Advanced Crop Science Vol3. No8 (2013) 515–524 ISSN: 2322-4282

Contents lists available at KnowledgesPublisher

Advanced Crop Science

journal homepage: www.KnowledgesPublisher.com/Home

Selection of robust estimator usedin analysis of sensory characteristics and identification of environments conducive to specialty coffee production

Elcio do Nascimento Chagas[a], Augusto Ramalho de Morais[b], Marcelo Angelo Cirillo[c] Luisa Pereira Figueiredo[d], Flávio Meira Borém[e]

[a] *Instituto Federal do Espírito Santo - Campus Alegre, Alegre - ES., Brazil.*
[b] *Departament of Exact Sciences, Universidade Federal de Lavras, Campus-UFLA, Lavras, MG, CEP 37200-000,Brazil.*
[c] *Departament of Exact Sciences, Universidade Federal de Lavras, Campus-UFLA, Lavras, MG, CEP 37200-000,Brazil.*
[d] *Department of Food Science, Universidade Federal de Lavras, Campus-UFLA, Lavras, MG, CEP 37200-000, Brazil.*
[e] *Department of Agricultural Engineering, Universidade Federal de Lavras, Campus-UFLA, Lavras, MG, CEP 37200-000, Brazil.*

ARTICLEINFO

Article history:
Received 20 April 2013
Revised form 23 April 2013
Accepted 27 April 2013
Available online 25 August 2013

Keywords:

Outliers

robust estimation

Minimum Covariance Determinant

ABSTRACT

We propose multivariate regression models to identify environments for specialty coffee production. Simulation studies were carried out, taking as criterion the relative statistical efficiency between robust and non robust estimators in the presence of outliers related to covariance matrix of data. After selecting the model, we considered the sensory variables Aroma, Flavor, Acidity, Body and Balance, and the following chemical variables: final contents of trigonelline, chlorogenic acids and caffeine in the analysis of 14 genotypes of Bourbon cultivar evaluated in experiments in different environments. We concluded that the regression model considering MCD method showed larger evidence for identifying environments for specialty coffee production. In the case of MVE method, the multiple model showed greater evidence for discriminating genotypes.

1- Introduction

Coffee is one of the most popular beverages, many companies conduct a sensory evaluation of coffee for both new product development and quality control in order to provide coffee which meets consumer expectations. However, differences in raw materials, formulations and manufacturing processes led to the existence of a wide range of products, most of them registered under Protected Designation of Origin or Protected Geographic Indication (Constanza et al., 2012). For example, Hayakawa et al. (2010)a submitted a proposal to develop the lexicon of brewed coffee, which can be used commonly by coffee tasters and professionals in different companies in Japan, and to investigate the recognition of each term by consumers. The above-mentioned coffee lexicons, Seo et al. (2009), Hayakama et al. (2010)b and SCAA (Lingle 2001) contribute little to a precise sensory evaluation in Japan.

Overall quality is a determining factor in the price of coffee beans. Growing demand for high quality coffee is a trend in the world coffee market, as consumers are now more demanding and able to discern differences between types of coffee including origin, certification and taste characteristics.

The concept of specialty coffee is intimately connected to the pleasure of drinking the beverage due either to specific attributes, production process or services related to consumption. Specialty coffees differ from commodity coffees due to physical and sensorial attributes and limited availability (Leroy et al., 2006). Beverage quality and complexity are major distinguishing characteristics of specialty coffees.

Bourbon cultivar is internationally recognized for its genetic potential for producing coffee with excellent beverage quality. It is used in the specialty coffee production around the world because of its unique sensory characteristics, including a high level of natural sweetness, chocolate-like taste, intense aroma and agreeable acid levels.

In Brazil, the first records of cultivation of Bourbon variety date back to 1859. Over many years the first seeds were multiplied and mutations and natural variability occurred, resulting in different genotypes also called Bourbon, with variations in fruit color (red or yellow), plant height and productivity (Fazuoli et al., 2005).

Although coffee has no significant nutritional value, it is considered an important nutraceutical product. Coffee is basically consumed for the sensation of pleasure and satisfaction associated with it, especially the balanced combination of flavors and aromas (Illy 2002). Coffee market is in a scenario of constant changes.

Various factors affect coffee quality, such as genetics (species and varieties) (Pereira et al., 2010), environment soil, climate, altitude (Barbosa et al., 2012,) Vaast et al., 2006; Avelino et al., 2005), genotype-by-environment interactions and post-harvest procedures, especially processing methods, drying, and storage (Dias et al. 2012; Ribeiro et al., 2011; Borém et al., 2008).These factors have been considered the major determinants for the sensory profile of coffee beverage.

A high beverage quality is essential for consumption and appreciation of coffee. Cup quality is related to several biochemical compounds in green beans, which are transformed into a thousand of compounds during roasting. Some compounds may affect quality, such as trigonelline, sugars, soluble solids, chlorogenic acids and caffeine (Clifford 1985 and Macrae, 1985).

Statistical modeling in analysis of specialty coffee production shows numerous factors that make it reasonable to assume the existence of atypical observations, known as outliers. Confirming the existence of such observations in a sample requires careful data analysis. However, the identification of outliers in multidimensional contexts is generally performed with graph-based methods. Major contributions were made by Gnanadesikan and Kettenring (1972) and more recently by Rocke and Woodruff (1996) and Filzmoser et al. (2008). Jolliffe (2002) states that, as there are too many variables, identification of outliers generally requires considering possible observations showing up in directions other than those detectable in simple plots for pairs of original variables.

An alternative to solve this problem is the application of robust methods, which are limited to a fraction of outliers in the sample. Hubert et a.l (2008) mention that such methods provide among others the following advantages: description of a structure that best fits a data set and identification of points deviating from the majority, which have great influence on the rest of the assembly.

For estimating the covariance matrix used in multiple analysis, robust estimation methods are based on resampling subsets of smaller size in relation to sample size considering the break point, which is the highest percentage of contamination an estimator can support when providing an accurate estimate. In this methodology, the most common robust estimators of covariance matrix are the methods Minimum Covariance Determinant Estimator (MCD) (Rousseeuw and Driessen, 1999) and Minimum Volume Ellipsoid Estimator (MVE) (Chen and Jackson, 2004, Cirillo and Barroso, 2011).

In the MCD method the selected subset will be used for estimating the covariance matrix showing the minor determinant as compared to matrices from other subsets. Similarly, with the method MVE the selected subset will result in a covariance matrix generated by an ellipsoid of minimal volume covering at least (N/2+1) sampling points, N being set as the sample size.

Based on this information, we conducted a comparative study of robust estimators of covariance matrix and the least squares estimator to be applied in multivariate regression models, in order to identify environments conducive for specialty coffee production, considering different genotypes of arabica coffee (*Coffea arabica* L.).

2- Materials and methods

We used the methods according to the following description: Definition of measure of relative efficiency used for comparing estimators of covariance matrix of data; Simulation studies for selecting estimators of covariance matrix of data in the presence of outliers; characterization of experiments conducted in different environments and regions for specialty coffee production, and application of multivariate regression models for identifying environments for specialty coffee production.

- Definition of measure of relative efficiency used for comparing estimates of covariance matrix of data:

Simulation studies were performed to evaluate the relative efficiency (1) defined between different estimators of covariance matrix, based on the use of determinants representing a measure of variability of p-variables in each sampling unit, which were summarized in a single value.

$$\mathbf{RE}(\hat{\Sigma}_{R_i}, \hat{\Sigma}) = \left(\frac{\left|\hat{\Sigma}\right|}{\left|\hat{\Sigma}_{R_i}\right|} \right)^{\frac{1}{p}} \quad \text{for} \quad i = 1, 2.$$

wherein $\hat{\Sigma}_{R_i}$ corresponds to the robust estimator of covariance matrix of data, $\hat{\Sigma}_{R_i}$ being the MCD estimator (Rousseeuw et al., 2004), $\hat{\Sigma}_{R_i}$ being the MVE estimator (Rousseeuw and Zomeren 1990), and Σcorresponding to the least squares estimator, henceforth mentioned as MMQ (Zen and Budescu, 2006).

- Simulation studies for selection of covariance matrix estimators in the presence of outliers:

Based on the measure RE (1) used as a criterion to select estimators of covariance matrix of data, we proceeded with simulation studies assuming the parametric values defined by covariance structures and vectors with averages specified in (2), ρ = 0.5.

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{p\times 1} ; \ \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & \rho^{1-2} & \cdots & \rho^{1-p} \\ \rho^{2-1} & 1 & \cdots & \rho^{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & 1 \end{bmatrix}_{p\times p} \text{ e } \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}_{p\times p}$$

To generate samples contaminated with outliers, we considered t-student and multiple lognormal distributions. Such distributions were chosen due to low and high kurtosis, respectively, compared to multiple normal distributions. Thus, multidimensional random variables contaminated by outliers were defined by G1i and G2i for i = 1, ..., n in (3) - (4), mixing probabilities arbitrarily set at δ = 0.05 and 0.4, sample sizes set at n = 25, 50, 100 and 200, and finally number of variables fixed at p = 5 and 10.

$$G_{1i} = \delta N_p(\boldsymbol{\mu}_2, \Sigma_2) + (1\text{-}\delta)\, t_p(\Sigma_2, \nu = 5) \quad (i=1,\ldots,n)$$

ν being the number of degrees of freedom in t-student distribution.

$$G_{2i} = \delta N_p(\boldsymbol{\mu}_1, \Sigma_1) + (1\text{-}\delta)\, W_p \text{ wherein}$$

$$W_p = \left[\exp(Z_1), \exp(Z_2), \ldots, \exp(Z_p)\right] \text{ and } Z \sim N_p(\mathbf{0}, \Sigma_1)$$

After sample generation for each combination of n, δ, and p, we calculated the average of estimates of relative efficiency (RE) in 1000 Monte Carlo simulations. Comparing the expected value of RE with the unit value allowed us to interpret E (RE)> 1, thus we concluded that robust estimators of covariance matrix are more efficient and therefore recommended in situations similar to those evaluated. To obtain results, we developed functions in software R Development Core Team (2011).

- Characterization of the experiments conducted in different environments and regions for specialty coffee production:

Field experiments were established in southern Minas Gerais state (municipalities of Lavras and Santo Antônio do Amparo) and region of Mogiana Paulista, state of São Paulo (municipality of São Sebastião da Grama). Therefore, different edaphoclimatic conditions of coffee producing regions were represented in the study. (Table 1).

Experimental results in 14 coffee genotypes (Table 2) (*Coffee arabica* L.) showed 11 genotypes from Bourbon cultivar, known for its high potential for specialty coffee production, and three genotypes from commercial cultivars widely cultivated in Brazil.

The group of genotypes Bourbon is a segregating population obtained from seeds collected in different coffee regions in Brazil. The other three cultivars, Mundo Novo IAC 502/9, Catuaí Vermelho IAC 144 and Icatu Precoce IAC 3282, used as a reference for assessing quality, came from research institutions and account for over 90% of commercial crops in Brazil. Climatic parameters of air temperature and precipitation were monitored continuously by automatic weather stations installed near the areas of study. A summary of edaphoclimatic characteristics for each altitude (environment) is shown in Table 3.

The three experiments (Table 3) were installed in randomized block design with three replications at spacing 3.5 × 0.8 m and experimental plot comprising 10 plants. Sensory attributes were obtained by sensory analysis conducted by certified judges according to the methodology proposed by the SCAA, Specialty Coffee Association of America (Lingle 2001). Contents of caffeine, trigonelline and chlorogenic acids were determined according to Casal et al. (2000) and the results given in percentage of dry matter (% ms). All laboratory analyzes were performed in triplicate.

- Application of multivariate regression models for identifying environments for specialty coffee production:

Given the specifications of the experiments, we proceeded with setting of multivariate regression models assuming *k* regressive variables and *p* response variables. Thus, the proposed multivariate regression model was given by (5) and adjusted in three situations with the estimators MMQ, MCD and MVE.

$$\mathbf{Y}_{(n\times p)} = \mathbf{X}_{(n\times(k+1))} + \boldsymbol{\beta}_{((k+1)\times p)} + \boldsymbol{\varepsilon}_{(n\times p)}$$

Wherein n is the sample size, Y(n×p) the matrix of response variables, X(n×(k +1)) the matrix of observations of regression variables, β((k +1)×p) the matrix of regression coefficients (unknown parameters), which is being estimated, and ε(n×p) the matrix of random error, so that E(ε)=0, Cov(ε)= Σ_{ε}. Thus, in the set of 14 genotypes of coffee relating to different altitudes, Y(14×5) is a matrix containing the mean values of sensory attributes Aroma (Y1), Flavor (Y2), Acidity (Y3), Body (Y4) and Balance (Y5). In addition, X(42× 4) corresponds to a matrix containing mean levels of trigonelline (X1), chlorogenic acids (X2) and caffeine (X3), whereas β(4×5) is a matrix of regression coefficients which were the unknown parameters to be estimated, and ε(42×5) is a matrix containing the components of random error associated with the mean values of sensory attributes.

After adjustment of model (5) considering non robust and robust estimates of covariance matrix, the measure RE (1) was applied and interpreted while keeping the unit value as a reference. Thus, it was possible to verify the performance of each estimator in the model fit for choosing a model to describe chemical and sensory profile of Bourbon coffee.

After selecting the model, we identified outliers by considering the Mahalanobis distances di (6) computed by

$$d_i = \sqrt{\left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0\right)^t \hat{\boldsymbol{\Sigma}}_0^{-1} \left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0\right)}$$

Being $\hat{\boldsymbol{\mu}}_0 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ the k-dimensional estimator of vector of averages; and of averages; and

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0\right)\times\left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0\right)^t$$

the estimator of covariance matrix by the method of least squares. Subsequently, we found robust Mahalanobis distances by substituting the estimates $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\Sigma}}_0$ by $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ which were obtained by considering MCD or MVE as robust estimates and selected through RE (1). Finally, estimation of critical values used in outlier detection in functional observations followed recommendations by Atkinson (1994), Peña and Prieto (2001).

3- Results and discussion

3-1- Monte Carlo evaluation of the relative efficiency between estimates of covariance matrix in methods MCD, MVE and least squares

Keeping the parametric specifications described in methodology, simulation results corresponded to the relative efficiency (RE), which was calculated by considering the ratio of determinants of covariance matrix estimated by MCD and MVE methods in relation to the determinant estimated by MMQ method. We assumed a low number of outliers (δ = 0.05) in the sample, which was generated by t-student and multiple lognormal distributions respectively characterized by small and high kurtosis when compared to multiple normal distribution.

Thus, the results described in Figures 1A and 1B for situation (δ = 0.05) show statistical evidence that estimation of covariance matrix by MMQ method should be used instead of the estimation provided by MCD method, since MMQ values of relative efficiency RE are closer to one. However, when comparing relative efficiency of MMQ and MVE estimators (Figures 1C and 1D) we found that MVE method is recommended instead of MMQ for all sample sizes evaluated.

Since MCD method yields more significant RE results than MVE method, which shows values far from one (Figures 1C and 1D), the applicability of these robust methods is in agreement with studies by Filzmoser and Todorov (2009), in which the authors claim that MCD was neglected in favor of MVE. However, the authors state that MCD estimator is not very efficient for normal or approximately normal models, especially if h is selected in a way that the highest breakpoint is reached. On the other hand, Rousseeuw and Zomeren (1990) state that MVE estimator is satisfactory with regard to computational cost, as the algorithm uses resampling processes through elementary sets. In practical terms, therefore, MVE method is recommended for regression analysis.

With regard to simulation scenarios based on a significant number of outliers (δ = 0.40), results for efficiency of robust covariance matrix estimators were more discrepant for relative efficiency (RE) with MMQ method, given the high kurtosis. Thus, results in Figures 2B and 2D show that the relative efficiency provided by MVE method (Figure 2D) is higher than by MCD method

(Figure 2B). Similarly, this effect was observed when considering small kurtosis (Figures 2A and 2C).

By increasing the number of variables (p = 10), results in Figure 3 were similar to previous situations (Figures 1 and 2). However, MMQ method showed higher efficiency when estimation of the determinant of covariance matrix, obtained by MVE method, occurred when samples were contaminated with outliers from lognormal distribution, characterized by high kurtosis.

3-2- Selection and diagnosis of multivariate regression models to identify environments conducive for specialty coffee production

Based on the multivariate regression model specified in (5), we obtained estimates of regression parameters considering the methods of least squares (MMQ), Minimum Covariance Determinant (MCD) and Minimum Volume Ellipsoid (MVE), as described in Table 4.

Based on the results described in Table 4, estimates of intercepts were similar for all response variables (Y1, .., Y5) classified by sensory attributes, which suggests that sample outliers are probably not leverage points (Barrett and Ling 1992). However, as some parameters estimates of the chemical variables final content of trigonelline (X1), chlorogenic acids (X2) and caffeine (X3) show differences due to estimation methods, there is evidence that outliers are influential points affecting parameter estimates.

To obtain more confirmatory results about the use of regression models with robust estimates of covariance matrix, results described in Table 5 show the measure of relative efficiency (RE) of MCD and MVE estimators in relation to MMQ estimator.

Results in Table 5 show that in fact MCD and MVE methods are recommended for estimation of covariance matrix of data for constructing regression models to identify environments for specialty coffee production. This fact is confirmed by the results of relative efficiency, which increased 21.89% to 38.29%. Thus, we proceeded with outlier identification keeping the robust estimates of regression model parameters. Figures 4 and 5 show estimates of robust Mahalanobis distances, as described in methodology.

In figure 4, outlier identification with robust distances computed by estimates of MCD method shows sample observations numbered 2, 17, 18, 24, 28 and 38. Observations 17, 18, 24 and 28 belong to the environment São Sebastião da Grama. In a study of interaction of these genotypes and environments with coffee beverage quality. São Sebastião da Grama yielded the highest levels of caffeine and trigonelline in coffee beans, as well as the best sensory quality.

Since higher altitudes enable the production of better quality coffee (Silva et al., 2006), it is expected that genotypes are better able to express their quality in São Sebastião da Grama (altitude 1300m) when compared to other environments. Thus, figure 4 shows the identification of this environment, which has greater ability for producing special coffee.

Therefore, outlier identification considering MCD method allows us to affirm that the multivariate regression model proposed to correlate sensory attributes and chemical variables, by assuming robust estimates of covariance matrix, is feasible to identify environments for specialty coffee production.

In figure 5, outlier identification with robust distances computed by estimates of MVE method shows the observations numbered 4, 12, 36 and 40 identified as outliers. Observations 2, 4 and 12 refer to the environment Santo Antônio do Amparo, observations 36, 38 and 40 to the environment Lavras, and the others to the environment São Sebastião da Grama. Thus, figure 5 shows that it was not possible to differentiate the environments through outliers, as shown in Figure 4. However, given the agreement in outlier identification when considering MCD and MVE methods in relation to the same observations (2,17,18,24,28 and 38), we can state that the proposed model was feasible to identify these genotypes (Mundo Novo IAC 502/9; Catuaí Vermelho IAC 144; Icatu Precoce IAC 3282; Bourbon Amarelo LCJ 10; Bourbon Amarelo).

4- Conclusion

Given the situations in the simulation study, we concluded that the measure of relative efficiency of robust estimators of covariance matrix was not affected by increasing number of variables. In terms of application, the regression model considering MCD method presented greater evidence for identifying environments for specialty coffee production. In the case of MVE method, the multiple model showed greater evidence for discriminating genotypes.

5- References

Atkinson, A.C. (1994). Fast very robust methods for the detection of multivariate outliers. J Am. Stat. Assoc .89, 1329-1339.

Avelino, J.; Barboza, B.; Araya, J.C.; Fonseca, C.; Davrieux F, Guyot, B.; Cilas, C. (2005). Effects of slope exposure, altitude and yield on coffee quality in two altitude terroirs of Costa Rica, Orosi and Santa Mara de Dota. J Sci Food Agric. 85, 1869-1876.

Azen, R.; Budescu, D. (2006). Comparing predictors in multivariate regression models: Na extension of dominance analysis. J. Educ. Behav. Stat. 31, 157-180.

Barret, B.; Ling, R.F. (1992). General classes of influence measures multivariate regression. J Am Stat Assoc .87,184-191.

Barbosa, J. N. Borem, F.M. ; Cirillo, M. A. ; Malta, M. R. ; Alvarenga, A. A. ; Alves, H. M. R. (2012). Coffee quality and its interactions with environmental factors in Minas Gerais, Brazil. Journal of Agricultural Science, v. 4, p. 181-190.

Borém, F.M.; Coradi, P.C; Saath, R.; Oliveira, J.A. (2008). Qualidade do café natural e despolpado após secagem em terreiro e com altas temperaturas. Ciênc. Agrotec. 32, 1609-1615.

Casal, S.; Oliveira, B.; Ferreira, M.A. (2000). HPLC/ diode-array applied to the thermal degradation of trigonelline, nicotinic acid and caffeine in coffee. Food. Chem. 68, 481-485.

Cirillo, M.A.; Barroso, L.P. (2012). Robust regression estimates in the prediction of latent variables in structural equation models. J. Modern Appl. Stat. Methods, v.11,p.42-53.

Clifford, M.N. (1985). Chemical and physical aspects of green coffee and coffee products. In Coffee: botany, biochemistry and production of beans and beverage, (Clifford. MN and Willson KC, Eds.) pp. 305-374, New York, Croom Helm.

Constanza M. L.; Elena Bru, E.; Vignolo, G.M.; Fadda, .S.G. (2012). Main Factors Affecting The Consumer Acceptance Of Argentinean Fermented Sausages. J. Sensory Studies 27, 304-313.

Dias, E.C.; Borém, F.M.; Pereira, R.G.F.A.; Guerreiro, M.C. (2012). Amino acid profiles in unripe Arabica coffee fruits processed using wet and dry methods. Eur. Food. Res. Technol. 234, 25-32.

Fazuoli, L.C.; Guerreiro, F.O.; Silvarolla, M.B.; Medina, F.H.P., Carvalho, A. (2005) Avaliação das cultivares de Mundo Novo, Bourbon Amarelo e Bourbon Vermelho de Coffea arabica L. em Campinas, SP. Bragantia. 64, 533-546.

Filzmoser, P. (2005). Identification of Multivariate Outliers: A Performance Study. Aust J Stat. 34, 127-138.

Filzmoser, P.; MARONNA, R.; WERNER, M. (2008). Outlier identification in high dimensions. Comput. Stat. Data. An. 52, 1694-1711.

Gao, S.; Mokhtarian, P.L.; Johnston, R.A. (2008). Nonnormality of Data in Structural Equation Transportation Research Record. Journal of the Transportation Research Board. 2082, 116-124.

Gnanadesikan R.; Kettenring, J.R. (1972). Robust estimates, residuals and outlier detection with multiresponse data. Biometrics. 28, 81-124.

aHayakawa, F.; Kazami, Y.; Wakayama, H.; Oboshi, R.; Tanaka, H.; Gou Maeda, C.H.; Iwawaki, H.; Miyabayashi, T. (2010). Sensory lexicon of brewed coffee for japanese consumers, untrained coffee professionals and trained coffee tasters. J. Sensory Studies. 25, 917-939.

bHayakawa, F., Ukai, N., Nishida, J., Kazami, Y. and Kohyama, K. (2010). Lexicon for the sensory description of French bread in Japan. J. Sensory Studies 25, 76–93.

Hubert, M.; Rousseeuw, P.J.; Aelst, S.V. (2008). High-breakdown robust multivariate methods. Statistical Science. 3, 92-119.

Illy, E.A. (2002). The complex flavors of coffee. Scientific American. 286, 48-53.

Jackson, D.A.; Chen, Y. (2004). Robust principal component analysis and outlier detection with ecological data. Environmetrics. 15, 129-139.

Jolliffe, I.T. (2002). Principal Component Analysis. New York: John Wiley.

Leroy, T.; Ribeyre, F.; Bertrand, B.; Charmetant, P.; Dufour, M.; Montagnon, C.; Marraccini, P.; POT, D. (2006). Genetics of coffee quality. Brazilian Journal of Plant Physiology. 18, 229-242.

Lingle, T.R. (2001). The coffee cupper's handbook: systematic guide to the sensory evaluation of coffees flavor. Long Beach, Specialty coffee Association of America.

Macrae, R. (1985). Nitrogenous compounds In: Coffee (Clarke R, ed.), London, Elsevier Applied Science.

Pereira, M.C.; Chalfoun, S.M.; Carvalho, G.R. D.; Savian, T.V (2010) Multivariate analysis of sensory characteristics of coffee grains (Coffea arabica L.) in the region of upper Paranaíba. Acta Sci Agron. 32, 635-641.

Pea, D.; Prieto, F.J. (2001). Multivariate outlier detection and robust covariance matrix estimation. Technometrics. 43, 286-310.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ribeiro, F.C.; Borém, F.M.; Giomo, G.S.; Lima, R.R.; Malta, M.R.; Figueiredo, L.P. (2011). Storage of green coffee in hermetic packaging injected with CO2. J. Stored Prod. Res. 47, 341-348.

Rocke, D.M.; Woodruff, D.L. (1996). Identification of outliers in multivariate data. J. Am. Stat Assoc. 91,1047-1061.

determinant estimator. Technometrics. 41, 212-223.

Rousseeuw, P.J.; Aelst, S.V.; Driessen, K.V.; Agulló, J. (2004). Robust multivariate regression. Technometrics. 46, 293-305.

Rousseeuw, P.J.; Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. J Am Stat Assoc. 85, 633-651.

Seo, H.S., Lee, S.Y. and Hwang, I.K. (2009). Development of sensory attribute pool of brewed coffee. J. Sensory Studies 24, 111–132.

Silva, R.F D.; Pereira, R.G.F.A.; Borém, F.M.; Silva, V.A. D. (2006). Altitude e a qualidade do café cereja descascado. Revista Brasileira de Armazenamento. 9,40-47.

Vaast, P.; Bertrand, B.; Perriot, J.J.; Guyot, B.; Granard, M. (2006) Fruit thinning and shade improve bean characteristics and beverage quality of coffee (Coffea arabica L.) under optimal conditions. J. Sci Food Agric. 86, 197-204.

Table 1- Locality of the experiments.

Town	State Region	Place
Lavras	southern Minas	Coffee Sector - DAG/UFLA
Santo Antônio do Amparo	southern Minas	Cerrado Grande Farm
São Sebastião da Grama	Mogiana Paulista	Recreio Farm

Table 2- Arabica coffee genotypes evaluated in the experiment.

ID	Genotype	Origin
1	Bourbon Amarelo	Epamig - Machado/MG
2	Mundo Novo IAC 502/9	Epamig - Machado/MG
3	Catuaí Vermelho IAC 144	Epamig - Machado/MG
4	Icatu Precoce IAC 3282	Procafé - Varginha/MG
5	Bourbon Amarelo1	Procafé - Varginha/MG
6	Bourbon Amarelo2	Santo Antônio do Amparo/MG
7	Bourbon Vermelho3	Campos Altos/MG
8	Bourbon Amarelo LCJ 9	IAC - Campinas/SP
9	Bourbon Amarelo4	São Sebastião do Paraíso/MG
10	Bourbon Amarelo LCJ 10	Oliveira/MG
11	Bourbon Amarelo5	Carmo de Minas/MG
12	Bourbon Amarelo6	Carmos de Minas/MG
13	Bourbon Trigo7	Alfenas/MG
14	Bourbon Amarelo8	Santo Antônio do Amparo/MG

IAC – Instituto Agronômico; Epamig – Empresa de Pesquisa Agropecuária de Minas Gerais; Procafé – Fundação Procafé;
Origin – refers to the institution, city and state (Brazil) where the genotypes were collected to be used in the experiment carried out in Lavras, São Sebastião da Grama and Santo Antônio do Amparo. 1,2,3,4,5,6,7 e 8 – unidentified lineage of Bourbon.

Table 3- Edaphoclimatic characteristics of each environment/altitude.

Environment	Altitude	Type of Soil	Average Annual Temperature	Average Annual Precipitation
Lavras	950 m	Red Latosol, clay texture	20.4ºC	1,460 mm
São Sebastião da Grama	1300 m	Yellow Latosol, clay texture	20ºC	1,560 mm
Santo Antônio do Amparo	1050 m	Red Latosol, clay texture	19.9ºC	1,700 mm

Table 4- Summary of estimates for parameters of regression models adjusted for sensory analysis of coffee quality.

Method	Variables	Coef.	Aroma (Y1)	Flavor (Y2)	Acidity (Y3)	Body (Y4)	Balance (Y5)
MMQ	Intercept	β0	7.7861	7.5941	7.3816	7.6731	6.9385
	X1	β1	-0.2282	-0.4047	0.0012	-0.3631	0.0224
	X2	β2	0.0224	-0.0638	-0.0120	0.0306	-0.1206
	X3	β3	-0.2553	0.3332	-0.0413	-0.2324	0.6530
MCD	Intercept	β0	7.4880	6.2767	7.1931	7.0121	6.7473
	X1	β1	-0.0865	-0.6924	0.0122	-0.6548	0.1331
	X2	β2	-0.0146	-0.0231	-0.0410	0.0579	-0.1031
	X3	β3	0.0525	1.4815	0.2488	0.4443	0.6301
MVE	Intercept	β0	7.1247	6.5749	7.3610	7.2030	6.8329
	X1	β1	0.3549	-0.3546	-0.0127	-0.6242	0.4823
	X2	β2	0.0507	-0.0070	0.0018	0.0560	-0.1052
	X3	β3	-0.3620	0.8466	-0.0830	0.2602	0.2622

Table 5- Relative Efficiency of MCD and MVE methods in relation to the method of least squares.

Method	Relative Efficiency
MCD	1.2189
MVE	1.3829

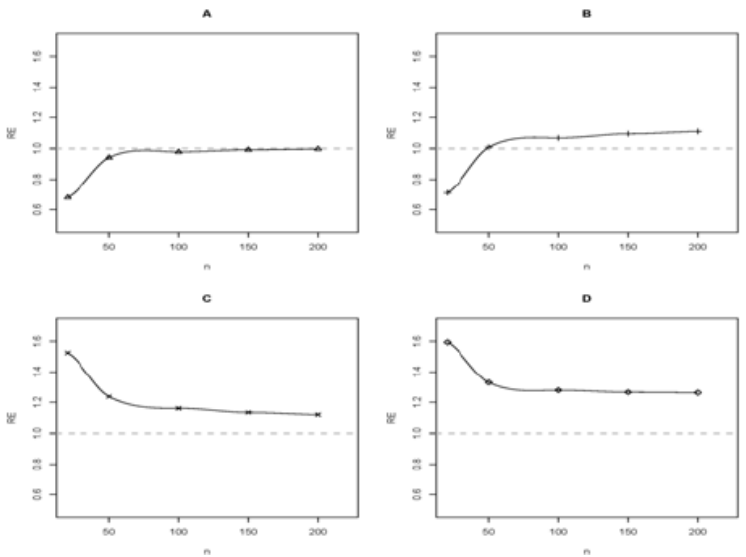

Fig. 1

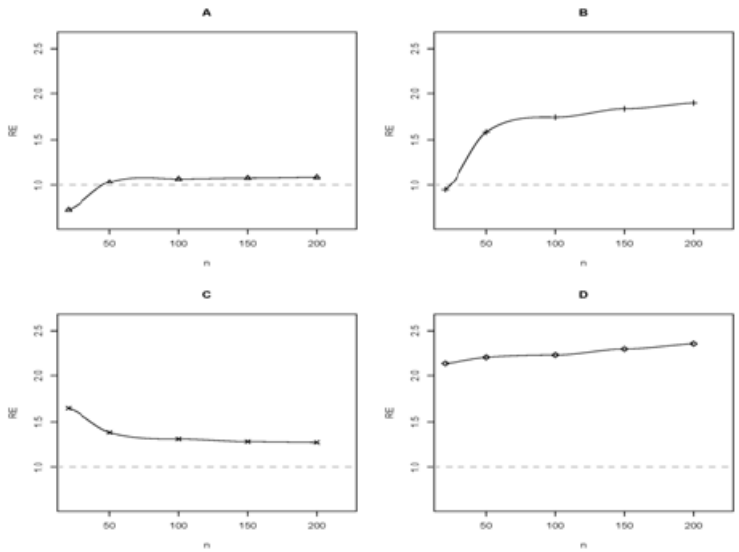

Fig. 2

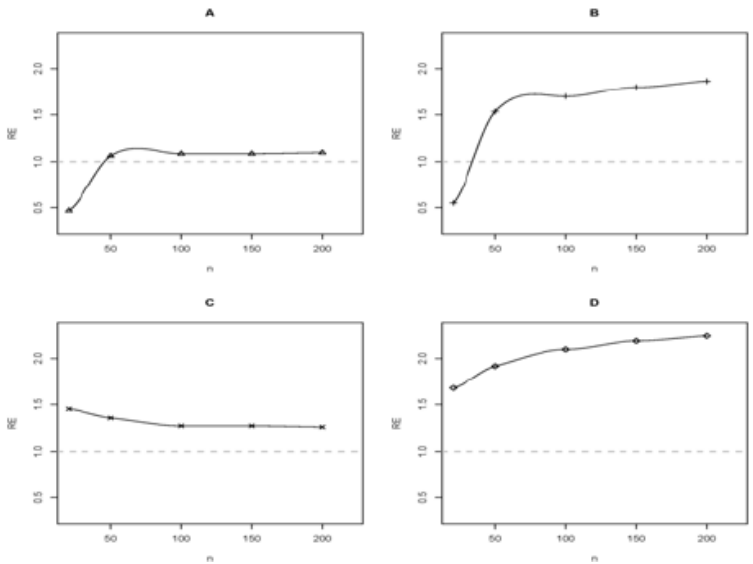

Fig. 3

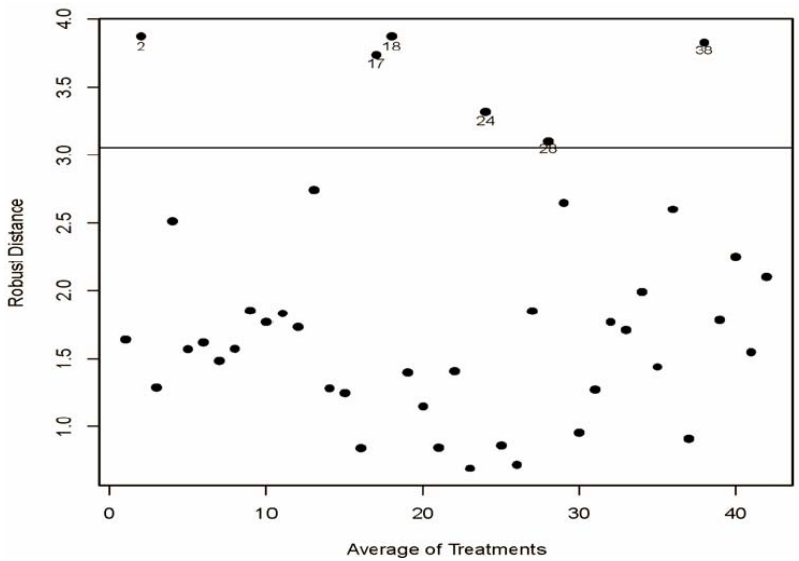

Fig. 4

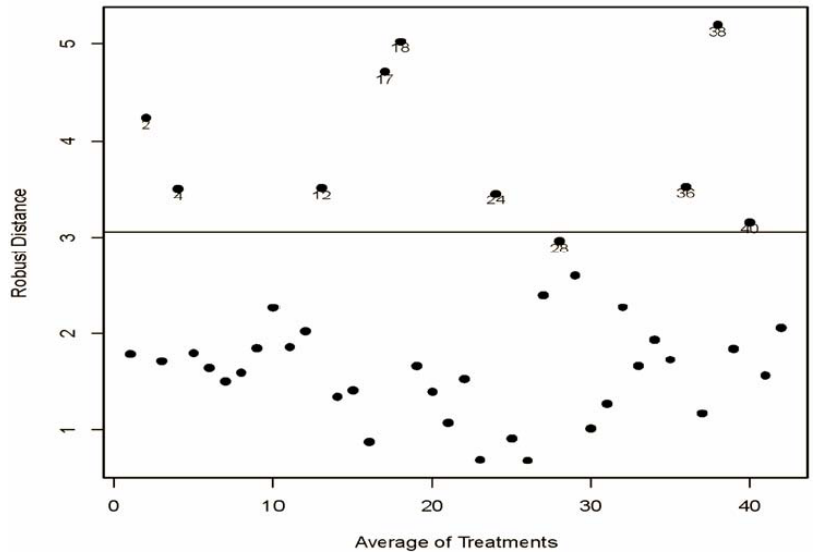
Robust Distance
Average of Treatments
0
10
20
30
40
1
2
3
4
5
2
4
12
17
18
24
28
36
38
40

Fig. 5