Utilização de estatística multivariada na predição da qualidade físico-química de leite

Use of multivariate statistics to predict the physicochemical quality of milk

Uso de estadísticas multivariadas para predecir la calidad fisicoquímica de la leche

Lenara Oliveira Pinheiro

ORCID: https://orcid.org/0000-0001-7867-1363

State University of Southwestern Bahia, Brazil

E-mail: lenarapinheiro@hotmail.com

Mário Roberto Júnior

ORCID: https://orcid.org/0000-0002-5615-0185

State University of Southwestern Bahia, Brazil

E-mail: mariorjr89@gmail.com

Clara Mariana Gonçalves Lima

ORCID: https://orcid.org/0000-0002-3150-9139

Federal University of Lavras, Brazil

E-mail: claramarianalima@gmail.com

Heliara Caires Sousa

ORCID: https://orcid.org/0000-0002-6693-8582

State University of Southwestern Bahia, Brazil

E-mail: heliara.cs@gmail.com

Jorge Pamplona Pagnossa

ORCID: https://orcid.org/0000-0002-7710-2254

Federal University of Lavras, Brazil

E-mail: jorgepampa@gmail.com

Leandro Soares Santos

ORCID: https://orcid.org/0000-0001-7431-318X

State University of Southwestern Bahia, Brazil

E-mail: leosoaressantos@yahoo.com.br

Sérgio Augusto de Albuquerque Fernandes

ORCID: https://orcid.org/0000-0003-3203-0792

State University of Southwestern Bahia, Brazil

E-mail: sfernandes@uesb.edu.br

Resumo

A análise multivariada envolve a aplicação de métodos estatísticos e computacionais para predizer respostas. Dentre os diversos métodos de análise estatística multivariada, a análise por componentes principais recebe destaque para efetuar a previsão da composição e qualidade de alimentos em geral. Objetivou-se, com o presente trabalho, caracterizar os produtores de leite do município de Itapetinga-BA, utilizando análise de componentes principais. Foram utilizadas 20 amostras de leite cru, coletadas na recepção do laticínio localizado em Itapetinga-BA. As variáveis analisadas foram: gordura, densidade, extrato seco desengordurado, proteína e lactose. Os dois primeiros componentes principais explicaram 87,24% da variação total. Verificou-se a formação de diferentes grupos distribuídos nos quatro quadrantes do sistema. O quadrante I destacou-se dos demais por formar um grupo composto por dez produtores da região analisada, caracterizando-se por apresentar amostras com maior teor de lactose e menor teor de gordura no leite. As variáveis lactose e gordura apresentam maior importância na caracterização do leite.

Palavras-chave: Métodos; Componentes principais; Produtores; Matriz de origem animal.

Abstract

Multivariate analysis involves the application of statistical and computational methods to predict responses. Among the various methods of statistical analysis multivariate, the analysis by main components is highlighted to predict the composition and quality of food in general. The objective of this work was to characterize the milk producers of the municipality of Itapetinga-BA, using principal component analysis. Twenty samples of raw milk were used, collected at the reception of the dairy located in Itapetinga-BA. The variables analyzed were: fat, density, defatted dry extract, protein and lactose. The first two main components explained 87.24% of the total variation. It was verified the formation of different groups distributed in the four quadrants of the system. First quadrant stood out from the others by forming a group composed of ten producers in the analyzed region, characterized by presenting samples with higher lactose content and lower fat content in milk. The lactose and fat variables are of greater importance in the characterization of milk.

Keywords: Methods; Principal components; Producers; Animal origin matrix.

Resumen

El análisis multivariado implica la aplicación de métodos estadísticos y computacionales para predecir respuestas. Entre los diversos métodos de análisis estadístico multivariante, se destaca el análisis por componentes principales para predecir la composición y calidad de los alimentos en general. El objetivo de este estudio fue caracterizar a los productores de leche en el municipio de Itapetinga-BA, utilizando el análisis de componentes principales. Se utilizaron veinte muestras de leche cruda, recolectadas en la recepción de la lechería ubicada en Itapetinga-BA. Las variables analizadas fueron: grasa, densidad, extracto seco desgrasado, proteínas y lactosa. Los primeros dos componentes principales explicaron el 87.24% de la variación total. Se verificó la formación de diferentes grupos distribuidos en los cuatro cuadrantes del sistema. El cuadrante I se destacó de los demás por formar un grupo compuesto por diez productores en la región analizada, caracterizado por presentar muestras con un mayor contenido de lactosa y un menor contenido de grasa en la leche. Las variables lactosa y grasa son más importantes en la caracterización de la leche.

Palabras clave: Métodos; Componentes principales; Productores; Matriz de origen animal.

1. Introduction

Multivariate analysis involves the application of statistical and computational methods to predict, reduce, group and classify a set of data of interest. The variables of these data, which must be interrelated, are used simultaneously (Souza et al., 2012). There are several methods to perform the statistical multivariate analysis among which are the analysis by main components and the partial square sums that are widely used in the development of analytical methods for the prediction of composition and quality of food in general (Viana, 2018).

According to Filho et al. (2010), the principal components analysis (PCA) is the chosen option in the analysis of production systems due to its capacity to synthetize large data tables and for indicating the variables responsible for the diversity of production systems.

Milk is a whitish secretion produced by the mammary glands of mammalian females; whose natural function is to feed the chicks at an early age. Some milk constituents are produced in secretory cells and others from blood (Guetouache et al., 2014).

Physicochemical parameters may change with the addition of substances in milk, then the study of these characteristics is important to determine milk quality (Conceição, 2018). In fact, the chemical composition of milk may vary according to the type of animal species and genetics, environmental conditions, lactation stage and nutritional status of the animals. As

examples, bovine milk is composed of water (87%), lactose (4% to 5%), protein (3%), fat (3% to 4%), minerals (0.8%) of and vitamins (0.1%) (Pereira, 2014).

In this sense, lactose is a disaccharide characteristic of milk, composed of glucose and galactose, and its synthesis, which requires α-lactalbumin milk protein, is directly related to the amount of milk produced, as it is responsible for transporting blood water to the mammary glands (Lucey et al., 2017).

Bovine milk contains about 200 different proteins and can be distinguished in two main fractions: caseins (αs1-casein, αs2-casein, β-casein and k-casein), which represent 80% of bovine milk proteins and are arranged in mycelial complexes conferring their milky appearance, and serum proteins (α-lactalbumin, β-lactoglobulin and albumin), which represent 20% of total bovine milk (D'AURIA et al., 2018). In addition, the milk serum (also known as whey) contains lactoferrin, immunoglobulins, glycoproteins and enzymes (Abbring et al., 2019; Maqsood et al., 2019).

Fat in milk occurs as small globules containing mainly triacylglycerols, surrounded by a membrane of complex structure composed of various components, such as proteins, glycoproteins, enzymes and lipids (Zhao et al., 2019). Triglycerides form 98% of the fat fraction of milk and the remaining 2% are monoacylglycerol, diacylglycerols, phospholipids, free fatty acids and cholesterol. In addition, the lipid fraction of milk is the most complex of all-natural fats, considering that triacylglycerols are formed by approximately 400 fatty acid esters (FAEs) (Pereira, 2014). Furthermore, milk fat consists of 70% saturated fatty acids and 30% unsaturated fatty acids. Among the saturated, the most important from a quantitative point of view are palmitic (30%), myristic (11%) and stearic (12%). Then, the unsaturated fraction, oleic acid is present in concentrations between 24% and 35%, while polyunsaturated acids constitute about 1.6% linoleic and 0.7% α-linolenic and trans fatty acids such as vaccenic acid 2.7% and conjugated linoleic acid 0.34% - 1.37% (Meena et al., 2019).

The profile of milk vitamins includes the fat soluble vitamins (A, D, E), associated with fat globules, and hydrosoluble vitamins (complex B and vitamin C). Milk can certainly be distinguished by its richness in B vitamins, contributing to daily intake of vitamins B6 and B2 indispensable to the normality of organic functions (Schmidt et al., 2017). In its mineral fraction, milk is recognized as a rich source of calcium, in addition to other elements such as phosphorus, magnesium, zinc and selenium (Pereira, 2014).

This work was conducted with the objective of employing the principal component analysis (PCA) to gather the most homogeneous form of milk producers in the city of Itapetinga-BA, Brazil, regarding the similarity of the physicochemical characteristics of milk

received by the industry, and to identify which characteristics most explain the differences between production systems.

2. Material and Methods

For this research, 20 milk samples were obtained from producers in the city of Itapetinga-BA, Brazil. They were acquired at the reception of a dairy located in the city, being coded and stored under refrigeration until the time of analysis. The physicochemical analysis of the samples were performed at the Laboratory of sheep culture of the State University of Southwest Bahia, Itapetinga-BA campus.

An ultrasound equipment (Lactoscan LA, Milkotronic Ltd, Bulgaria) was used using the following physicochemical parameters: fat (G), protein (PTN), defatted dry extract (ESD), lactose (LAC) and density (D).

The Statistical Analysis System (SAS) program , version 5.1, was used to perform the multivariate statistical analysis.

Data treatment was performed from the determinations of the dispersion measures of the variables, namely standard deviation and coefficient of variation. Based on this result, the data were standardized, which consisted of the insertion of a new variable with mean and standard deviation eliminating the differences between the units of measurements of the studied variables. The data was standardized in Microsoft Office Excel 2010®.

From the standardized data set, the principal component analysis was performed, resulting in an ordering diagram according to the similarity of the variables considered. In order to choose the number of principal components, the method proposed by Cattel (1996) was adopted, which suggests that the amplitude of the self-values be represented graphically according to the number of self-values. The selection of the number of retained components was based on the breakpoint of the chart, when there is a sharp drop in the amplitude of the autovalues.

After determining the number of main components, the scores for each main component were estimated. The software OriginPro8 was used to improve the visualization of the graphs of the correlation matrix autovalues and the dispersion plot of the samples.

3. Results and Discussion

Table 1 shows the results of the self-values, as well as the percentage of variance explained by each component and the cumulative variance of them.

Table 1 - Autovalues and explanation percentage of total variance.

Components	Autovalue	% of Variance	% Cumulative
1	3.3224	66,45	66,45
2	1.0396	20,79	87,24
3	0.5064	10,13	97,37

Source: Authors

As can be seen in Table 1, as well as can be graphically visualized in Figure 1, using the Catter method, the first two principal components (PC) were chosen, with an explanatory power of 87.24% of the total variance, demonstrating that the main component technique was effective to summarize the number of characteristics responsible for defining the groups. This situation presents a decrease in the characterization work as a consequence, with improvement in precision, besides making less complex the analysis and interpretation of the data.

Figure 1 - Self-values of the correlation matrix.
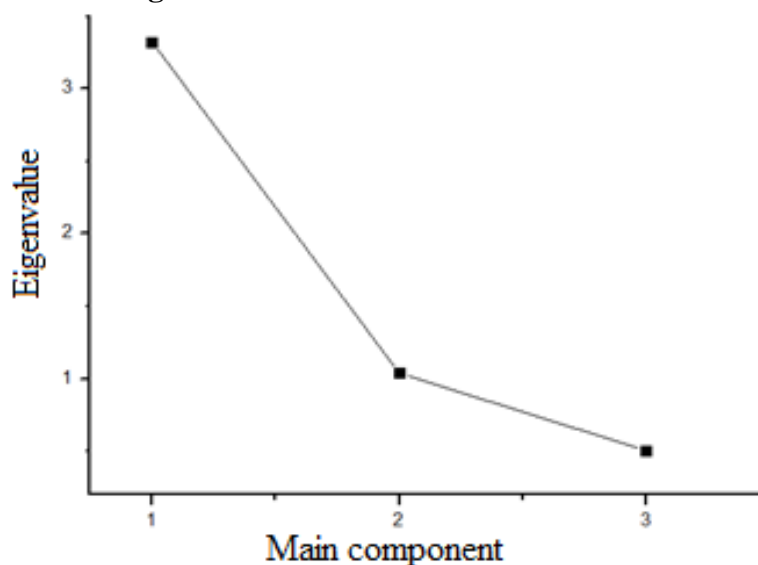
Source: Authors

Table 2 shows the matrix of the weights, in which can be verified the variables that best correlate with each component. The first principal component (PC1 - 66.45%) presented a positive correlation with the variables G, D, ESD, PTN and LAC, and lactose obtained a

lower contribution. The second principal component (PC2 - 20.79%) presented a negative correlation only with variable G, presenting a positive correlation for the other variables, especially lactose with high correlation. PC1 presented the best distribution of the data, with the physicochemical variables with greater importance to determine the characteristics of each studied group.

Table 2 - Correlation between variables and main components.

| Variables | Autovectors | |
	PC1	PC2
G	0,4056	-0,2676
D	0,5052	0,1153
ESD	0,5376	0,0042
PTN	0,5373	0,0025
LAC	0,0487	0,9565

Source: Authors

In regard of the Figure 2, it can be concluded that the variables D, ESD and PTN are the most representative, considering PC1, because they are located at the end of the x axis and, therefore, the farthest from the origin of the Cartesian axis, therefore, has the greatest influence. Among the variables G and LAC there is no correlation, due to the distance formed between them.
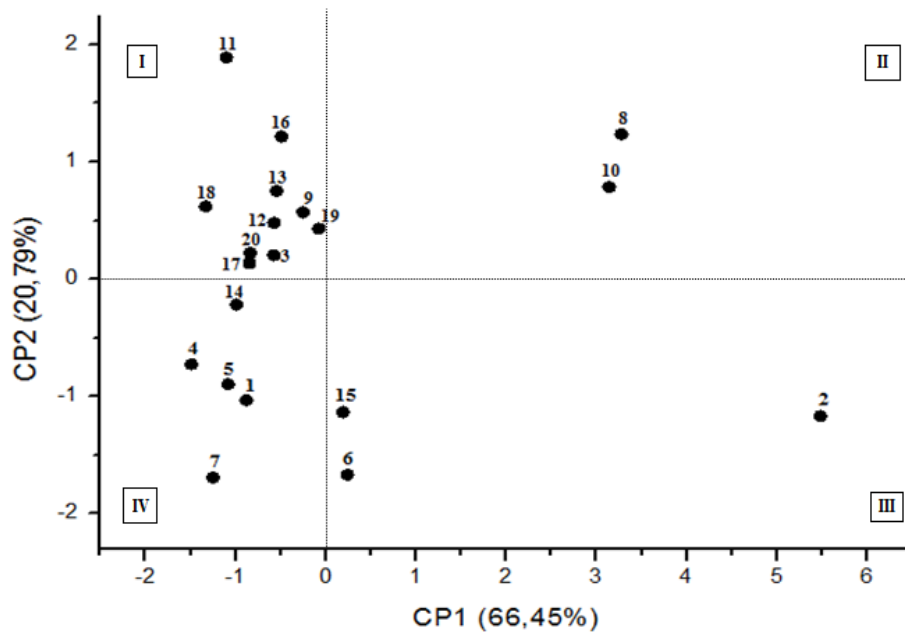
Figure 2 - Score graph considering the two components obtained for milk samples with the variables studied.
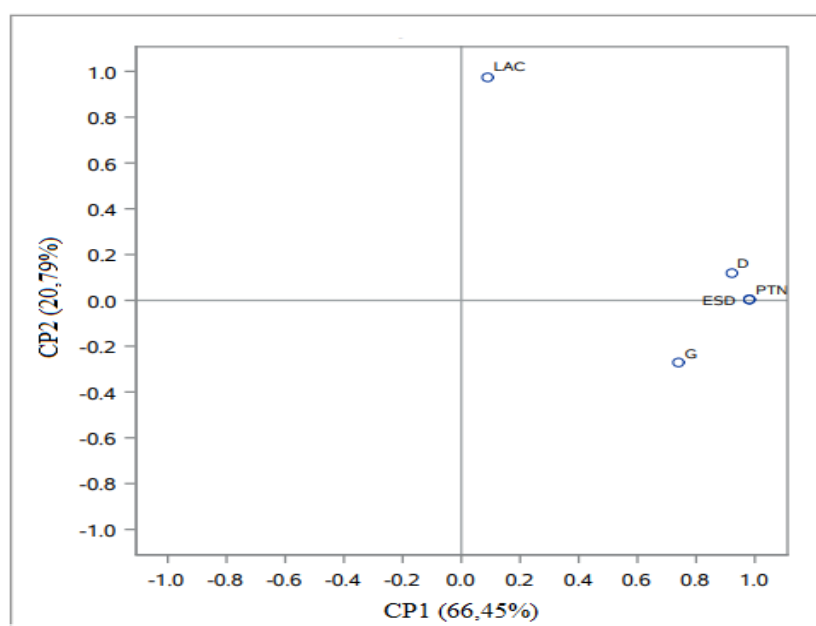
Source: Authors

From the acquisition of PC scores, the dispersion of the mean values of milk samples was analyzed according to the producers (Figure 3).

Figure 3 - Dispersion graph of milk samples in relation to the main components PC1 and PC2.

Source: Authors

The score graphic analysis led us to infer the occurrence of group formation from the set of samples. In first quadrant, it is observable the group with high correlation with PC2, indicating that this group is the producer of highest level of lactosis and lowest level of fat. Besides, samples with positive correlation with PC1 and PC2 can be found in second quadrant, showing the same pattern in lactosis, density, no fat dry extract and protein levels.

In the third quadrant, the samples which exhibit positive correlation with PC1 are found, indicating low level of lactosis in these samples. Also, sample number 2 differed significantly from others by not forming any group. In fourth quadrant, the samples in which negative correlation with PC2 and PC1 are found, showing low content of fat and lower level of lactosis.

4. Conclusion

The Principal Component Analysis allowed the characterization and grouping of milk producers in the city of Itapetinga-BA in different groups. The distribution of samples in the scatterplot facilitated the visualization of samples with higher and lower concentrations of fat, protein, defatted dry extract, lactose and density, in addition to pointing out the tendency of similar data approximation, as well as indicating the influence of lactose variables. and fat in the formation of groups.

The procedure presented in this work has great potential when associating physical-chemical analysis by ultrasound with multivariate tools, as it is a fast and safe process, and can be used in the food industry as an alternative assessment of the physical-chemical characteristics of milk from different producers , aiming to improve its quality. However, qualified labor is required to generate the results and perform the data evaluation.

The expansion of the studies can be carried out with the proposal of increasing samples and expanding the analyzed parameters to improve the knowledge of the composition and quality of milk in the studied region.

References

Abbring, S., Hols, G., Garssen, J., & van Esch, B. C. (2019). Raw cow's milk consumption and allergic diseases–the potential role of bioactive whey proteins. *European journal of pharmacology*, 843, 55-65.

Cattell, R. B. (1996). "The screen test for the number of factors". Multivariate behavioral research, 1(2), 245-276.

Conceição, D. G. (2018). *Utilização do FTIR aliado à análise quimiométrica como ferramenta de triagem para identificação de adulterantes no leite cru* (Masters dissertation, Universidade Estadual do Sudoeste da Bahia).

D'Auria, E., Mameli, C., Piras, C., Cococcioni, L., Urbani, A., Zuccotti, G. V., & Roncada, P. (2018). Precision medicine in cow's milk allergy: proteomics perspectives from allergens to patients. *Journal of proteomics*, 188, 173-180.

Guetouache, M., Guessas, B., & Medjekal, S. (2014). Composition and nutritional value of raw milk. *Journal Issues ISSN*, 2350, 1588.

Lucey, J. A., Otter, D., & Horne, D. S. (2017). A 100-year review: Progress on the chemistry of milk and its components. *Journal of Dairy Science*, 100(12), 9916-9932.

Maqsood, S., Al-Dowaila, A., Mudgil, P., Kamal, H., Jobe, B., & Hassan, H. M. (2019). Comparative characterization of protein and lipid fractions from camel and cow milk, their functionality, antioxidant and antihypertensive properties upon simulated gastro-intestinal digestion. *Food chemistry*, 279, 328-338.

Meena, S., Rajput, Y. S., Sharma, R., & Singh, R. (2019). Effect of goat and camel milk vis a vis cow milk on cholesterol homeostasis in hypercholesterolemic rats. *Small Ruminant Research*, 171, 8-12.

Pereira, P. C. (2014). Milk nutritional composition and its role in human health. *Nutrition*, 30(6), 619-627.

Schmidt, A., Schreiner, M. G., & Mayer, H. K. (2017). Rapid determination of the various native forms of vitamin B6 and B2 in cow's milk using ultra-high performance liquid chromatography. *Journal of Chromatography* A, 1500, 89-95.

Souza, A. M. D., & Poppi, R. J. (2012). Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. *Química Nova*, 35(1), 223-229.

Viana, C. C. R. (2018). *Caracterização de fórmulas infantis para lactantes usando espectroscopia no infravermelho médio*. (Masters dissertation, Universidade Federal de Juiz de Fora).

Zhao, L., Du, M., Gao, J., Zhan, B., & Mao, X. (2019). Label-free quantitative proteomic analysis of milk fat globule membrane proteins of yak and cow and identification of proteins associated with glucose and lipid metabolism. *Food Chemistry*, 275, 59-68.

Porcentagem de contribuição de cada autor no manuscrito

Lenara Oliveira Pinheiro – 30%

Mário Roberto Júnior – 15%

Clara Mariana Gonçalves Lima – 15%

Heliara Caires Sousa – 10%

Jorge Pamplona Pagnossa – 10%

Leandro Soares Santos – 10%

Sérgio Augusto de Albuquerque Fernandes – 10%

Porcentagem de contribuição de cada autor no manuscrito