On characterization of sensory data in presence of missing values: The case of sensory coffee quality assessment

Sobre la caracterización de datos sensoriales en presencia de valores perdidos: El caso de la evaluación sensorial calidad del café

Andrés F. Ochoa-Muñoz[1*] Jefferson A. Peña-Torres[2] Cristian E. García-Bermúdez[1]
Kevin F. Mosquera-Muñoz[3] Jeison Mesa-Diez[4]

Recibido 29 de septiembre de 2021, aceptado 17 de agosto de 2022
Received: September 29, 2021 Accepted: August 17, 2022

ABSTRACT

Multiple factor analysis was used to examine organoleptic coffee assessments such as aroma, aftertaste, flavor, acidity, balance, body, uniformity, sweetness, clean cup, and other organoleptic-related properties used in Coffee Quality Assessment. The Sensory analysis was performed using missing values (NA) scenarios with 5%, 10%, 20%, and 30% of NA. The results suggest that RI-MFA is robust to NA presence of and appears to be appropriate when sensory data are present. Simulation scenarios deleting or replacing values from real-world datasets could be a good strategy; different domains, samples, types of variables, and distributions could prove much closer to reality.

Keywords: Statistical multivariate analyses, missing values, multiple factorial analysis, sensorial data.

RESUMEN

Se utilizó un análisis factorial múltiple para examinar las evaluaciones organolépticas del café como aroma, sabor residual, sabor, acidez, balance, cuerpo, uniformidad, dulzura, limpieza de taza, entre otras propiedades organolépticas. Utilizando escenarios de valores faltantes (NA) con 5%, 10%, 20% y 30% de NA, se realizó el análisis sensorial. Los resultados sugieren que RI-MFA es robusto a la presencia de NA y parece ser apropiado cuando están presentes los datos sensoriales. Los escenarios de simulación que eliminan o reemplazan valores de conjuntos de datos del mundo real pueden ser una buena estrategia; diferentes dominios, muestras, tipos de variables y distribuciones podrían estar mucho más cerca de la realidad.

Palabras clave: *Análisis estadístico multivariante, valores faltantes, análisis factorial múltiple, datos sensoriales.*

[1] Universidad de Valparaíso. Instituto de Estadística. Valparaíso, Chile.
E-mail: andres.ochoa@postgrado.uv.cl; cristian.garcia@postgrado.uv.cl
[2] Universidad del Valle. Escuela de Ingeniería de Sistemas y Computación. Cali, Colombia.
E-mail:jefferson.amado.pena@correounivalle.edu.co
[3] Universidad del Valle. Departamento de Economía. Cali, Colombia.
E-mail: kevin.mosquera@correounivalle.edu.co
[4] Universidad del Valle. Escuela de Estadística. Cali, Colombia.
E-mail: jeison.mesa@correounivalle.edu.co
* Autor de correspondencia: andres.ochoa@postgrado.uv.cl

INTRODUCTION

Coffee is an important drink, different in every way and with different sensory characteristics. Consumer demand for products by quality characteristics is also growing [1]. Therefore, it is vital to produce high-quality and stable coffee that satisfies consumer demand and preferences [2]. Differences determined by the growing region's specific environmental conditions, temperature, altitude, latitude, and humidity, directly influence the grains [3]. Although several species of coffee are known today, there is a particular interest in Robusta and Arabica coffee. Both are hardy crops; these particular coffee plant species are resistant to disease, insects, and weather. Moreover, these species have economic and cultural importance in several countries in the world [4-6]. With a strong and long tradition in Colombia, coffee exportation is a relevant commercial activity [7].

Organoleptic quality is one of the most important characteristics of the successful marketing of coffee. Nowadays, a common way to evaluate coffee quality is through trained testers in a sensory panel [8, 9]. Organoleptic quality is one of the most essential characteristics of successful marketing of coffee. Nowadays, a common way to evaluate coffee quality is through trained testers in the sensory panels [8], [9]. The sensory panel typically evaluates sensory characteristics such as aroma, flavor, and natural and chemical factors, important for the consumer of special and regular coffees [8]. Following the criteria of the Speciality Coffee Association (SCA) [10, 11]. The sensory characteristics such as aroma, flavor, and natural and chemical factors are important for the consumer of special and regular coffees [8]. The sensory assessment of coffee characteristics is important because that could be related to consumers' acceptance and purchase [12, 13]. Moreover, given the importance of coffee, several computer and statistical techniques can be used in sensory data analysis [14, 15]. In this context, a judge or an electronic device [16] describes all sensations perceived and sets a quantitative or qualitative evaluation of the coffee beverage characteristics [17-21].

Multiple Factor Analysis (MFA) [22] is one of the most popular techniques to study multiple variables or factors such as those evaluated in sensory studies. It is an alternative both to Principal Component Analysis (PCA) and Simple and Multiple Correspondence Analysis (SCA, MCA) that permits synthesizing, representing, and interpreting relationships between several sets of features [23]. This technique has visualization tools and mathematical indices that can be helpful for coffee quality researchers. MFA has been widely studied as part of multivariate statistics and can analyze quantitative and qualitative variables grouped by type and interest within the study. MFA is a standard statistical technique that works with all available data. In particular, MFA has been used in sensorial data studies on wine quality, orange juices, coffee variety, and surveys-based studies [24-27]. Even though there are several strategies to avoid or treat them, it is common for this type of data to suffer losses or contain errors that prevent analysis. A variable may only have a small number of missing responses, but in combination, all datasets could have missing values and could be a problem in the study [28].

The presence of Missing Values or data not available (NA) is a problem that has been approached from different perspectives, among which is the Regularized Iterative method (RI-MFA) which uses the mean if the factor is quantitative or with the most acceptable value according to the proportions if it is qualitative [29]. RI-MFA is considered the best strategy when data contains missing values, but evaluating the performance is vital for proper usage of the techniques dealing with missing values [30, 31]. the performance evaluation should incorporate controlled simulations with incomplete data to better understand the RI-MFA technique. Imputation and missing data generation techniques are underdeveloped and improved, and scenario simulation can uncover issues that may arise when missingness is induced on complete data.

Given the above, this study aims to apply the methods of MFA and RI-MFA to analyze the sweetness, flavor, bitterness, fragrance or aroma, saltness, body, acidity, mouthfeel, aftertaste, and cup balance of a sensory coffee assessment. Although these properties are not found in the SCA protocol, organizations such as the Coffee Quality Institute (CQI), which focuses on improving coffee quality and the producers0 lives, include other standards, protocols, and variables. The dataset included the evaluation of coffee from 1341 coffee-producing units around the world located in 36 countries and was obtained using the web scraping technique from Coffee Quality Institute in 2018. Coherent approximations have been obtained with MFA and RI-MFA multivariate methods on the coffee sensory dataset when considering samples with desirable and probable scenarios of 5%, 10%, 20%, and 30% missing values.

MATERIAL AND METHODS

The dataset was collected and downloaded from Coffee Quality Institute (CQI) in 2018 using a web scraping custom program to study the quality of coffee. The trained and accredited CQI panelist indicates which coffee is common or special according to sensory particularities. In the dataset, the organoleptic coffee assesses aroma, aftertaste, flavor, acidity, balance, body, uniformity, sweetness, clean cup, copper points, and others related to coffee data, (see Table 1).

The worldwide coffee producers were considered individuals of interest in this study. The data collected contains information about two species of coffee Arabica and Robusta, produced in 37 countries worldwide. A detailed description of the data and descriptive statistics of the sample scores were obtained. A radar graph of the attributes for presenting the available information was also obtained.

Initially, all dataset values were considered. MFA is used to explore multiple data tables of the same set of observations. In the MFA context, the dataset is a table {X, ..., Xn} and each column is an observed individual feature, which has been recorded in a row. MFA number and feature type can vary from another matrix with the same individual. The main table is divided into groups (subtables) according to the nature of the variables. Commonly, a large number of variables are available whose relations are of interest to the researcher.

MFA analyzes j data tables, K1, ...Kj, where Ki contains a dataset portion with quantitative or qualitative observations on the same individuals. MFA takes root in PCA or MCA, in which weights are assigned to variables. MFA follows three stages:

- **Partial analyses**: Each Ki is associated with a cloud of individuals and is processed with Principal Component Analysis (PCA) if quantitative or Multiple Correspondence Analysis (MCA) if qualitative. In this stage, MFA compares the groups through the individuals, group typologies in a common space, and identifies particular individuals in the group.
- **Ponderation**: In this stage, MFA balances the Ki influence. The matrix of variance-covariance associated with each Ki is decomposed in the previous stage, and its largest eigenvalues $\lambda i1$ are derived. Each variable belonging to Ki is weighted by $1/\sqrt{\lambda i1}$
- **PCA stage**: MFA is considered a particular weighted PCA. In this stage, PCA is performed on the merged and weighted data table K that contains all Ki to obtain the principal components **F**.

After analysis of variance, groups, and inertia, the study scenarios were created. The RI-MFA technique is used to deal with multiple tables containing different missing rates. After analysis of variance, groups, and inertia, the study scenarios were created. The RI-MFA technique is used to deal with multiple tables containing different missing rates, in order to obtain the best possible estimation for the NA. RI-MFA technique is carried out by an iterative execution of six stages:

- **Imputation:** NA is imputed with the mean by variable in the quantitative Ki. The technique

Table 1. Organoleptic and non-organoleptic Coffee attributes.

Group	Attribute (Code)	Description
Source	Specie (SPC) Country (CRT)	Information of specific genotype and their origin
Flavor	Fragrance/aroma (FRA) Flavor (FLV). aftertaste (AFT) acidity (ACD) mouthfeel (MOF) balance (EQ).	Recorded by a trained judge. According to feeling appreciation during the tasting.
Cup presentation	Uniform cup (UCP) Clean cup (CCP) Bittersweet (BTT).	Recorded by a trained judge. According to visual appreciation during the tasting.
Global judge	Copper points (PCP) Quality score (SCO).	Global quantification of the sample

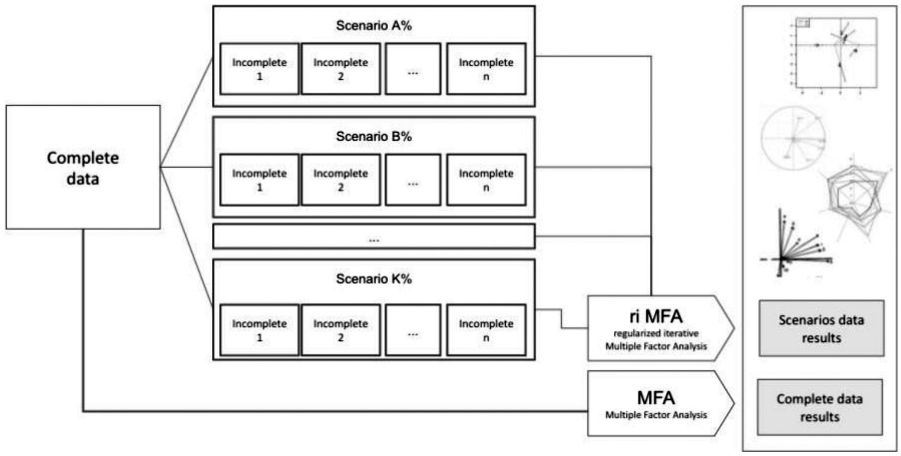

Figure 1. Schematic overview of the missing scenarios and RI-MFA inertia comparison.

makes a full disjoint table in qualitative variable groups and the NA is replaced according to the proportion of the values by column.

- **Partial analyses:** Likes MFA according to the type of Ki variables (quantitative or qualitative) a PCA or MCA is performed.
- **Juxtaposed table construction:** The technique builds a juxtapose table (Z) according to the ƛi1 ponderation.
- **PCA stage:** Likes MFA calculates a global PCA on a juxtaposed table (Z).
- **Cross-validation:** Technique takes q dimension to cross-validation and juxtaposes table is rebuilt. Only NA position in the table is imputed according to $Z = \psi^q u^q \lambda i1$.
- **Convergence:** RI-MFA is an iterative technique that repeats the Juxtaposed table construction and cross-validation stage until $|Zt - Z\,t{+}1| < e$.

Data analysis was carried out with R language [32]. A multivariate exploratory data analysis was conducted to understand the main relationship between subjects and sensory quality variables. The MFA function from FactoMineR [33] and imputeMFA from the missMDA [16] packages were used in experiments. Radar plot, web scraping, and heatmaps were generated using Python language. Code and scripts are available on Github [34]. Figure 1 shows a schematic overview of the scenario creation and RI-MFA comparison. The presence of 5%, 10%, 20%, and 30% missing values on the coffee sensorial dataset called *scenarios candidates* was imputed with RI-MFA.

RESULTS AND DISCUSSION

We conducted an exploratory data analysis to understand the coffee dataset and summarize the main characteristics. The scores given for the coffee characteristics were primarily numerical. We considered the complete dataset with two coffee species Arabica and Robusta; Arabica is produced in 36 countries, whereas Robusta is produced in 5 countries around the world. The sensory dataset contains 1341 assessments and 12 coffee features without missing values. In Table 2 means and deviation standard (SD) are given, and the means were taken as the reference value.

Data were expressed in terms of individuals and scores evaluated by species and country to determine

Table 2. Sensory panel assessments (means ± SD from the complete dataset).

Parameter	Arabica	Robusta
FRA	7.56 ± 0.38	7.70 ± 0.30
FLV	7.52 ± 0.40	7.63 ± 0.30
AFT	7.40 ± 0.41	7.56 ± 0.34
ACD	7.53 ± 0.38	7.66 ± 0.26
MOF	7.52 ± 0.41	7.51 ± 0.73
EQ	9.90 ± 0.60	7.54 ± 0.53
BTT	7.52 ± 0.36	7.68 ± 0.32
UCP	9.83 ± 0.62	9.90 ± 0.24
CCP	9.83 ± 0.82	9.93 ± 0.21
PCP	7.50 ± 0.48	7.76 ± 0.33
SCR	82.09 ± 3.68	80.87 ± 2.44

the central tendency and dispersion quality behavior as part of descriptive statistical analysis. Moreover, with MFA, continuous and categorical variables from the dataset were analyzed. The scores given for the characteristics Uniform cup (UCP), and Clean Cup (CCP) are near to max value; the panelist records indicate that samples do not have relevant differences. Thus, an analysis with these attributes could lead to uninteresting findings, whereas other characteristics such as Fragrance (FRA), Flavor (FLV), Copper points (PCP), and Score (SRC) are of interest.

According to Figure 2, Arabica (red) has the highest scores for Balance (EQ) and Quality Score (SRC). Robusta (blue) is superior in other characteristics. One relevant producing unit is located in Papua New Guinea, Japan, Ethiopia, and the United States. The coffee flavor (FLV) is one of the most important factors in differentiating quality. It is a list of attributes with greater weight in judging [35-38]. The Copper points (PCP) and Quality score (SRC) are global and subjective appreciation. Panelists determined the different sensory characteristics among the different samples, and according to several studies on genotype and environment influence, the source of the sample has a strong relationship with the quality of coffee.

Initially, descriptive sensory analysis performed on a worldwide coffee quality dataset demonstrated statistically that producer unit performance is similar. With the entire dataset, the results MFA show two selected components that explain about 79% of the total variation in the data set. The first dimension (Dim 1) explains 63.71% of variation, whereas the second (Dim 2) the 15.77%. The first part of the analysis consisted of identifying the producer unit's behavior according to coffee sensory attributes on the bottom-left of Figure 2. Lower Balance (EQ), Uniform Cup (UCP), and Clean Cup (CCP) scores lead to the location of these units in the plot, whereas the contributions of all units are shown in Figure 3. The above allowed us to represent the mentioned variables on the factorial plane visually. For instance, three producing units of the arabica located in Guatemala, Ethiopia, and Honduras are farther to balance (EQ), Uniform cup (UCP), and Clean Cup (CCP). The highest number of units are located at the center, with an average evaluation.; the producer units with the sample's best-rated attributes are in the top right corner of the plot. Multivariate missing scenarios start with a complete data set of (n) individuals and (m) variables. A missing value scenario has a percentage of Missing Values (NA). The percentage refers to the number of actual values removed from the complete dataset.

In comparison with the complete data, the MFA of imputed data shows an increment in the percentage of inertia near 74% when a percentage of 30% is

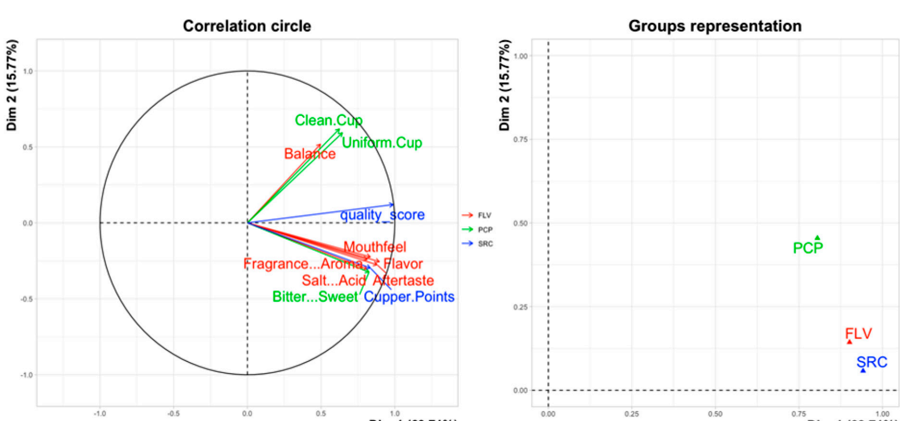

Figure 2. Top. Radar chart of Arabica and Robusta (Left) and first 15 producer units with highest scores from the dataset (Right); Bottom-left. Summarization of the correlation between Coffee attributes. Bottom-right. Contribution of eigenvalues Flavor (FLV), Copper Points (PCP), and Origin (SRC) associated with the total variance.

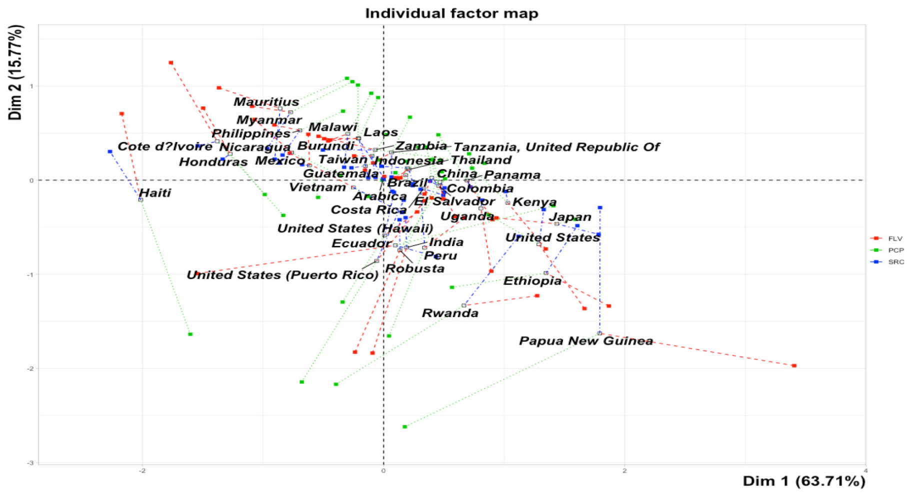

Figure 3. Different units are connected by lines where the line length is proportional to the divergence of the rated attributes.

set, which can be explained by similarity added during the imputation process after coffee-producing units are generally more similar to each other than in the real data. Figure 4 shows the plotting of the inertia percentage reached by MFA with imputed data and the missing values scenario. Despite the low relevance of the scenarios with a percentage higher than 50, the plot presents the tendency of cumulative inertia based on missing value proportion.

Cumulative inertia may suggest that the imputed data play a similar role to the observed data, which increases the probability of finding similar data and therefore results with a theoretical sense but not logical.

The analysis of missing values in different scenarios reveals that specific producing units or individuals lose their similar profile, and several are located far from each other on the factor map. Figure 5 shows the factor map of two imputed datasets in the presence of 30, 35, and 45 missing values. These values could be explained through plausible values, which the method imputes. These values do not test scores for

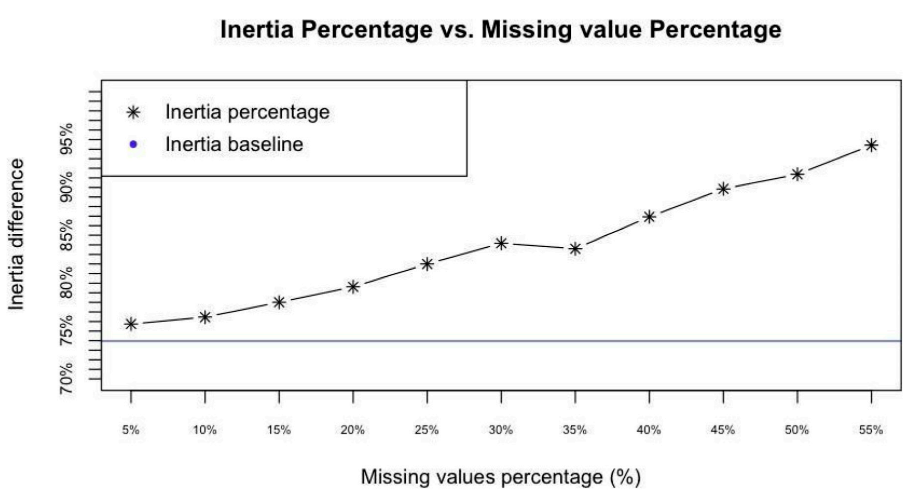

Figure 4. Percentage of explained inertia of the MFA of several missing scenarios.

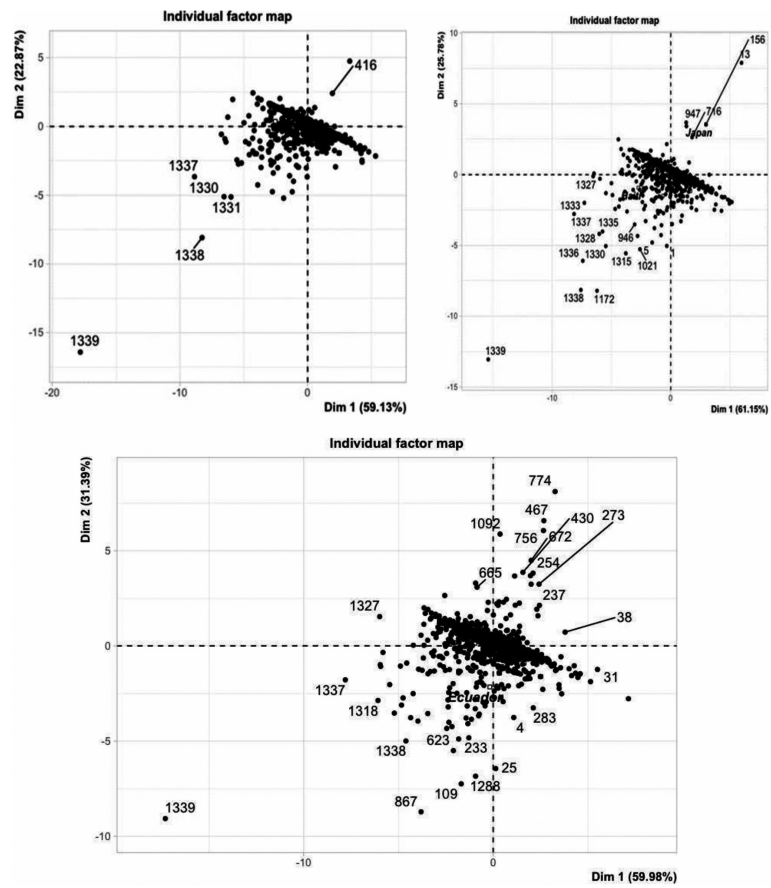

Figure 5. Factor maps for three imputed datasets with 30, 35, and 45 missing values presence.

individuals but offer an intermediate measurement from available information.

Figure 6 is presented correlation circles to 30, 35, and 45 missing values presented as a sample of these situations. As shown in this figure, the relationship between input and output variables remains positive, when missing values are at least 30%, the strength of the relationship increases, resulting in a decrease in the distance between the

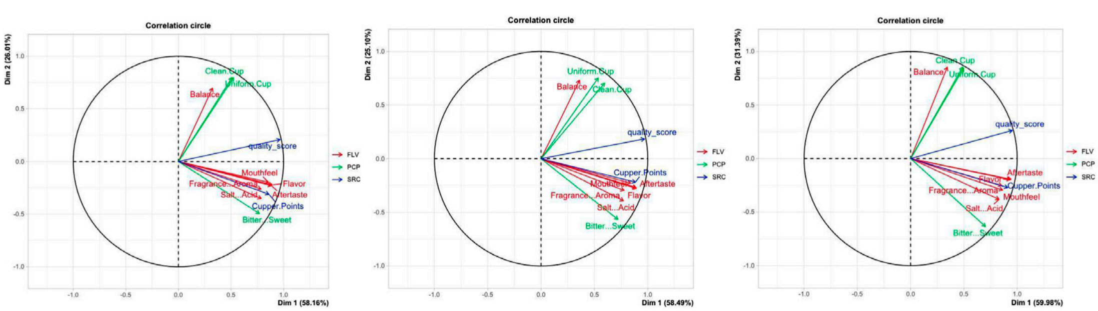

Figure 6. Correlation circle representing the projection of the input variables and the output variable in the imputed dataset with 30, 35, and 45 missing values presence.

factor axes and the variables. These results are the same for 35% and 45%. This finding suggests that the relationship between variables is overestimated with a direct positive relationship between bias and the percentage of imputed data.

There are many possible approaches to dealing with missing values in multivariate data. The comparative purpose is reached when in each RI-MFA imputed data set, a comparison of the inertia ratio was made from the MFA result. In this form, we make a modest contribution by providing empirical evidence to discuss how the imputation influences the inertial ratio when the dataset has missing values.

CONCLUSION

We obtained coherent results in this study when missing values were imputed using RI-MFA. These results are expected for multivariate analysis that as the percentage of NA grows, the inertia of all data points, which reveals component importance, also increases. In practice, it may be reasonable because the imputation method uses the available information and makes features according to other records in the dataset. However, analysts often obtained lower inertia when individuals have varied little.

Both MFA and RI-MFA are robust to the presence of NA and appear to be appropriate when sensory data have NA. Both MFA and RI-MFA perform well in the presence of NA and appear to be appropriate when sensory data contains NA. Based on the simulation study, it was observed that RI-MFA is a good strategy to estimate missing values, it would be interesting to see the behavior of the RI-MFA in other data sets and to combine this methodology with the cluster analysis. RI-MFA imputation may have good predictive accuracy in 1 to 10% of missing values; higher percentages may lead to severely biased inference when the imputed variables are used in subsequent factor analyses.

In this case study, it is shown that a correct analysis requires a broader knowledge of missing values and careful critique of the imputation mechanism. The imputation is implemented in many software packages and appears to be the solution in all cases where missing values are present. Future studies and experiments should incorporate other imputation methods, estimators, interpolators, and pooling estimates strategies which, following our schematic scenario configuration, assess the robustness of the MFA method in sensorial data analysis. Alternatively, future research should also include Nonlinear estimation by Iterative Partial Least Square (NIPALS), based on available information. This inclusion is because adequate results have been found in recent studies and are an alternative solution to the problem of NA [39-41].

REFERENCES

[1] T. Oberthür, P. Läderach, H. Posada, M.J. Fisher, L.F. Samper, J. Illera, L. Collet, E. Moreno, R. Alarcón, A. Villegas, H. Usma, C. Perez and A. Jarvis. “Regional relationships between inherent coffee quality and growing environment for denomination of origin labels in Nariño and Cauca, Colombia”. Food Policy. Vol. 36 Issue 6, pp. 783-794. 2011. DOI: 10.1016/j.foodpol.2011.07.005.

[2] L. Figueiredo, F. Borém, M. Cirillo, F. Ribeiro, G. Giomo and T. Salva. “The Potential for High Quality Bourbon Coffees from Different Environments”. Journal of Agricultural Science. Vol. 5 Nº 10. 2013. DOI: 10.5539/jas.v5n10p87.

[3] A.D. Bote and J. Vos. “Tree management and environmental conditions affect coffee (Coffea arabica L.) bean quality”. NJAS. Wageningen Journal of Life Sciences. Vol. 83, pp. 39-46. 2017. DOI: 10.1016/j.njas.2017.09.002.

[4] L.H. Ziska, B.A. Bradley, R.D. Wallace, C.T. Bargeron, J.H. LaForest, R.A. Choudhury, K.A. Garrett and F.E. Vega. “Climate Change, Carbon Dioxide, and Pest Biology, Managing the Future: Coffee as a Case Study”. Agronomy. Vol. 8 Nº 8. 2018. DOI: 10.3390/agronomy8080152.

[5] V. Belchior, B.G. Botelho, L.S. Oliveira and A.S. Franca. “Attenuated Total Reflectance Fourier Transform Spectroscopy (ATR-FTIR) and chemometrics for discrimination of espresso coffees with different sensory characteristics”. Food Chemistry. Vol. 273, pp. 178-185. 2019. DOI: 10.1016/j.foodchem.2017.12.026.

[6] M.B. dos S. Scholz, C.S.G. Kitzberger, S.H. Prudencio and R.S. dos S.F. da Silva. “The typicity of coffees from different terroirs

determined by groups of physico-chemical and sensory variables and multiple factor analysis". Food Research International. Vol. 114, pp. 72-80. 2018, DOI: 10.1016/j.foodres.2018.07.058.

[7] F. Ceballos-Sierra and S. Dall'Erba. "The effect of climate variability on Colombian coffee productivity: A dynamic panel model approach". Agricultural Systems. Vol. 190. 2021. DOI: 10.1016/j.agsy.2021.103126.

[8] J. Adhikari, E. Chambers and K. Koppel. "Impact of consumption temperature on sensory properties of hot brewed coffee". Food Research International. Vol. 115, pp. 95-104. 2019. DOI: 10.1016/j.foodres.2018.08.014.

[9] M.J. Chapko and H.-S. Seo. "Characterizing product temperature-dependent sensory perception of brewed coffee beverages: Descriptive sensory analysis". Food Research International. Vol. 121, pp. 612-621. 2019. DOI: 10.1016/j.foodres.2018.12.026.

[10] T.R. Lingle and S.N. Menon. "Cupping and Grading-Discovering Character and Quality". In the Craft and Science of Coffee, pp. 181-203. 2017. DOI: 10.1016/B978-0-12-803520-7.00008-6.

[11] A.M. Feria-Morales. "Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control". Food Quality and Preference. Vol. 13 Issue 6, pp. 355-367. 2002. DOI: 10.1016/S0950-3293(02)00028-9.

[12] D.F. Barbin, A.L. de S.M. Felicio, D.-W. Sun, S.L. Nixdorf and E.Y. Hirooka. "Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview". Food Research International. Vol. 61, pp. 23-32. 2014. DOI: 10.1016/j.foodres.2014.01.005.

[13] A. Pinsuwan, S. Suwonsichon, P. Chompreeda and W. Prinyawiwatkul. "Sensory Drivers of Consumer Acceptance, Purchase Intent and Emotions toward Brewed Black Coffee". Foods. Vol. 11 Issue 2. 2022. DOI: 10.3390/foods11020180.

[14] N. Gutiérrez-Guzmán, A. Cortés-Cabezas and E. Chambers IV. "A novel tasting platform for sensory analysis of specialty coffee". Coffee Sci. Vol. 13 Nº 3, pp. 401-409. 2018. ISSN: 1984-3909. URL: http://www.coffeescience.ufla.br/index.php/Coffeescience/article/view/1497

[15] S.E. Yeager, M.E. Batali, J.-X. Guinard and W.D. Ristenpart. "Acids in coffee: A review of sensory measurements and meta-analysis of chemical composition". Critical reviews in food science and nutrition, pp. 1-27. 2021. DOI: 10.1080/10408398.2021.1957767.

[16] M. Pardo and G. Sberveglieri. "Coffee analysis with an electronic nose". IEEE Transactions on Instrumentation and Measurement. Vol. 51 Issue 6, pp. 1334-1339. 2002. DOI: 10.1109/TIM.2002.808038.

[17] N. Buck, D. Wohlt, A. R. Winter and E. Ortner. "Aroma-Active Compounds in Robusta Coffee Pulp Puree-Evaluation of Physicochemical and Sensory Properties". Molecules. Vol. 26 Issue 13, pp. 3925. 2021. DOI: 10.3390/molecules26133925.

[18] N.Grujić-Letić, B. Rakić, E. Šefer, M. Milanović, M. Nikšić, I. Vujić and N. Milić. "Quantitative determination of caffeine in different matrices". Macedonian pharmaceutical bulletin. Vol. 62 Nº 1, pp. 77-84. 2016. ISSN: 14098695. DOI: 10.33320/maced.pharm.bull.2016.62.01.007.

[19] S. Cortés-Diéguez, C. Otero-Cerviño, H. Rodeiro-Mougán and J.A. Feijóo-Mateo. "Quantitative descriptive analysis of traditional herbal and coffee liqueurs made with grape marc spirit (Orujo)". Foods. Vol. 9 Issue 6, pp. 753. 2020. DOI: 10.3390/foods9060753.

[20] R.C.E. Dias, P. Valderrama, P.H. Março, M.B. dos S. Scholz, M. Edelmann and C. Yeretzian. "Quantitative assessment of specific defects in roasted ground coffee via infrared-photoacoustic spectroscopy". Food Chemistry. Vol. 255, pp. 132-138. 2018. DOI:10.1016/j.foodchem.2018.02.076.

[21] N. Motisi, P. Bommel, G. Leclerc, M.-H. Robin, J.-N. Aubertot, A. Arias Butron, I. Merle, E. Treminio and J. Avelino. "Improved forecasting of coffee leaf rust by qualitative modeling: Design and expert validation of the ExpeRoya model". Agricultural Systems. Vol. 197. 2022. DOI: 10.1016/j.agsy.2021.103352.

[22] J. Pagès. "Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes". Revue de Statistique Appliquée. Vol. 50 Nº 4, pp. 5-37. 2002. URL: http://www.numdam.org/item/RSA_2002__50_4_5_0/

[23] M.J. Greenacre. "Correspondence analysis". WIREs Computational Statistics. Vol. 2 Issue 5, pp. 613-619. 2010. DOI: 10.1002/wics.114.

[24] H. Abdi and D. Valentin. "Multiple factor analysis (MFA)". Encyclopedia of Measurement and Statistics, pp. 657-663. 2007.

[25] J. Pagès. "Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley". Food Quality and Preference. Vol. 16 Issue 7, pp. 642-649. 2005. DOI: 10.1016/j.foodqual.2005.01.006.

[26] J.-C. Barbe, J. Garbay and S. Tempère. "The Sensory Space of Wines: From Concept to Evaluation and Description. A Review". Foods. Vol. 10 Issue 6. 2021. DOI: 10.3390/foods10061424.

[27] A. Dabija, M.E. Ciocan, A. Chetrariu and G.G. Codină. "Maize and sorghum as raw materials for brewing, a review". Applied Sciences. Vol. 11 Issue 7. 2021. DOI: 10.3390/app11073139.

[28] S. Guerra, C. Lagazio, L. Manzocco, M. Barnabà and R. Cappuccio. "Risks and pitfalls of sensory data analysis for shelf life prediction: Data simulation applied to the case of coffee". LWT - Food Science and Technology. Vol. 41 Issue 10, pp. 2070-2078. 2008. DOI: 10.1016/j.lwt.2008.01.011.

[29] F. Husson and J. Josse. "Handling missing values in multiple factor analysis". Food Quality and Preference. Vol. 30 Issue 2, pp. 77-85. 2013. DOI: 10.1016/j.foodqual.2013.04.013.

[30] R.M. Schouten, P. Lugtig and G. Vink. "Generating missing values for simulation purposes: a multivariate amputation procedure". Journal of Statistical Computation and Simulation. Vol. 88 Issue 15, pp. 2909-2930. 2018. DOI: 10.1080/00949655.2018.1491577.

[31] M.S. Santos, R.C. Pereira, A.F. Costa, J.P. Soares, J. Santos and P.H. Abreu. "Generating Synthetic Missing Data: A Review by Missing Mechanism". IEEE Access. Vol. 7, pp. 11651-11667. 2019. DOI: 10.1109/ACCESS.2019.2891360.

[32] R core Team. "The R project for Statistical Computing". 2021. URL: https://www.R-project.org/

[33] S. Lê, J. Josse and F. Husson. "FactoMineR: An R Package for Multivariate Analysis". Journal of Statistical Software. Vol. 25 Issue 1, pp. 1-18. 2008. DOI: 10.18637/jss.v025.i01.

[34] A. Ochoa. "AndresOchoaRSA/AFMCafe". GitHub. Date of visit: July 16, 2021. URL: https://github.com/AndresOchoaRSA/AFMCafe

[35] J.C. Herrera and C. Lambot. "The Coffee Tree-Genetic Diversity and Origin". In The Craft and Science of Coffee, pp. 1-16. 2017. DOI: 10.1016/B978-0-12-803520-7.00001-3.

[36] I. Laukaleja and Z. Kruma. "Quality of specialty coffee: balance between aroma, flavour and biologically active compound composition: review". In International Scientific Conference: Research for Rural Development 2018. Jelgava, Latvia. May 16-18, 2018. URL:https://agris.fao.org/agris-search/search.do?recordID=LV2019000266

[37] A. Samoggia and B. Riedel. "Coffee consumption and purchasing behavior review: Insights for further research". Appetite. Vol. 129, pp. 70-81. 2018. DOI: 10.1016/j.appet.2018.07.002.

[38] L. Marie, C. Abdallah, C. Campa, P. Courtel, M. Bordeaux, L. Navarini, V. Lonzarich, A. Skovmand Bosselmann, N. Turreira-García, E. Alpizar, F. Georget, J.-C. Breitler, H. Etienne and B. Bertrand. "G×E interactions on yield and quality in Coffea arabica: new F1 hybrids outperform American cultivars". Euphytica. Vol. 216, pp. 1-17. 2020. DOI: 10.1007/s10681-020-02608-8.

[39] N. Patel, P. Mhaskar and B. Corbett. "Subspace based model identification for missing data". AIChE Journal. Vol. 66 Issue 1, pp. 1-13. 2020. DOI: 10.1002/aic.16538.

[40] V.M. González Rojas. "Inter-Battery Factor Analysis via PLS: The Missing Data Case". Revista Colombiana de Estadística. Vol. 39 Nº 2, pp. 247-266. 2016. DOI: 10.15446/rce.v39n2.52724.

[41] A.F. Ochoa- Muñoz, V.M. González-Rojas y C.E. Pardo-Turriago. "Missing data in multiple correspondence analysis under the available data principle of the NIPALS algorithm". DYNA. 86 Nº 211, pp. 249-257. 2019. DOI: 10.15446/dyna.v86n211.80261.