Formation of homogeneous groups of bovine milk production units via multivariate statistical techniques

Diego Prado de Vargas[1] ⓘ **José Laerte Nörnberg[2]** ⓘ **Renius de Oliveira Mello[3]** ⓘ
Rudolf Brand Scheibler[4] ⓘ **Fernanda Cristina Breda Mello[3]**
Fábio Antunes Rizzo[4] ⓘ **Tatiana Kátia Bertussi Bellato[5]**

[1]Departamento de Biologia e Farmácia, Universidade de Santa Cruz do Sul (UNISC), Avenida Independência, 2293, 96815-900, Santa Cruz do Sul, RS, Brasil. E-mail: diegodevargas@hotmail.com. *Corresponding author.
[2]Departamento de Tecnologia e Ciência dos Alimentos, Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.
[3]Departamento de Zootecnia, Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.
[4]Departamento de Zootecnia, Universidade Federal de Pelotas (UFPel), Pelotas, RS, Brasil.
[5]Universidade do Oeste de Santa Catarina (UNOESC), São Miguel do Oeste, SC, Brasil.

ABSTRACT: *Milk supply chain in Brazil exhibits significant production system heterogeneity in all federal units. Thus, the objective of this study was to form homogeneous groups of bovine milk production units based on the chemical and microbiological quality of the milk via multivariate statistical techniques. A total of 1,541 milk producing units (MPUs), corresponding to 44,089 samples, were analyzed. The first three principal components accounted for 81.38% of the total variation in the data. Principal component 1 (PC1) was associated with the chemical quality of milk (fat, protein [PROT] and total dry extract [TDE] content), while PC2 and PC3 were associated with microbiological quality (somatic cell count [SCC] and total bacterial count [TBC]). The concurrent analysis of the two two-dimensional projections characterized the different productive strata by their quality attributes and identified the positive/negative points of milk microbiological characteristics in each production group. Thus, the dimensionality of the set of 1,541 MPUs was reduced to 15 homogeneous production groups. This method optimizes the use of the dairy industry monthly database and characterizes all the heterogeneities present in dairy production systems.*
Key words: *fat, lactose, protein, total dry extract, typology.*

Formação de grupos homogêneos de unidades de produção de leite bovino através de técnicas estatísticas multivariadas

RESUMO: *A cadeia produtiva brasileira de leite possui expressiva heterogeneidade de sistemas de produção em todas as unidades federativas. Assim, objetivou-se formar grupos homogêneos de unidades de produção de leite bovino através de técnicas estatísticas multivariadas, com base na qualidade química e microbiológica do leite. Foram utilizadas 1.541 unidades produtoras de leite (UPL), totalizando 44.089 amostras analisadas. Os três primeiros componentes principais explicaram 81,38% da variação total dos dados. O componente principal 1 associou-se à qualidade química do leite (gordura, proteína e extrato seco total), enquanto os componentes principais 2 e 3 com a qualidade microbiológica (contagem de células somáticas e contagem bacteriana total). Com a análise conjunta das três projeções bidimensionais, caracterizaram-se os distintos estratos produtivos quanto aos seus atributos de qualidade e identificaram-se os pontos positivos/negativos das características microbiológicas do leite de cada um dos grupos de produção. Assim, obteve-se uma redução da dimensionalidade do conjunto de 1.541 UPL em 15 grupos de produção homogêneos, otimizando a utilização da base de informações mensais das indústrias lácteas, e caracterizando a totalidade das heterogeneidades presentes em sistemas de produção leiteiros.*
Palavras-chave: *gordura, lactose, proteína, extrato seco total, tipologia.*

INTRODUCTION

Brazil, a major global producer of milk, is ranked fourth in world production and produces 34.1 million tons of milk per year (FAO, 2013). Considering the growth in the demand for milk in the foreign market and the potential of Brazil to meet a large part of this demand, achieving the internationally required milk quality standard is extremely important.

However, the Brazilian supply chain has significant heterogeneity in its production systems in all federal units; thus, meeting the current demand for milk has been problematic (ALEIXO et al., 2007). Awareness of the heterogeneity of these systems is becoming increasingly important for effective communication with rural producers and for improvement of the quality of national milk (HOSTIOU and DEDIEU, 2012).

To assist producers with the pricing of milk and direct technical assistance at rural property level, dairy industry has constructed a database of the monthly collections of the physical-chemical and microbiological characteristics of milk. In terms of milk composition, the following two main aspects are considered: the centesimal composition, which includes the fat, protein (PROT), lactose (LACT), total dry extract (TDE) and defatted dry extract (DDE) content, and the hygienic-sanitary component, which includes the somatic cell count (SCC) and total bacterial count (TBC). However, according to BODENMÜLLER FILHO et al. (2010), the appropriate use of the monthly collection database requires analysis tools to simplify the use of the database. Furthermore, according to ALEIXO et al. (2007), multivariate data analysis techniques such as principal component analysis (PCA) combined with cluster analysis (AAG) are statistical tools that could be potentially used to resolve these problems.

The analysis of milk production groups and milk typology become indispensable tools for the dairy industry. However, interpretation of the data is difficult for rural producers. Providing more specific technical assistance for data analysis at this level could improve the quality of this raw material. The objective of this study was to form homogeneous groups of bovine milk production units based on the chemical and microbiological quality of milk via multivariate statistical techniques.

MATERIALS AND METHODS

In total, 1,706 milk producing units (MPUs) were collected on a monthly basis from June 2008 to December 2011. Samples were analyzed to determine their fat, PROT, LACT, DDE and TDE content and SCC and TBC, which resulted in the analysis of 54,696 records. To reflect the particular uniformity of each MPU, results concerning the collective expansion tanks were excluded from the statistical analyses. Records were considered as monthly classes, and properties with less than four controls and with three standard deviations above or below the mean in the month were excluded. After the original database was edited, 44,089 records of 1,541MPUs were used in the statistical analyses. These data were from 15 municipalities in the east-central mesoregion of Rio Grande do Sul, which belongs to the microregion of Lajeado-Estrela.

The SCC data were transformed into somatic cell linear scores using the following equation: (SCLS) = $[\log_2(SCC/100)]+3$ (SHOOK, 1993). The TBC variable was defined as the natural logarithm of the initial TBC.

Subsequently, the multivariate analysis of variance (MANOVA) was performed by the general linear model (GLM) procedure and MANOVA command (SAS, 2002) according to the following statistical model: $Y_{ijk} = \mu_k + H_{ik} + e_{ijk}$ in which Y_{ijk} = the observed value of the variable within the k-th MPU and i-th replicate; μ_k = the overall mean of the k-th variable; H_{ik} = the fixed effect of the i-th MPU in the k-th variable; and e_{ijk} = the random effect associated with observation Y_{ijk}.

Because of the correlation between DDE and the other variables analyzed, DDE was excluded from the statistical model. In the multivariate analysis used to test the hypothesis asserting the means of treatment (MPU) vectors were zero, H_0: $\mu_1 = \mu_2$ $K = \mu_{1541}$, the Wilks test was performed using the following equation:

$$\Lambda = \frac{|E|}{|A|}$$

where is the determinant of the matrix E that refers to the sum of the squares and residual products and is the determinant of the matrix A that refers to the sum of the squares and total products. Afterward, the PRINCOMP procedure was used to perform PCA (SAS, 2002).

Subsequently, the first three dimensions of the principal components were used to group the milk properties according to their similarities (cluster analysis). The number of homogeneous MPU groups was obtained using the co-expressed correlation coefficient (CCC), pseudo-F and pseudo-t^2, which were expressed in relation to the first three dimensions of the principal components. The procedure PROC CLUSTER was used to perform the cluster analysis (SAS, 2002).

The SAS® System for Windows™ version 9.0 (SAS Institute Inc., Cary, NC, USA) was used to perform the statistical analyses.

RESULTS AND DISCUSSION

The Wilks test showed a significant difference (P <0.05) in the mean vectors of the different MPUs upon performing multivariate analysis of the data. The first three principal components explained 81.38% of the total variation in the data: principal component 1 (PC1), PC2 and PC3 explained 38.66%, 25.02% and 17.70% of the data, respectively. In the three-dimensional space of the principal components, production units were grouped by similarity, permitting the reduction of dimensionality from 1,541 to 15MPUs, which were

represented by three two-dimensional graphs (PC1 x PC2, PC1 x PC3 and PC2 x PC3) (Figure 1).

PC1 and PC2 exhibited a cumulative variability of 63.68%, demonstrating a smaller loss of information than the results in the literature, including those reported by ALEIXO et al. (2007) and BODENMÜLLER FILHO et al. (2010). These authors reported cumulative variabilities of 45.00%, 45.70% and 56.51%, respectively, for the first two principal components.

However, the dimensional set used must contain at least 70% of the total variance of the data. Thus, in this study, considerations were in relation to the first three principal components.

In addition, through visual inspection and according to SMITH et al. (2002), the angle between the vectors of the loads explains the correlation between the variables: if this angle is near zero, the correlation is positive; if this angle is near 180°, the correlation will be negative; and, finally, if this angle is near 90°, these variables are mostly unrelated.

The first two-dimensional representation provided evidence that the SCC and TBC are directly correlated with each other and, in turn, are inversely correlated with the LACT content (Figure 1). Likewise, as the values of SCC and TBC increased, VARGAS et al. (2014) and VARGAS et al. (2013), respectively, observed the same behavior for the LACT variable. The SCC and TBC are indicative of the hygienic-sanitary quality of milk. The high negative correlation between these microbiological indicators and LACT might be due to the decreased synthesis of this constituent due to alterations in the mammary gland as well as the decrease in LACT due to its absorption into the bloodstream. In addition, the use of this carbohydrate by breast pathogens might reduce the LACT content in milk (VARGAS et al., 2013; VARGAS et al., 2014). Therefore, in addition to the SCC and TBC, the LACT content could function as a variable indicative of the sanitary quality of milk.

Interpretation of the principal components was performed by determining the correlations between the variables and the components. Thus, in the dimensions of the Cartesian plane, the variables that explained the variability on the x-axis (PC1) were fat (r=0.84, P <0.001), PROT (r=0.75, P <0.001) and TDE (r=0.99, P <0.001). Variables that explained the variability on the y-axis (PC2) included LACT (r=-0.86, P <0.001), the SCC (r=0.50, P <0.001) and the TBC (r=0.62; P <0.001). Variables that explained the variability on the z-axis (PC3) were the SCC (r=0.78, P <0.001) and TBC (r=-0.67, P <0.001). Therefore, the existing correlations showed the first principal component (x-axis) was linked to the chemical quality of the milk and the second (y-axis) and third (z-axis) principal components were linked to microbiological quality (Figure 1).

Furthermore, the first (PC1 x PC2) and second Cartesian planes (PC1 x PC3), which described the variability of the data in further detail (63.68% and 56.36%, respectively), enabled the placement of the clusters in relation to the different quality parameters of the milk. Thus, in both bi-dimensional projections, groups 4, 6, 7, 8, 10, 12, 13 and 14 were located in quadrants 1 and 4 and 1, 2, 3, 5, 9, 11 and 15 were located in quadrants 2 and 3. Therefore, the first and second Cartesian planes exhibited the highest and lowest values, respectively, of fat, PROT and TDE (Figure 1). However, considering averages of these variables, none of the groups formed were in disagreement with the regulatory standards for chemical quality of milk, established by the Ministry of Agriculture, Livestock and Food Supply (MAPA) Normative Instruction No. 62 (IN 62, BRASIL, 2011) (Table 1).

With regard to microbiological quality indicators, the differential correlations between the TBC variable and PC2 and PC3 were highlighted. TBC and PC2 were directly correlated (r=0.62, P <0.001), and TBC and PC3 were inversely correlated (r=-0.67, P <0.001). Thus, the positive/negative points of the microbiological quality of the productive strata were identified via the joint analysis of the two-dimensional planes PC1 x PC2 and PC1 x PC3, which resulted in technical assistance more specific than that from using the first Cartesian plane decision-making method.

In PC1 x PC2, the clusters located in the first (4, 6 and 10) and second quadrants (1, 2, 5 and 11) exhibited low microbiological quality. However, when the data was arranged in PC1 x PC3, these groups were distinguished with regard to the negativity of hygienic-sanitary standards. Groups 1, 5, 6 and 10 were considered inferior due to their high SCC (quadrants 1 and 2), while groups 2, 4 and 11 were considered inferior due to their high TBC values (quadrants 3 and 4). Thus, to obtain the desired improvements, the groups that presented high SCC values (1, 5, 6 and 10) at the producer level should have this parameter monitored periodically and continuously in individual cows to assist in the identification of animals responsible for the high counts of the expansion tank and, thus, more appropriately to direct the actions of the producers. In this sense, according to MAIA et al. (2013), the development of microbiological cultures of milk would allow the identification of pathogens that cause mastitis and, consequently, aid in determining treatment strategies.
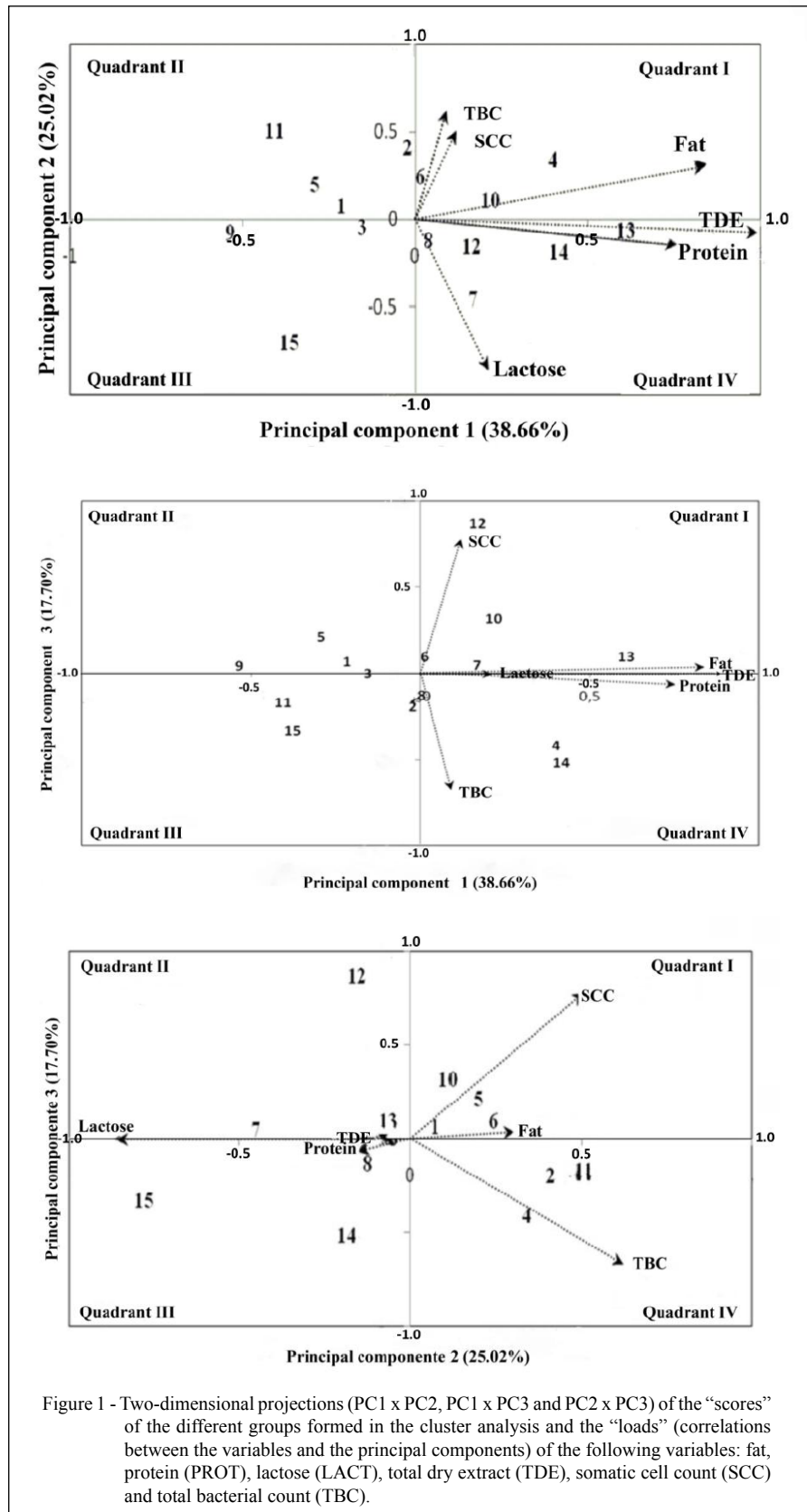
Figure 1 - Two-dimensional projections (PC1 x PC2, PC1 x PC3 and PC2 x PC3) of the "scores" of the different groups formed in the cluster analysis and the "loads" (correlations between the variables and the principal components) of the following variables: fat, protein (PROT), lactose (LACT), total dry extract (TDE), somatic cell count (SCC) and total bacterial count (TBC).

Table 1 - Number of producers (No. of producers); the mean fat, protein (PROT), lactose (LACT), mineral, total dry extract (TDE), defatted dry extract (DDE), somatic cell count (SCC) and total bacterial count (TBC) values; and the respective standard deviation and coefficient of variation (CV) of the groups formed using cluster analysis as a function of their different quadrants.

Group	No. of producers	Fat (%)	PROT (%)	LACT (%)	Minerals (%)	TDE (%)	DDE (%)	SCC (cells mL^{-1})[1]	TBC (UFC mL^{-1})[2]
1	127	3.46	3.01	4.32	0.94	11.74	8.27	579.000	4,426.712
2	74	3.73	3.07	4.22	0.97	11.99	8.26	792.000	5,473.356
3	74	3.48	3.04	4.41	0.95	11.87	8.40	595.000	1,239.746
4	173	3.96	3.23	4.29	1.00	12.48	8.52	894.000	4,688.176
5	70	3.38	3.00	4.33	0.94	11.65	8.26	889.000	725.000
6	203	3.61	3.08	4.32	0.97	11.98	8.37	897.000	2,580.787
7	43	3.50	3.17	4.52	0.97	12.16	8.66	385.000	2,107.026
8	144	3.55	3.10	4.37	0.97	12.00	8.44	433.000	4,623.804
9	113	3.17	2.96	4.34	0.93	11.39	8.22	547.000	2,663.781
10	89	3.62	3.15	4.40	0.98	12.15	8.53	945.000	819.000
11	123	3.51	2.96	4.18	0.94	11.58	8.08	968.000	4,370.221
12	105	3.45	3.09	4.46	0.96	11.96	8.51	628.000	528.000
13	95	3.88	3.27	4.44	1.02	12.60	8.73	677.000	1,786.089
14	73	3.88	3.23	4.44	1.00	12.55	8.67	423.000	4,114.262
15	35	3.05	3.10	4.46	0.96	11.57	8.51	184.000	3,909.601
Mean	-	3.55	3.10	4.37	0.97	11.98	8.43	655.93	2,799,042.18
Standard deviation	-	0.25	0.10	0.09	0.03	0.37	0.19	0.02	0.03
CV	-	7.05	3.14	2.15	2.64	3.06	2.21	5.17	7.56
Total observations	1541	44.089	44.089	44.089	44.09	44.089	44.089	44.089	44.089

[1]Data de-transformed from the somatic cell linear score (SCLS = [log$_2$ (SCC / 100)] + 3). [2]Data de-transformed from the natural logarithm of the normal TBC. [3]CV

Due to their high TBC values, productive groups 2, 4 and 11 should result in increased care concerning milk contamination by the resident microbiota outside the udder. WINCK and THALER NETO (2012) showed cleaning udders before milking affects the TBC: the producers who pre-immersed teats in disinfectants exhibited better results. In addition, the study showed care is required regarding personal hygiene, milker training and water, which is a potential contaminant source for milk because of its importance in milking activities.

The groups located in quadrants 3 (3, 9 and 15) and 4 (7, 8, 12, 14 and 13) in the two-dimensional projection PC1 x PC2 demonstrated high microbiological quality. However, the following groups were even more distinctive when placed in the PC1 x PC3 projection: groups 8, 14 and 15 (quadrants 3 and 4) exhibited low SCC values and groups 3, 7, 9, 12 and 13 (quadrants 1 and 2) exhibited a desirable TBC. Among these groups, when evaluating the two-dimensional projection PC2 x PC3, which explained 42.72% of the variability in the data and described only the microbiological quality of the milk, groups 12 and 15 differed more

in relation to other strata. Group 12, comprising 105 MPUs, exhibited the lowest values of TBC (higher scores in relation to PC3), and group 15, comprising 35 MPUs, exhibited the lowest values of SCC (lower scores in relation to PC2). Therefore, these productive strata could be used as a positive reference in the technical assistance provided to the milk producer in order to alleviate the high SCC and TBC observed in other units.

Notably, all groups failed with respect to the SCC limit (400,000 cells mL^{-1}) and/or the TBC limit (100,000 colony-forming units [CFU] mL^{-1}), proposed by the last stratification of IN 62, which was implemented on July 30, 2016. Furthermore, six (2, 4, 5, 6, 10 and 11) and 13 groups (2, 3, 4, 6, 7, 8, 9, 10, 11, 13, 14 and 15) exceeded the SCC and TBC limits proposed by extinct IN 51, which is now extinct, of 750,000cells mL^{-1} and 750,000CFU mL^{-1}, respectively (Table 1).

Among the major obstacles to increasing the Brazilian export of dairy products are those related to sanitary and sanitary embargo. Thus, at the industrial level, a review of the SCC and TBC

quality payment system is needed. In addition, at the producer level, an increase in educational actions with the purpose of increasing knowledge and awareness of issues aiming to improve milk quality is needed.

In this sense, at an industrial level, the use of reward and penalty systems to identify the main problems of raw material quality is possible with assistance from the data gleaned from the groupings placement in the Cartesian plane. This method would facilitate the logistics of capturing milk with quality chemical and microbiological characteristics, representing a dilution in production costs, as these indicators are directly related to the industrial yield.

CONCLUSION

Multivariate statistical techniques were used to form 15 homogeneous groups based on the chemical and microbiological data of the bovine milk of 1,541 production units. These data were obtained from 15 municipalities in the central eastern mesoregion of Rio Grande do Sul, Brazil which belongs to the microregion of Lajeado-Estrela. The first three principal components accounted for 81.38% of the total variation in the data. The two-dimensional representations showed that the SCC and TBC were directly correlated with each other and inversely correlated with the LACT content. In addition, the different productive strata were characterized by their quality attributes, and the positive and negative microbiological characteristics of the milk were identified. In total, 105MPUs showed low TBC values, and 35MPUs showed low SCC values. Thus, the multivariate statistical techniques used herein were used in the generation of hypotheses concerning heterogeneous groups. Variation was reduced by categorizing data into similar groups using the qualitative variables of milk, and this type of information can aid in the technical assistance of rural producers.

ACKNOWLEDGMENTS

REFERENCES

ALEIXO, S.S. et al. Multivariate analysis that can be used to determine of dairy producers homogeneous groups. **Revista Brasileira de Zootecnia**, v.36, n.6, p.2168-2175, 2007 (supl.). Available from: <http://dx.doi.org/10.1590/S1516-35982007000900029>. Accessed: mai. 20, 2017. doi: 10.1590/S1516-35982007000900029.

BODEN MÜLLER FILHO, A. et al. Typology of production systems based on the milk characteristics. **Revista. Brasileira de.**

Zootecnia, v.39, n.8, p.1832-1839, 2010. Available from: <http://dx.doi.org/10.1590/S1516-35982010000800028>. Accessed: jul. 20, 2016. doi: 10.1590/S1516-35982010000800028.

BRASIL. **Instrução Normativa n. 51 de 18 de setembro de 2002**. Dispões sobre regulamentos técnicos aplicados ao leite cru refrigerado e pasteurizado. Diário Oficial da União, Brasília, 20 set. 2002. Seção 1, n. 183, p.13-22.

BRASIL. **Instrução Normativa n. 62 de 29 de dezembro de 2011**. Dispõe sobre regulamentos técnicos de produção, identidade e qualidade do leite tipo A, leite Cru refrigerado, leite pasteurizado e do regulamento técnico de coleta de leite cru refrigerado e seu transporte a granel.

BUENO, V. F. F. et al. Somatic cell count: relationship to milk composition and period of the year in Goiás State, Brazil. **Ciência Rural**, v.35, n.4, p.848-854, 2005. Available from: <http://dx.doi.org/10.1590/S0103-84782005000400016>. Accessed: Jul. 15, 2015. doi: 10.1590/S0103-84782005000400016.

HOSTIOU, N. et al. A method for assessing work productivity and flexibility in livestock farms. **Animal**, v.6, n.5, p.852-862, 2012. Available from: <https://doi.org/10.1017/S1751731111002084>. Accessed: Jan. 12, 2017. doi: 10.1017/S1751731111002084.

MAIA, P. V. et al. *Escherichia coli* J5 vaccination during pre-calving and mastites and milk production of crossbred cows. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.65, n.5, p.1367-1375, 2013. Available from: <http://dx.doi.org/10.1590/S0102-09352013000500014>. Accessed: jun. 28, 2015. doi: 10.1590/S0102-09352013000500014.

SHOOK, G. E. Genetic improvement of mastitis through selection on somatic cell count. **The Veterinary Clinics of North America: Food Animal Practice**, v.9, n.3, p.563-581, 1993. Available from: <https://doi.org/10.1016/S0749-0720(15)30622-8>. Accessed: Aug. 11, 2017. doi: 10.1016/S0749-0720(15)30622-8.

SMITH, R. R. et al. Characterization of dairy productive systems in the Tenth Region of Chile using multivariate analysis. **Agricultura Técnica**, v.62, n.3, p.35-395, 2002. Available from: <http://dx.doi.org/10.4067/S0365-280720020003000004>. Accessed: Aug. 15, 2017. doi: 10.4067/S0365-280720020003000004.

STATISTICAL ANALYSIS SYSTEM - SAS. **The SAS system for windows**. v.9.0 Cary: SAS Institute Inc., 2002.

VARGAS, D. P. et al. Correlations between total bacterial count and quality parameters of milk. **Revista Brasileira de Ciência Veterinária**, v.20, n.4, p.241-247, 2013. Available from: <http://dx.doi.org/10.4322/rbcv.2014.00>. Accessed: Mai. 28, 2017. doi: 10.4322/rbcv.2014.00.

VARGAS, D. P. et al. Correlations between somatic cell count and physical-chemical parameters and microbiology of milk quality. **Ciência Animal Brasileira**, v.15, n.4, p.473-483, 2014. Available from: <http://dx.doi.org/10.1590/1809-6891v15i420637>. Accessed: Mai. 28, 2017. doi: 10.1590/1809-6891v15i420637.

WINCK, C. A. et al. Profile of dairy farms in Santa Catarina State in relation to Normative Instruction 51. **Revista Brasileira de Saúde e Produção Animal**, v.13, n.2, p.296-305, 2012. Available from: <http://dx.doi.org/10.1590/S1519-99402012000200001>. Accessed: Mai. 28, 2017. doi: 10.1590/S1519-99402012000200001.