Tasters' performance in a coffee quality contest in Brazil

Aracy Camilla Tardin Pinheiro[1], Marcos Antônio Fabri Júnior[2], Cosme Damião Cruz[3],
José Luis dos Santos Rufino[4], Ney Sussumu Sakiyama[1]

[1]Universidade Federal de Viçosa/UFV, Departamento de Agronomia, Viçosa, MG, Brasil
[2]Empresa de Assistência Técnica e Extensão Rural/EMATER- MG, Escritório Regional de Lavras. Lavras, MG, Brasil
[3]Universidade Federal de Viçosa/UFV, Departamento de Biologia Geral, Viçosa, MG, Brasil
[4]Centro de Excelência dos Cafés das Matas de Minas, Viçosa, MG, Brasil
Contact authors: aracycamilla@hotmail.com; fabri.junior@emater.mg.gov.br; cdcruz@ufv.br; rufinojoseluis@gmail.com; sakiyama@ufv.br
Received in May 7, 2021 and approved in September 20, 2021

ABSTRACT

The objective of this work was to evaluate the performance of coffee tasters in five annual editions of Minas Gerais Coffee Quality Contest. The repeatability coefficients of the tasters' scores for sensory attributes were estimated, as well as the minimum numbers of tasters required for consistent sensory results, and the groups of tasters by (dis)similarity of sensory scores. For the repeatability analysis, the treatments (coffees) were tested with the repetitions, constituted by the tasters. The repeatability coefficients were estimated using the analysis of variance, principal component and structural analysis methods. The minimum number of tasters was obtained based on pre-established determination coefficients. Euclidean distance matrices between tasters were determined, which were used as a measure of dissimilarity for cluster analysis by the Tocher optimization method. The tasters' performance in five annual editions of Minas Gerais Coffee Quality Contest is reliable using COE or SCA sensory analysis protocols. Although not fully calibrated, most tasters are grouped with similar cupping results. Unless efficient calibration prior to the contest is adopted, the number of tasters to be used in the next contest editions can not be drastically and randomly reduced, since the estimated minimum number varied over the years. Calibration activities are suggested to improve two main aspects of the Minas Gerais Coffee Quality Contest: distinguishing the best coffees and trainning tasters.

Key words: Specialty coffee; Sensory analysis; Coffee beverage; Repeatability analysis.

1 INTRODUCTION

The commercial value of coffee is influenced by quality. The beverage quality is evaluated by professional tasters, who perform sensory analysis to score coffee quality attributes according to a chosen protocol (D'Alessandro, 2015; Gutierrez; Barrera, 2015; Pereira et al., 2017; Pereira et al., 2018).

Protocols, such as those established by Specialty Coffee Association - SCA (Specialty Coffee Association - SCA, 2015) and Cup of Excelence - COE (Alliance for Coffee Excelllence - ACE, 2020), are used for commercial, research, and contest purposes. Protocols and related guidelines and standards are also used for coffee samples preparation (e.g. roasting and grinding grade, sample amount, and water temperature and volume), tasting room preparation (e.g. without interfering smell, wind, and noise), and tasters capacity building.

Nevertheless, taster judgement may be biased by personal preferences, individual perception of quality, temporary low sensory perception, lack of calibration, prior information about the sample, undue interaction with other members, and other factors (Di Donfrancesco; Guzman; Chambers, 2014; Pereira et el., 2017).

The reliability of the sensorial analysis is associated with the homogeneity of scores among tasters and it seems to be more important than the number of tasters (Ferreira et al., 2018). Coffee quality contests are usually run through a group of tasters. The ideal number of tasters to evaluate coffee sensory quality is unknown. A small number of tasters may be inaccurate, while a large number may be more expensive without corresponding increase in accuracy. Then, the number may vary depending on the the the objectives of the contest and its feasibility.

For instance, the annual Minas Gerais Coffee Quality Contest objectives are to distinguish farmers who produce best coffees in the state, promote added value, publicize Minas Gerais coffees, but also to train tasters for sensory analysis of specialty coffees (Empresa de Assistência Técnica e Extensão Rural do Estado de Minas Gerais - EMATER, 2021). This contest is composed by large number of tasters, which allows one to statistically study its composition. Repeatability coefficient (Cruz; Regazzi; Carneiro, 2012) can be used to study the number of tasters and Tocher cluster analysis (Cruz; Ferreira; Pessoni, 2011) can be used to study the reliability of the coffee sensory panel.

Thus, the objective of this work was to evaluate the performance of coffee tasters in five annual editions of Minas Gerais Coffee Quality Contest.

2 MATERIAL AND METHODS

The work was carried out with data from the Minas Gerais Coffee Quality Contest, organized by the Secretary of State for Agriculture, Livestock and Supply (SEAPA) and the Technical Assistance and Rural Extension Company of the State of Minas Gerais (Emater-MG), by the Federal Institute

of Sul de Minas (IF Sul de Minas), by the Federal University of Lavras (UFLA) and by the Support, Teaching, Research and Extension Foundation (FAEPE), in the years 2013, 2014, 2015, 2016 and 2018. Data for the year 2017 were not available.

The contest received samples of Arabica coffee, from coffee growers in the municipalities of the state of Minas Gerais, produced by them in the current year, which fell into two categories: natural and peeled cherry. The natural coffee category (or dry method processing) is the form of preparation by which the freshly harvested coffee, after the washing/separation process, is taken to the terrace to dry in the sun and/or to the dryer, without removing the skin of the fruit.

The peeled cherry coffee category (or wet method processing), now called CD, includes the peeled cherry coffee, which refers to the preparation method in which the fruits are washed, pass through a peeler, separating the green fruits from the ripe ones, following, later, for drying. This category also includes pulped and/or demucilated cherry coffee, which is the form of preparation in which the fruits are washed, they pass through a peeler, in which the green and ripe fruits are separated, later they are taken to a fermentation tank or they pass through a demucilator, finally, they proceed to drying.

The sensory evaluation forms from the final phase of the Minas Gerais Coffee Quality Contest for the years 2013, 2014, 2015, 2016 and 2018 were used, this phase brought together the best coffees. The previous phases consisted of a physical classification step and a sensory classification step. The first phase, of physical classification, had an eliminatory character, type 2 coffees were classified, according to the Official Brazilian Coffee Classification Table, sieves 16 and above, with maximum leakage of 5% and humidity between 10% and 12%. In the next phase, the coffees were submitted to sensory analysis, in which samples with sensory analysis scores below 80 points were disqualified. The samples of the finalist coffees used in the study consisted of the coffees selected in the sensory analysis. The number of samples and the number of tasters varied in the years studied and can be found in Table 1, where you can also find the sensory protocol used, which will be described later.

Table 1: Number of samples, number of tasters and sensory protocol used in the years of contests under study.

Year	Number of samples	Number of tasters	Sensory Protocol
2013	40	12	COE
2014	39	12	COE
2015	28	11	COE
2016	19	19	SCA
2018	38	14	SCA

In the sensorial analysis, the characteristics related to the organoleptic patterns of the beverage were evaluated in each sample, through the "cupping test". Sensory analysis was performed by skilled tasters in specialty coffees (EMATER, 2021), all samples were coded, so that professionals had no information on the samples evaluated. Each taster performed a sample determination, each sample being made up of five cups.

In 2013, 2014 and 2015 the Cup of Excellence (COE) evaluation protocol (ACE, 2020) was used, in which the following attributes are evaluated: clean cup, sweetness, acidity, body, flavor, aftertaste, balance and overall. In this methodology, each sample has a starting score of 36 points, to which the scores for each attribute are added (0 to 8 points), making up the final score. In the years 2016 and 2018, the SCA (Specialty Coffee Association) protocol was used, in which scores from 0 to 10 points are attributed for the attributes fragrance/aroma, uniformity, clean cup, sweetness, flavor, acidity, aftertaste, body, balance and overall, the sum of the notes of all attributes constitutes the final score.

The experimental design used consisted of a DIC (completely randomized design). For the repeatability analysis, the treatments (coffees finalists in the contest) were tested with the repetitions, constituted by the tasters, in the five years of the contest. Classified data was used, which is the previous classification of the original data. From this classification, data were obtained and submitted to processing to calculate the repeatability coefficient. The repeatability coefficients (r) were estimated using the analysis of variance (ANOVA) methods, in which the temporary effect of the environment is removed from the error, by the principal component method (CP), based on the correlation matrices and the method of structural analysis (AE), based on the intraclass correlation matrices. The number of measurements, that is, tasters, necessary to predict the real value of individuals, based on the pre-established determination coefficients (R^2) (0.80, 0.85, 0.90 and 0.95), it was obtained according to the methodology described by Cruz, Regazzi and Carneiro (2012).

Euclidean distance matrices were determined between the tasters for the five years of the contest. In determining the matrices in the years in which the COE protocol was used, all attributes of the drink were used, in addition to the final score, however in the last years studied, 2016 and 2018, with the use of the SCA protocol, three attributes were excluded, the clean cup, sweetness and uniformity, due to the low variation between the tasters for these attributes. These distance matrices were used as a measure of dissimilarity for the cluster analysis of the tasters.

Tocher optimization method was used to study the similarity in the evaluation of the tasters, by forming groups of tasters more similar to each other in the evaluation of coffees. In this method, the set of tasters was divided, for each year of the contest, into non-empty and mutually exclusive groups. For that, it was adopted the criterion that the average of the dissimilarity measures, within each group, must be less than the average distances between any groups (Cruz; Ferreira; Pessoni, 2011).

To graphically visualize the differences in the sensory analysis of coffees between the groups of tasters, for each year studied, the sensory profiles of the groups were constructed, obtained by the Tocher optimization method. The notes of the sensory attributes were plotted on radar type diagrams with a single graphic scale, being constructed with the average of the notes of the sensory attributes for the groups of tasters. In the years with the use of the COE protocol, all eight attributes were used, but in the years in which the SCA protocol was used, the attributes clean cup, sweetness and uniformity were disregarded for the construction of the profiles. Sensory profiles were performed using Microsoft® Office Excel© software.

Statistical analyzes were performed in the Genes Software (Cruz, 2013).

3 RESULTS

Table 2 shows the repeatability coefficients for the first three years of studies, using the COE protocol, which ranged from 0.2268 to 0.5981. In 2013, the coefficients ranged from 0.2268 to 0.5207, this year the lowest coefficient was observed for the years studied with the COE protocol, for the final score; the highest coefficient occurred for the clean cup attribute. In 2014, the coefficients varied from 0.4032, for body to 0.5981 for balance. In 2015 the coefficients ranged from 0.2781 for the sweetness attribute to 0.5546 for the balance attribute.

Table 2: Estimates of repeatability coefficients and respective determination coefficients (in percentages in parentheses) of tasters for the sensory attributes of coffees evaluated in the contest of the years 2013, 2014 and 2015, using the COE protocol.

Attributes	Anova	AE	CP
	----------------------------------2013----------------------------------		
Clean cup	0.5008 (92.33)	0.5076 (92.52)	0.5207 (92.87)
Sweetness	0.4106 (89.31)	0.4174 (89.58)	0.4307 (90.08)
Acidity	0.4126 (89.40)	0.4195 (89.66)	0.4438 (90.54)
Body	0.4392 (90.38)	0.4461 (90.62)	0.4544 (90.90)
Flavor	0.2746 (81.96)	0.2808 (82.41)	0.2947 (83.37)
Aftertaste	0.3957 (88.71)	0.4026 (88.99)	0.4114 (89.34)
Balance	0.3487 (86.53)	0.3554 (86.87)	0.3681 (87.48)
Overall	0.4922 (92.08)	0.4990 (92.28)	0.5152 (92.73)
Final score	0.2268 (77.87)	0.2325 (78.43)	0.2494 (79.95)
	----------------------------------2014----------------------------------		
Clean cup	0.5214 (92.89)	0.5284 (93.08)	0.5380 (93.32)
Sweetness	0.4446 (90.57)	0.4517 (90.81)	0.4509 (90.79)
Acidity	0.4556 (90.94)	0.4627 (91.18)	0.4629 (91.18)
Body	0.4032 (89.02)	0.4102 (89.30)	0.4183 (89.61)
Flavor	0.4548 (90.92)	0.4619 (91.15)	0.4716 (91.46)
Aftertaste	0.4411 (90.45)	0.4482 (90.69)	0.4577 (91.01)
Balance	0.5782 (94.27)	0.5849 (94.41)	0.5981 (94.70)
Overall	0.5566 (93.78)	0.5635 (93.40)	0.5792 (94.29)
Final score	0.5405 (93.38)	0.5474 (93.55)	0.5644 (93.96)
	----------------------------------2015----------------------------------		
Clean cup	0.4679 (90.63)	0.4778 (90.96)	0.5324 (92.61)
Sweetness	0.2781 (80.90)	0.2871 (81.59)	0.3819 (87.17)
Acidity	0.4099 (88.43)	0.4199 (88.84)	0.4318 (89.32)
Body	0.4013 (88.05)	0.4112 (88.48)	0.4270 (89.13)
Flavor	0.3268 (84.22)	0.3363 (84.79)	0.4135 (88.58)
Aftertaste	0.4156 (88.68)	0.4259 (89.08)	0.5289 (92.51)
Balance	0.5208 (92.28)	0.5307 (92.56)	0.5546 (93.20)
Overall	0.4768 (90.93)	0.4868 (91.25)	0.5542 (93.20)
Final score	0.3196 (83.79)	0.3291 (84.37)	0.3449 (85.28)

AE: structural analysis; CP: principal component.

Table 3 shows the repeatability coefficients for the years 2016 and 2018, using the SCA protocol. In 2016, the coefficients ranged from 0.1987, for the body attribute to 0.9459 for the sweetness and uniformity attributes, in 2018 the variation was between 0.2825 and 0.9371, for the final score and the attribute sweetness, respectively. The repeatability coefficient for the body attribute, in the year 2016, obtained through the use of the variance analysis estimation methodology, was the lowest among all the years studied.

Table 4 shows the number of tasters needed for different determination coefficients, using the COE protocol.

In 2013, with 80% reliability, it took between four and fourteen tasters, depending on the attribute and the estimation methodology used (Table 4). In 2014, for 80% confidence, it took between three and six tasters for the different sensory attributes of coffee, in different estimation methods, for the same level of reliability. In 2015, this variation was between four and eleven tasters.

In order to estimate the final grade, in 2013, between thirteen and fourteen tasters were required. This was the only year studied in which the number of tasters needed to achieve 80% reliability exceeded the number of tasters used. This result demonstrated that the tasters used were less calibrated and disagreed in the score of the finalist coffees.

In 2014, for the 80% confidence level, four tasters were needed to determine the final grade. In 2015, considering this same requirement, between eight and nine tasters were required to determine the final grade of coffees.

Among the attributes of the beverage, the ones that required the least number of tasters for 2013 were clean cup, overall and body. In 2015, the clean cup and overall attributes are also among those that required a smaller number of tasters, which also occurred for the balance attribute. In 2014, as previously mentioned, the tasters presented greater uniformity in the evaluation, between three and six tasters were required in all attributes.

Table 3: Estimates of repeatability coefficients and respective determination coefficients (in percentages in parentheses) of tasters for the sensory attributes of coffees evaluated in the contest of the years 2016 and 2018, using the SCA protocol.

Attributes	Anova	AE	CP
	----------2016----------		
Fragance/Aroma	0.2479 (86.23)	0.2597 (86.95)	0.2789 (88.02)
Uniformity	0.9112 (99.49)	0.9155 (99.52)	0.9459 (99.70)
Clean Cup	0.9065 (99.46)	0.9110 (99.49)	0.9409 (99.67)
Sweetness	0.9112 (99.49)	0.9155 (99.52)	0.9459 (99.70)
Flavor	0.2885 (88.51)	0.3011 (89.11)	0.3124 (89.62)
Acidity	0.2680 (87.43)	0.2803 (88.10)	0.3193 (89.91)
Body	0.1987 (82.49)	0.2093 (83.41)	0.2259 (84.72)
Aftertaste	0.3002 (89.07)	0.3130 (89.64)	0.3278 (90.26)
Balance	0.2363 (85.44)	0.2479 (86.23)	0.3134 (89.66)
Overall	0.2006 (82.66)	0.2113 (83.58)	0.2547 (86.65)
Final Score	0.4176 (93.16)	0.4318 (93.52)	0.4344(93.59)
	----------2018----------		
Fragance/Aroma	0.3609 (88.77)	0.3679 (89.07)	0.4227 (91.11)
Uniformity	0.8581 (98.83)	0.8614 (98.86)	0.9065 (99.27)
Clean Cup	0.9089 (99.29)	0.9111 (99.31)	0.9247 (99.42)
Sweetness	0.9287 (99.45)	0.9305 (99.47)	0.9371 (99.52)
Flavor	0.3340 (87.53)	0.3408 (87.86)	0.3532 (88.43)
Acidity	0.2979 (85.59)	0.3044 (85.97)	0.3069 (86.11)
Body	0.3039 (85.94)	0.3105 (86.31)	0.3279 (87.23)
Aftertaste	0.3863 (89.81)	0.3933 (90.08)	0.4121 (90.75)
Balance	0.3260 (87.13)	0.3328 (87.47)	0.3656 (88.97)
Overall	0.4034 (90.44)	0.4105 (90.70)	0.4424 (91.74)
Final Score	0.2825 (84.65)	0.2890 (85.05)	0.3037 (85.93)

AE: structural analysis; CP: principal component.

Table 4: Estimate of the number of tasters needed to evaluate the coffees in the contest for the years 2013, 2014 and 2015, using the COE protocol, regarding the sensory attributes of clean cup (BEB), sweetness (DOC), acidity (ACI), body (COR), flavor (SAB), aftertaste (RET), balance (BAL), overall (GER) and final score (FIN), considering three different methods and coefficients of determination of 80%, 85%, 90% and 95%.

Method	R^2	BEB	DOC	ACI	COR	SAB	RET	BAL	GER	FIN
					2013					
Anova	0.80	3.88	5.58	5.54	4.97	10.25	5.94	7.26	4.02	13.20
	0.85	5.50	7.91	7.84	7.04	14.51	8.41	10.28	5.69	18.70
	0.90	8.73	12.56	12.46	11.17	23.05	13.36	16.32	9.04	29.71
	0.95	18.43	26.52	26.29	23.59	48.66	28.20	34.46	19.07	62.71
CP	0.80	3.68	5.29	5.01	4.80	9.57	5.72	6.87	3.76	12.04
	0.85	5.22	7.49	7.10	6.80	13.56	8.11	9.73	5.33	17.06
	0.90	8.29	11.90	11.28	10.81	21.54	12.88	15.45	8.47	27.09
	0.95	17.49	25.12	23.81	22.81	45.48	27.19	32.61	17.88	57.19
AE	0.80	3.99	5.74	5.69	5.11	10.57	6.11	7.47	4.13	13.64
	0.85	5.65	8.14	8.07	7.24	14.97	8.65	10.58	5.85	19.32
	0.90	8.97	12.92	12.81	11.49	23.77	13.74	16.81	9.29	30.68
	0.95	18.94	27.28	27.05	24.26	50.19	29.01	35.48	19.61	64.78
					2014					
Anova	0.80	3.57	4.86	4.65	5.75	4.66	4.93	2.84	3.10	3.31
	0.85	5.06	6.88	6.58	8.15	6.60	6.98	4.02	4.39	4.69
	0.90	8.03	10.92	10.45	12.94	10.08	11.08	6.39	6.97	7.44
	0.95	16.96	23.06	22.07	27.32	22.14	23.39	13.49	14.72	15.71
CP	0.80	3.44	4.87	4.64	5.56	4.48	4.74	2.69	2.91	3.09
	0.85	4.87	6.90	6.57	7.88	6.35	6.71	3.81	4.12	4.37
	0.90	7.73	10.96	10.44	12.52	10.08	10.66	6.05	6.54	6.95
	0.95	16.32	23.14	22.04	26.42	21.29	22.51	12.77	13.81	14.66
AE	0.80	3.67	5.00	4.78	5.92	4.80	5.07	2.92	3.19	3.40
	0.85	5.20	7.08	6.77	8.39	6.79	7.18	4.13	4.51	4.82
	0.90	8.26	11.24	10.76	13.32	10.79	11.40	6.57	7.17	7.65
	0.95	17.44	23.73	22.71	28.12	22.78	24.07	13.86	15.13	16.15
					2015					
Anova	0.80	4.37	9.93	5.53	5.73	7.89	5.39	3.54	4.22	8.15
	0.85	6.19	14.07	7.83	8.11	11.18	7.64	5.01	5.97	11.55
	0.90	9.84	22.34	12.43	12.89	17.76	12.13	7.96	9.49	18.35
	0.95	20.76	47.17	26.25	27.21	37.50	25.62	16.80	20.03	38.73
CP	0.80	3.51	6.47	5.26	5.37	5.67	3.56	3.21	3.22	7.60
	0.85	4.98	9.17	7.46	7.60	8.04	5.05	4.55	4.56	10.76
	0.90	7.91	14.57	11.84	12.07	12.77	8.02	7.23	7.24	17.09
	0.95	16.69	30.75	25.00	25.49	26.95	16.92	15.26	15.29	36.08
AE	0.80	4.55	10.39	5.76	5.97	8.24	5.62	3.68	4.39	8.51
	0.85	6.45	14.71	8.16	8.46	11.68	7.96	5.21	6.22	12.06
	0.90	10.24	23.37	12.95	13.43	18.54	12.64	8.28	9.88	19.16
	0.95	21.61	49.33	27.35	28.35	39.15	26.69	17.48	20.85	40.44

AE: structural analysis; CP: principal component.

Table 5 shows the number of tasters required, considering different levels of confidence, for the years 2016 and 2018, using the SCA protocol. In 2016, depending on the sensory attribute and the estimation method, one to seventeen tasters were required, with a reliability of 80%. In the same condition, in 2018, the variation was smaller, with a requirement between one and eleven tasters.

In 2016, for the final score, six tasters were needed, for 80% confidence. In 2018, considering the same level of reliability, it took between ten and eleven tasters to determine the final score.

In the evaluation of the other attributes of the beverage, only one taster was required for the evaluation of uniformity, clean cup and sweetness, in the two years with the use of the SCA protocol. For the others, the largest requirement in the number of tasters occurred for body and overall, in 2016, and for acidity and body, in 2018.

Euclidean distances were used as a measure of dissimilarity and Tocher optimization method for cluster analysis, which provided the formation of groups among the studied tasters, in the five years of study, as shown in Table 6. In the first three years of study, using the COE protocol, most tasters evaluated coffees in a similar way, with four groups being formed in 2013 and 2015, in 2014 three groups were formed. In these three years, the largest number of tasters were gathered in the same group, the others being formed by only one taster in each group.

Table 5: Estimate of the number of tasters needed to evaluate the coffees in the contest for the years 2016 and 2018, using the SCA protocol, regarding the sensory attributes aroma (ARO), uniformity (UNI), clean cup (BEB), sweetness (DOC), flavor (SAB), acidity (ACI), body (COR), aftertaste (RET), balance (BAL), overall (GER) and final score (FIN), considering three different methods and determination coefficients of 80%, 85%, 90% and 95%.

Method	R²	ARO	UNI	BEB	DOC	SAB	ACI	COR	RET	BAL	GER	FIN
						2016						
Anova	0.80	11.40	0.37	0.39	0.37	9.28	10.27	15.11	8.78	12.14	14.93	5.26
	0.85	16.16	0.52	0.55	0.52	13.15	14.55	21.41	12.44	17.19	21.16	7.46
	0.90	25.66	0.83	0.88	0.83	20.89	23.11	34.01	19.75	27.30	33.60	11.84
	0.95	54.17	1.75	1.86	1.75	44.10	48.78	71.79	41.70	57.64	70.93	25.00
CP	0.80	10.34	0.23	0.25	0.23	8.80	8.53	13.71	8.20	8.76	11.71	5.21
	0.85	14.65	0.32	0.36	0.32	12.47	12.08	19.42	11.62	12.42	16.58	7.38
	0.90	23.27	0.52	0.57	0.52	19.81	19.19	30.84	18.46	19.72	26.34	11.72
	0.95	49.12	1.09	1.19	1.09	41.81	40.51	65.10	38.97	41.63	55.61	24.74
AE	0.80	12.14	0.39	0.41	0.39	9.87	10.92	16.13	9.33	12.92	15.94	5.58
	0.85	17.20	0.55	0.59	0.55	13.98	15.47	22.85	13.21	18.31	22.58	7.90
	0.90	27.31	0.88	0.93	0.88	22.20	24.58	36.30	20.98	29.08	35.86	12.55
	0.95	57.66	1.85	1.96	1.85	46.87	51.88	76.62	44.30	61.39	75.70	26.50
						2018						
Anova	0.80	6.87	0.64	0.39	0.30	7.74	9.14	8.88	6.17	8.02	5.74	9.84
	0.85	9.74	0.91	0.55	0.42	10.96	12.95	12.58	8.74	11.36	8.14	13.94
	0.90	15.46	1.45	0.88	0.67	17.41	20.56	19.98	13.88	18.05	12.92	22.15
	0.95	32.65	3.06	1.85	1.42	36.75	43.41	42.19	29.31	38.10	27.28	46.75
CP	0.80	5.46	0.41	0.33	0.27	7.33	9.03	8.20	5.71	6.94	5.04	9.17
	0.85	7.74	0.58	0.46	0.38	10.38	12.80	11.62	8.08	9.83	7.14	12.99
	0.90	12.29	0.93	0.73	0.60	16.48	20.32	18.45	12.84	15.62	11.34	20.63
	0.95	25.95	1.96	1.55	1.27	34.80	42.90	38.95	27.10	32.97	23.95	43.56
AE	0.80	7.08	0.66	0.40	0.31	7.98	9.43	9.16	6.36	8.27	5.92	10.16
	0.85	10.03	0.94	0.57	0.44	11.30	13.36	12.98	9.00	11.71	8.38	14.39
	0.90	15.94	1.49	0.90	0.69	17.95	21.21	20.61	14.30	18.61	13.31	22.85
	0.95	33.64	3.14	1.90	1.46	37.89	44.78	43.51	30.19	39.28	28.10	48.25

AE: structural analysis; CP: principal component.

Table 6: Grouping, by Tocher optimization method, of the tasters responsible for the sensory evaluation of coffees participating in the contest in the five years of study, based on Euclidean distance.

Year	Group 1	Group 2	Group 3	Group 4
2013	1. 2. 5. 6. 7. 8. 9. 10. 12	3	4	11
2014	2. 3. 4. 5. 6. 7. 9. 10. 11. 12	1	8	-
2015	2. 3. 4. 5. 6. 7. 8. 9	1	10	11
2016	1. 2. 3. 4. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15.16. 17. 18. 19	5	-	-
2018	3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 14	1	2	13

Using the SCA protocol, two groups were formed in 2016 and four groups in 2018 (Table 6). In 2016, eighteen of the nineteen tasters evaluated the coffee in a similar way, staying in the same group, only the taster number 5 was different in the evaluation of the coffees. In 2018, the tasters also showed similarity in the assessment, eleven tasters were gathered in the same group, there were only three separate tasters in this group, each of them formed a different group.

The sensory profiles of the groups of tasters formed in the first three years of study, using the COE protocol, through the Tocher grouping, can be seen in Figure 1. These profiles allow to graphically visualize the main differences in the evaluation between the groups of tasters, for all the attributes of the beverage.

In 2013, group 1 presented a more balanced sensory profile among the scores of the evaluated attributes, being the result of the average evaluation of nine of the twelve studied tasters (Figure 1A). Group 2, which was composed by the taster 3, presented a less balanced sensory profile, with bigger differences in the evaluation of the body attribute, in which it provided, on average, the lowest scores among all groups. In the evaluation of group 3, composed of the taster 4, the sensory profile stood out for the highest average score for the overall attribute. Group 4, which represents the taster 11, presented a profile with higher average scores for the attributes body, acidity and sweetness.

In 2014, group 1, formed by ten of twelve tasters, more convergent in the evaluation, and group 3, composed of taster 8, presented sensory profiles balanced between the attributes, with greater differences between them for the evaluation of clean cup and flavor, in which the taster in group 3 assigned lower grades (Figure 1B). The sensory profile of group 2, composed of the taster 1, was less balanced when compared to the others, with the lowest scores for the evaluation of the attributes clean cup, sweetness, acidity, body and balance and prominence, with the highest score, for the aftertaste.
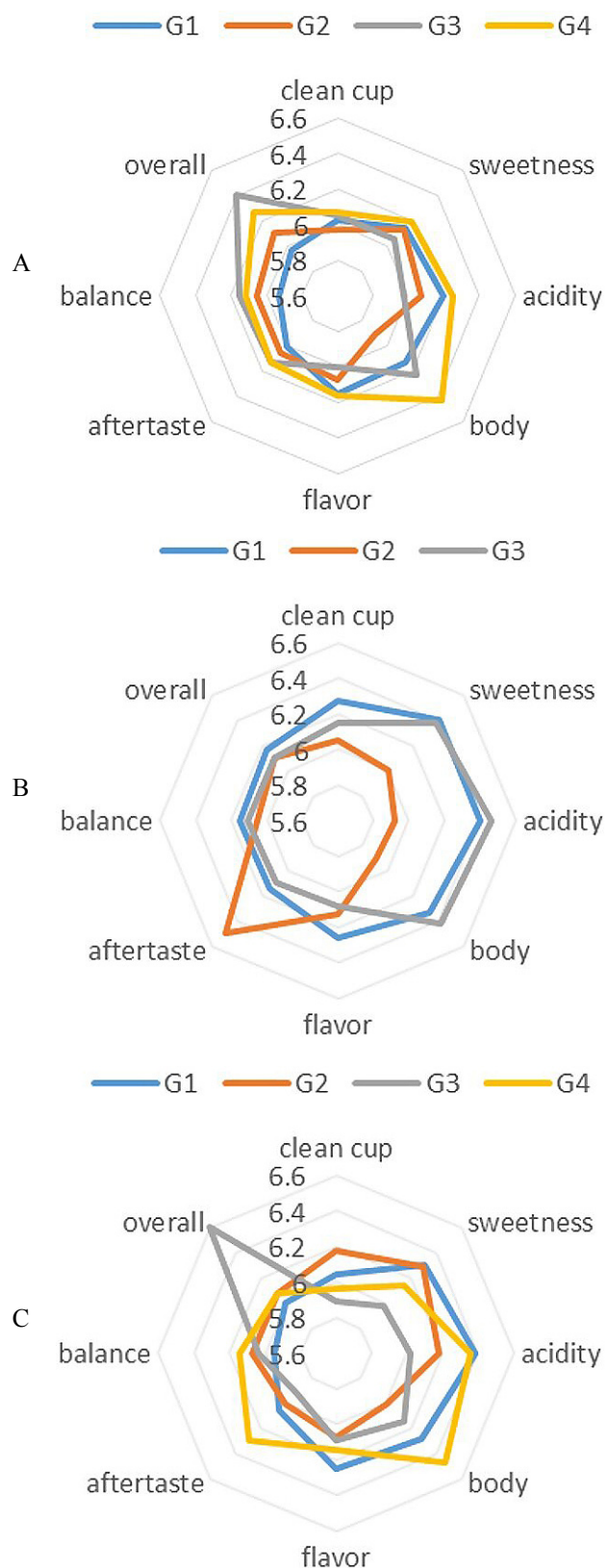

Figure 1: Sensory profile of the groups of tasters, formed through the Tocher optimization method, in the Minas Gerais Coffee Quality Contest, using the COE protocol, in the years: 2013 (A), 2014 (B) and 2015 (C). G1: group 1; G2: group 2; G3: group 3; G4: group 4.

In 2015, group 1, formed by the 8 most converging tasters in the sensory evaluation of coffee, also presented a balanced profile, although there was an emphasis on the highest notes for the attributes sweetness and acidity and the highest score for flavor, in relation to the other groups (Figure 1C). Groups 2 (taster 1), 3 (taster 10) and 4 (taster 11) had less balanced profiles. In the profile of group 2 there was the highest score in the evaluation of the clean cup and the lowest score for the body attribute, in comparison with the others. Group 3 evaluated the attributes clean cup, sweetness and acidity with the lowest average scores, however, it presented the highest score for the overall. The profile of group 4 stood out for presenting the highest average scores for the balance, aftertaste and body attributes.

Figure 2 shows the sensory profiles of the assessments made by the groups of tasters in the years 2016 and 2018, formed using Tocher optimization grouping, using the SCA protocol.
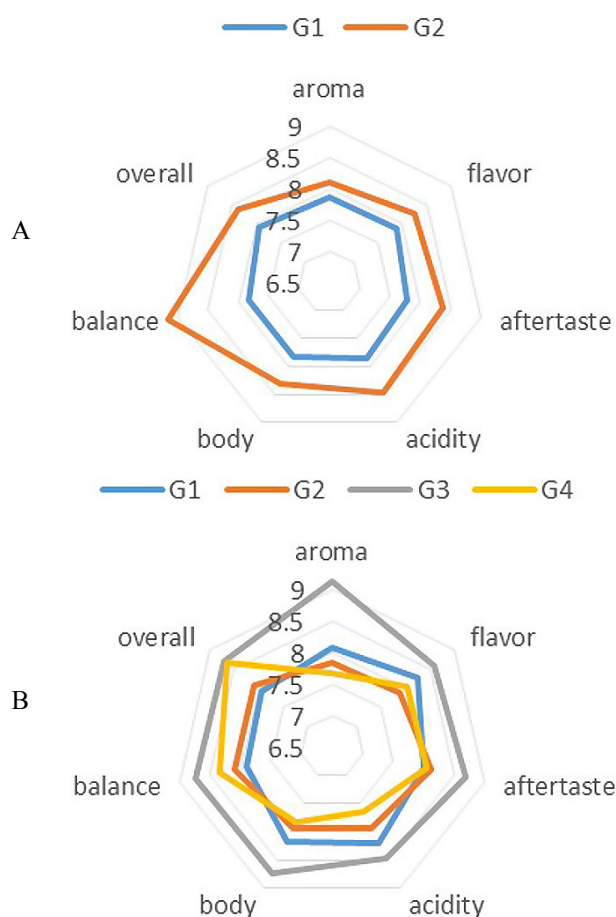

Figure 2: Sensory profile of groups of tasters, formed through the Tocher optimization method, in the Minas Gerais Coffee Quality Contest, using the SCA protocol, in the years: 2016 (A) and 2018 (B). G1: group 1; G2: group 2; G3: group 3; G4: group 4.

In the year 2016, group 1 presented a balanced sensory profile, as well as the profile of group 2, however, there was in the evaluation of the taster 5, which represents group 2, greater attribution of scores for the balance attribute, when compared with the other attributes (Figure 2A). Group 2 assigned higher average scores for all attributes compared to group 1.

In 2018 the profiles of group 1, formed by eleven tasters more convergent in the sensory evaluation of coffee, and group 2, constituted by taster 1, presented the most balanced profiles between the notes of the attributes, the biggest differences in the evaluation of both were the highest scores in group 1 for the attributes aroma, acidity, body and flavor and the lowest scores for balance and overall. Group 3, formed only by the taster 2, presented the highest scores for all the attributes of the beverage, with emphasis on the aroma score. Group 4, composed of the taster 13, presented a sensory profile with some similarities with group 2, for the aftertaste, flavor, aroma and body scores, however, the highest scores for the overall and balance attributes differed these two groups.

4 DISCUSSION

The highest repeatability coefficients in this study were obtained using the SCA protocol, for the attributes sweetness, uniformity and clean cup (Table 3). This fact occurred due to the way these attributes are evaluated, differently from the other seven attributes (fragrance/aroma, flavor, acidity, body, balance, aftertaste and overall), which are subjectively evaluated, according to their quality, for these three the taster makes an objective assessment, granting 2 points per cup that is normal for these characteristics. Thus, there is less variation between the tasters in the scores of these attributes.

The lowest coefficients with the use of the SCA protocol were obtained for the body attribute, in the year 2016 and for the final score, in the year 2018 (Table 3). The lowest coefficients of repeatability with the COE protocol were obtained for the final score in 2013 and for sweetness in 2015 (Table 2).

In the year 2014, with the use of twelve tasters, all the repeatability coefficients reached values above 0.4, which may indicate that the tasters were more trained or calibrated in the evaluation of coffee, since the same number of tasters was employed in the previous year (2013), with different results.

Regarding the estimation methodology, the highest repeatability coefficients, for all attributes and the final grade, were obtained with the principal component methodology. Other authors obtained similar results when studying the repeatability for characteristics in other species (Bergo et al., 2013; Negreiros et al., 2014). The variance analysis methodology provided the lowest repeatability coefficients in all the years studied (Tables 2 and 3), corroborating with results obtained by Negreiros et al. (2014).

Bergo et al. (2013) considered values above 0.4 for the repeatability coefficient to be reliable, however, in this study,

coefficients of lower magnitudes were observed in four of the five years analyzed for some attributes of the coffee and even for its final score (Tables 2 and 3). Despite obtaining repeatability coefficients considered low, the determination coefficients obtained were higher than 80% for all coffee attributes in all years studied, except for the final score in 2013, which ranged from 77.87 to 79.95%, depending on the estimation method used.

In general, the determination coefficient ranged from 77.87 to 94.7% in the years studied. These results demonstrate that there is good reliability in the number of tasters used to express the real sensory quality of coffee in the studied contest, using both the COE and SCA protocol. In addition, for Negreiros et al. (2014), the definition of the ideal determination coefficient should privilege, in addition to the minimum reliability expected in the data, the availability of resources and labor for evaluations. As there are no other reference studies to assess the minimum degree of accuracy requirement for the number of tasters, 80% was considered to be a good level in this study.

The tasters were less uniform and, consequently, there was a need for a higher number of tasters for the attributes flavor and balance, in 2013 and, sweetness and flavor, in 2015. Flavor is a complex attribute, defined as a mixed experience of olfactory, taste and tactile sensations perceived during tasting (Carvalho et al., 2016; Teixeira, 2009). The complexity linked to the definition of this attribute may be the reason for the bigger variation in the opinion of the tasters in the attribution of grades.

Regarding the number of tasters, in general, it is possible to notice a large variation in the number needed between the sensory attributes of coffee, regardless of the methodology used (Tables 4 and 5). This variation may occur due to the greater or lesser capacity of the tasters in the evaluation of the different attributes, in addition to their level of experience. For Ferreira et al. (2018), who observed the existence of variability in the grades attributed by three different tasters in different attributes of the drink, before carrying out the analysis of the attributes for any technical purpose, it is indicated that the descriptive analysis of the data is performed to verify the need for elimination, of the scores of at least one taster in at least one attribute. According to the authors, there would be no need to always be considered the same taster for the different attributes of the drink.

There was also a wide variation in the number of tasters needed between the years studied (Tables 4 and 5). In 2014, 2015, 2016 and 2018, it would be possible to reduce the number of tasters in the evaluation of the contest's coffees, without loss of reliability, considering the level of 80%. In 2013, however, for the same level of reliability, between 13 and 14 tasters were needed in the final grade evaluation, which is higher than the original number, 12.

There was an increase in the maximum number of tasters required, that is, a higher number required among all attributes in the same year of study, with an increase in the number of tasters used, for the same level of reliability, of 80%. In 2016, when the largest number of tasters was studied, nineteen, seventeen tasters were needed, for 80% confidence. In 2013 and 2014, a maximum of fourteen and six tasters were required, respectively, studying twelve tasters. In 2015, with eleven tasters evaluated, the maximum number required for 80% confidence was eleven. In the last year evaluated, 2018, with fourteen tasters, eleven tasters were needed. This maximum number is also the highest among the estimation methodologies used, with variation between them.

Other studies have pointed out the need for different numbers of tasters from those found in this study. Pereira et al. (2018) indicated that the use of 6 tasters is efficient to perform the sensory analysis, for that purpose 10 tasters were used, which evaluated 20 samples of Arabica coffee, with a minimum score of 80 points, following the SCA and BSCA protocol for Arabica group coffees. According to these authors, the modeling applied in this study allows to conclude that it is possible to recommend the minimum number of tasters, for these conditions, however, they stated that the approach is limited to the study and thus, suggest that it is necessary to use 6 or more tasters in scientific studies and in routine tests for marketing purposes.

Without considering an ideal number of tasters for sensory analysis of coffee, other studies used a smaller number of tasters. Silveira et al. (2015) used three tasters to assess the sensory quality of coffee at different altitudes, faces of sun exposure and fruit color. Ribeiro et al. (2016) used four trained and qualified professional tasters, certified as judges of special coffees, to study the association between chemical descriptors of the bean with the sensory quality of the coffee drink, on expressions resulting from the interactions of the genotype, environment and coffee processing. Tolessa et al. (2016) used three professional tasters to analyze the interactive effects of altitude, shading levels and harvest periods on Ethiopia's coffee quality.

In the sensory analysis of coffee, according to Di Donfrancesco, Guzman and Chambers (2014), the use of expert tasters presents some problems, such as the influence of external factors and the change in an individual's perceptual skills through illness and other factors. For Pereira et al. (2018), the taster tends to prefer a sensory profile over others or even according to commercial and industrial standards, in order to meet the demand of certain customers. These factors may affect the evaluation of the tasters.

In addition, there are other aspects that may interfere with the performance of tasters in sensory analyzes, which were not evaluated in this study, such as, for example, the interaction between them during the analyzes. Pereira et al. (2017) observed distortions in the performance of coffee tasters

when there is interaction between them, they concluded that the existence of conversations, comments and noise interfered in the sensorial analysis, decreasing the quality of the tasters' judgment on the levels of the attributes of the coffee beverage, probably due to their lack of concentration.

Considering that there are factors that can affect the performance of tasters in the sensory analysis, it is important to assess the consistency or similarity in the assessment of the tasters, not just the number of these professionals, for a more reliable assessment. Chambers, Bowers and Dayton (1981) emphasized that, in addition to the minimum number of tasters, it is necessary to study the consistency of who is carrying out the analysis. For these authors, a panel of three members, trained and experienced, presented a lower residual mean square than a semi-trained panel, composed of eight members, for a study of sensory analysis of birds, which demonstrates, according to them, that the consistency between the tasters must be respected and observed.

Ferreira et al. (2018) evaluated whether three trained tasters could constitute the minimum number of tasters to ensure the credibility of the sensory analysis of coffee. These authors concluded that, regardless of the number of tasters used, the reliability of the scores is related to the variability, and the lower the variability of the scores in the same situation studied for coffees, the greater the reliability on them. According to the authors, the reliability is directly related to the training and the technical capacity of the taster, less related to the number of tasters and more to the homogeneity of the scores attributed by each taster for the same evaluation conditions.

In this study, the convergence or similarity in the assessment observed for most of the studied tasters, in all years of study (Table 6), can be explained due to the fact that only skilled tasters were used. For Pereira et al. (2018) the use of professionals such as Q-graders and R-graders (professionals of tasting and classification of coffees that receive a world certification linked to the Coffee Quality Institute - CQI, for arabica and robusta coffee, respectively) is necessary, since these professionals are previously trained to carry out such activities, the authors also affirm that it is necessary to demand more veracity from the sensory analysis of coffee.

In the cluster analysis between the tasters, it was possible to notice that the majority of the tasters evaluated, in the five years of study, were grouped, which indicates that they evaluated the coffees in a similar way. However, our results also indicated that the number of tasters in the contest can not be drastically and randomly reduced, since the estimated minimum number varied over the years. Improving the calibration between tasters would probably improve the consistency of tasters over the years, allowing to establish a fixed small number of tasters for the contest. Calibration is usually achieved through independent training courses for specific cupping protocols or though the adoption of a calibration step of the tasters prior to the contest.

5 CONCLUSIONS

The tasters' performance in five annual editions of Minas Gerais Coffee Quality Contest is reliable using COE or SCA sensory analysis protocols.

Although not fully calibrated, most tasters are grouped with similar cupping results.

Unless efficient calibration prior to the contest is adopted, the number of tasters to be used in the next contest editions can not be drastically and randomly reduced, since the estimated minimum number varied over the years.

Calibration activities are suggested to improve two main aspects of the Minas Gerais Coffee Quality Contest: distinguishing the best coffees and trainning tasters.

6 ACKNOWLEDGMENTS

7 AUTHORS' CONTRIBUTION

ACTP systematized and analyzed the data, interpreted the results and wrote the manuscript, MAF Jr provided the data and reviewed the results, CDC supervised all statistical analyzes and reviewed the results, JLSR reviewed the results and the manuscript and NSS guided, reviewed and approved the final version of the work. All authors discussed the results and contributed to the final manuscript.

8 REFERENCES

ALLIANCE FOR COFFEE EXCELLLENCE - ACE. **COE Cupping Form**. 2020. Available in: < https://allianceforcoffeeexcellence.org/wp-content/uploads/2017/11/cupping_form_new-3.pdf>. Access in: October, 22, 2021.

BERGO, C. L. et al. Estimativa de repetibilidade de caracteres de produção em pupunheiras para palmito da raça Putumayo. **Revista Brasileira de Fruticultura**, 35(3):829-836, 2013.

CARVALHO, A. M. et al. Relationship between the sensory attributes and the quality of coffee in different environments. **African Journal of Agricultural Research**, 11(38):3607-3614, 2016.

CHAMBERS, E.; BOWERS, J. A.; DAYTON, A. D. Statistical designs and panel training/experience for sensory analysis. **Journal of Food Science**, 46(6):1902-1906, 1981.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. **Modelos biométricos aplicados ao melhoramento genético**. 2.ed. Viçosa, Minas Gerais: Editora UFV, 2012. 514p.

CRUZ, C. D. GENES: A software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, 35(3):271-276, 2013.

CRUZ, C. D.; FERREIRA, F. M.; PESSONI, L. A. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 2011. 620p.

D'ALESSANDRO, S. C. Identificação de cafés especiais. In: SAKIYAMA, N. S. et al. **Café arábica**: Do plantio à colheita. Viçosa, Minas Gerais: Editora UFV, p.268-291, 2015.

DI DONFRANCESCO, B. D.; GUZMAN, N. G.; CHAMBERS, E. Comparison of results from cupping and descriptive sensory analysis of Colombian brewed coffee. **Journal of Sensory Studies**, 29(4):301-311, 2014.

EMPRESA DE ASSISTÊNCIA TÉCNICA E EXTENSÃO RURAL DO ESTADO DE MINAS GERAIS - EMATER. **XVIII Concurso de qualidade dos cafés de Minas Gerais**. 2021. Available in: < https://www.emater.mg.gov.br/doc/intranet/upload/cafe2018/regulamento_do_concurso_dos_cafes.pdf >. Access in: October, 22, 2021.

FERREIRA, W. M. P. et al. Requisitos para credibilidade da análise sensorial do café. **Revista de Ciências Agrárias**, 41(1):271-280, 2018.

GUTIERREZ, N.; BARRERA, O. Selección y entrenamiento de un panel en análisis sensorial de café *Coffea arabica* L. **Revista de Ciencias Agrícolas**, 32(2):77-87, 2015.

NEGREIROS, J. R. S. et al. Estimativa de repetibilidade para caracteres de qualidade de frutos de laranjeira-doce. **Pesquisa Agropecuária Brasileira**, 49(1):40-48, 2014.

PEREIRA, L. L. et al. The consistency in the sensory analysis of coffees using Q-graders. **European Food Research and Technology**, 243(9):1545-1554, 2017.

PEREIRA, L. L. et al. Propositions on the optimal number of q-graders and r-graders. **Journal of Food Quality**, 3285452:7, 2018.

RIBEIRO, D. E. et al. Interaction of genotype, environment and processing in the chemical composition expression and sensorial quality of Arabica coffee. **African Journal of Agricultural Research**, 11(27):2412-2422, 2016.

SPECIALTY COFFEE ASSOCIATION - SCA. **Protocols**. 2015. Available in: < http://www.scaa.org/PDF/resources/cupping-protocols.pdf>. Access in: October, 22, 2021.

SILVEIRA, A. S. et al. Sensory analysis of specialty coffee from different environmental conditions in the region of Matas de Minas, Minas Gerais, Brazil. **Revista Ceres**, 63(4):436-443, 2016.

TEIXEIRA, L. V. Análise sensorial na indústria de alimentos. **Revista do Instituto de Laticínios Cândido Torres**, 64(366):12-21, 2009.

TOLESSA, K. et al. 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, 2016.