainly in altitudes greater than $1,200m$. Moreover, model estimates for harvesting years 1 and 3 has also shown that there is a greater agreement between those harvesting years for yellow grains and κ estimates for genotype Yellow Catuai and Yellow Bourbon were 0.603 and 0.453, respectively.

For the harvesting years 1 and 4 (2010/11 and 2013/14) the processing type covariate has a negative influence in the agreement between such harvesting years. The kappa coefficient was estimated in $\hat{k}_{14} = 0.0046$, indicating that the agreement between the harvesting years was that expected randomly.

Therefore the selected model is only explained by altitude, processing type and genotypes Acaiá and Yellow Bourbon. The wet processing type combined with high altitudes shows that the deemed genotypes Acaiá and Yellow Bourbon disagree for both harvesting years and the Kappa estimates were –0.00518 and –0.0496, respectively. Moreover, the wet processing type was significant so that for genotype Yellow Bourbon cultivated in altitudes above $1,200m$ considering shade slope reached κ estimates equals 0.623 and 0.214, for pairs of harvesting years 2 – 4 and 3 – 4, respectively.

Analyzing all six pairs of harvesting years it is noticed that the harvesting pair 2 – 3 has only altitude as the covariate contributing in the model for identification of kappas. The Kappa estimative in this model was $\hat{k}_{23} = -0.0375$ indicating disagreement. In other words, harvesting years 2 and 3 have differences in sensorial characteristics.

We observe that altitude has an important role to determine concordance correlations estimates of pairs of crop seasons based on groups of raters. Under the natural processing way, samples of genotypes such as Yellow Bourbon and Catuai have the highest estimates of κ_{st}.

The results highlighted for the yellow coffee fruit varieties (Yellow Bourbon) presented in this manuscript corroborate with Giomo & Borém (2011) when stating that even if any cultivar of *Coffea arabica L.* has the potential to produce high quality coffees, more valued sensory profiles, with a high reputation in the specialty coffee market, have been found more frequently in ancient cultivars such as 'Yellow Bourbon'

Our analysis proposal identified that the covariable altitude combined with the genotypes representing the yellow fruits (Yellow Bourbon and Catuaí) had the best κ agreement values for the pairs of crops.

The κ values for sensory agreement of the samples for the pairs of crops when combined with the type of processing also present higher concordances/agreements when compared to other genotypes.

According to Borém *et al.* (2020), in the study on the sensory characteristics of coffee corresponding to the cultivation environment with altitudes above $1,050m$, it was observed that the genotypes representing red fruits did not present such correspondence. However, for varieties of red coffee fruits, this correspondence only occurred when submitted to a wet processing method. In our analysis, it was noticed that the highest κ values were due to yellow bourbon genotypes and that for red coffee fruit varieties, κ values show a slight improvement if combined with the wet processing method.

Our proposed methodology is focused in obtaining marginal models to study longitudinal correlations considering an experiment in sensory analysis where there are many pooled samples of repeated measures groups.

The method has allowed measuring the agreement between responses and therefore identify preferences on rated samples.

Furthermore, using marginal models for all distinct pairs in time has provided a practical interpretation of descriptive analyses for covariates concordance correlations. Through the sensory analysis study it was also possible to identify if the time effects induce sensory changes on special coffee from data categorization when raters are equally well trained (balanced).

The logistic models approach are presented to identify the goodness of fit in GEE using Kappa (Akanda *et al.,* 2005) where if the fitted model predicts the categorical response, $k = 1$ or in studies of covariate selections in logistic models (Klar *et al.,* 2000). Moreover, is applied to determine of the intraclass correlation coefficients measures the tendency for correlated measurements to be more alike, whereas ignores the ordering of categories with ordinal responses when it models agreement with κ (Williamson *et al.,* 2000).

The Kappa modelling method applying marginal models in GEE has allowed comparing the evolution of time effects interpretations in sensory analysis of special coffee. It also supported the identification of covariates with positive or negative contribution to estimate de concordance correlations as well as to discriminate what pairs of crop seasons are more similar when classified in categories of grades.

In essence, the method proposed in this manuscript was discussed as a possibility for using Kappa modeling by Williamson *et al.* (2000). In discussions on modeling kappa for measuring dependent categorical agreement data suggested that the method can be considered a stratified analysis, in which κ is estimated for each covariate cross-classification, modeling the marginal distribution and the association distribution, and can be interpreted accordingly.

The effect of fitting of logistic regression models has implications about kappa given that κ is a function of marginal probabilities. The marginal distribution must be correctly modeled to ensure unbiased estimation of κ and all covariates entered into the association modelling should also be examined in the marginal model. However, the differences $\hat{P}_{oist} - \hat{P}_{eist}$ allow identifying if pairs of agreed responses occur by random effect or if there is difference between what is observed on the sample and what is expected to occur to classify that sample in pairs of time s and t, $s < t$.

Conclusions

The data categorization of sensorial assessments using K-means methods based on univariate case enabled to group the grades in different classes with dimensions asymptotically equals and contributed to homogeneous distribution of grades categories. In addition, the categorization using pairs of harvesting years does not depend of same classes limit.

The methodology of using Kappa modeling for all distinct pairs of crops allowed the construction of marginal models that explain the interaction between environmental and genetic covariates. It was verified that altitude contributed to the highest Kappa values.The altitude covariate combined with the genotypes Yellow Bourbon and Catuai presented pairs of crops that sampling sensory similar longitudinal characteristics.

The results show that the type of grain processing plays an important role in determining the estimates of agreement correlations of the pairs of crops. For the pair of harvesting years 1 – 4 the processing type covariate has shown a significant contribution in the identification of disagreement in sensorial assessment between the deemed harvesting years where the kappa estimates were negative.

The categorization aimed to notice that the pairs of harvesting years 2010/11 – 2011/12 and 2011/12 – 2013/14 have shown kappa estimates greater than 0.5 for genotypes Yellow Acaia, Bourbon and Catuai in higher altitudes considering shade slope. These results reports that the deemed harvesting years are similar in sensorial characteristics.

Finally, kappa estimates for pair of harvesting years 2011/12 – 2013/14 are approximately equal 1 when the wet processing type is unique covariate with significant effect in the models. Thus, those harvesting years are similar in sensorial attributes for genotypes Acaia and Yellow Bourbon.

The differentiated use of the kappa modeling proposed in this work helped in making interpretations about the sensory changes that occurred in the longitudinal samples. Our proposal, made possible the analysis in terms of sensorial qualities of the drink. The methodology helped to detect pairs of crops that showed sensory similarities between the samples and discriminated the variables that make the samples different.

Acknowledgments

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior - CAPES.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Akanda, M. A. S.; Khanam, M. et al. Goodness-of-fit tests for gee models using kappa-like statistic to diabetes mellitus study. *Journal of Applied Sciences*, **5** (9), 1597–1601 (2005).

[2] Avelino J.; Barboza, B.; Araya, J. C.; Fonseca, C.; Davrieux, F.; Guyout, B., & Cilas, C. Effects of slope exposure, altitude and yield on coffee quality in two altitude terroirs of costa rica, orosi and santa maría de dota. *Journal of the Science of Food and Agriculture*, **85** (11), 1869–1876 (2005).

[3] Borém, F. M. *Projeto protocolo de identidade, qualidade e rastreabilidade para embasamento da indicacao geografica dos cafes da mantiqueira*. [S. I.], (2007).

[4] Boré, F. M.; Cirillo, M. A.; De Carvalho Alves, A. P.; Dos 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*, **35** (2), e12552 (2020).

[5] Bor'em, F. M.; Luz, M. P. S.; Sáfadi, T.; Volpato, M. M. L.; Alves, H. M. R.; Borém, R. A. T.; & Maciel, D. A. Meteorological variables and sensorial quality of coffee in the mantiqueira region of minas gerais. *Coffee Science*, **14** (1), 38-47 (2019).

[6] Borém, F. M. and Shuler, J. Handbook of Coffee Post-harvest Technology: A Comprehensive Guide to the Processing, Drying, and Storage of Coffee. *Gin Press*. 282p. (2014)

[7] Cohen, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, **20** (1), 37–46 (1960).

[8] Decazy,F.; Avelino, J.; Guyot, B.; Perriot, J.-J.; Pineda, C.; & Cilas, C. Quality of different honduran coffees in relation to several environments. *Journal of food science*, **68** (7), 2356–2361 (2003).

[9] Donner, A. & Klar, N. The statistical analysis of kappa statistics in multiple samples. *Journal of clinical epidemiology*, **49** (9), 1053–1058 (1996).

[10] Donner, A.; Shoukri, M. M.; Klar, N.; and Bartfay, E. Testing the equality of two dependent kappa statistics. *Statistics in Medicine*, **19** (3), 373–387 (2000).

[11] Duarte, G. S., Pereira, A. A., and Farah, A. Chlorogenic acids and other relevant compounds in brazilian coffees processed by semi-dry and wet post-harvesting methods. Food Chemistry, **118** (3), 851–855 (2010).

[12] Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychological bulletin*, **76**, (5), 378-382 (1971).

[13] Giomo, G. and Borém, F. Cafés especiais no brasil: opção pela qualidade. *Informe Agropecuário*, Belo Horizonte, **32** (261), 7–16 (2011).

[14] Gonin, R.; Lipsitz, S. R.; Fitzmaurice, G. M.; and Molenberghs, G. Regression modelling of weighted κ by using generalized estimating equations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **49** (1), 1–18 (2000).

[15] Heagerty, P. J. and Zeger, S. L. Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, **91** (435), 1024–1036 (1996).

[16] Joet, T.; Laffargue, A.; Descroix, F.; Doulbeau, S.; Bertrand, B.; Dussert, S. Influence of environmental factors, wet processing and their interactions on the biochemical composition of green arabica coffee beans. *Food chemistry*, **118** (3), 693–701 (2010).

[17] Kerber, R. *Chimerge: Discretization of numeric attributes*. pages 123–128 (1992).

[18] Klar, N.; Lipsitz, S. R.; and Ibrahim, J. G. An estimating equations approach for modelling kappa. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, **42** (1), 45–58 (2000).

[19] Knopp, S.; Bytof, G.; and Selmar, D. (2006). Influence of processing on the content of sugars in green arabica coffee beans. *European Food Research and Technology*, **223** (2), 195-201 (2006).

[20] Kurgan, L. A. and Cios, K. J. Caim discretization algorithm. *IEEE transactions on Knowledge and Data Engineering*, **16** (2), 145–153 (2004).

[21] Liang, K.-Y; and Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika*, **73** (1), 13–22 (1986).

[22] Liang, K.-Y; and Zeger, S. L.; and Qaqish, B. Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **54** (1), 3–24 (1992).

[23] Lingle, T. R. (2011). *The coffee cupper's handbook: a systematic guide to the sensory evaluation of coffee's flavor*. Specialty Coffee Association of America Long Beach. 2011.

[24] Ma, Y.; Tang, W.; Feng, C.; and Tu, X. M. Inference for kappas for longitudinal study data:applications to sexual health research. *Biometrics*, **64** (3), 781–789 (2008).

[25] Scholz, M. B.; Kitzberger, C. S. G.; Prudencio, S. H., et al. The typicity of coffees from different terroirs determined by groups of physico-chemical and sensory variables and multiple factor analysis. *Food Research International*, **114**, 72–80 (2008).

[26] Tolessa, K.; D'heer, J.; Duchateau, L.; and Boeckx, P. Influence of growing altitude, shade and harvest period on quality and biochemical composition of ethiopian specialty coffee. *Journal of the Science of Food and Agriculture*, **97** (9), 2849–2857 (2017).

[27] Tsai, C.-J.; Lee, C.-I.; and Yang, W.-P. A discretization algorithm based on class-attribute contingency coefficient. *Information Sciences*, **178** (3), 714–731 (2008).

[28] Williamson, J. M.; Kim, K.; and Lipsity, S. R. Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, **90** (432), 1432–1437 (1995).

[29] Williamson, J. M.; Lipsity, S. R; and Manatunga, A. K. Modeling kappa for measuring dependent categorical agreement data. *Biostatistics*, **1** (2), 191–202 (2000).

[30] Zeger, S. L.; and Liang, K.-Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42** (1), 121–130 (1986).

Appendix

Presentation of data organization

Table 3. Illustrative table structure of data organization

Altitude	Slope	Processing	Raters	Crop Seasons	Genotype	Scores
1	1	0	1	1	1	78.0
1	.	.	.	1	1	79.0
⋮					⋮	
1	2	1	4		1	
2	1	0	1		2	
⋮	⋮	⋮	⋮		⋮	
2	2	1	4		2	
3	1	0	1		3	
⋮	⋮				⋮	
1	1			2	4	
⋮					⋮	
2		1			1	
⋮		⋮			⋮	
3	2	1	4	4	1	83.0

Altitude	$1 :\leq 1000m$; $2 : 1000m - 1200m$; $3 :\geq 1200m$
Slope	0 : shade; 1 : sum
Processing	0 : dry; 1 : wet
Raters	$1, 2, 3,$ and 4
Crop seasons	$1 : 2010/11$; $2 : 2011/12$; $3 : 2012/13$; $4 : 2013/14$
Genotype	1 : Acaia; 2 : Yellow Bourbon; 3 : Catuai; 4 : Mundo Novo
Scores	Range: $73 - 100$.