Multivariate data analysis for semen quality in two frog species

Análise de dados multivariada para qualidade do sêmen para duas espécies de rãs

Análisis de datos multivariados para la calidad del semen en dos especies de ranas

Marcelo Maia Pereira
ORCID: https://orcid.org/0000-0002-1898-2722
Fundação Instituto de Pesca do estado do Rio de Janeiro, Brazil
E-mail: mmaiap2001@yahoo.com.br

Abstract
Despite the growing body of research on frog reproduction and semen quality, studies employing statistical methods and analyses are scarce. The present study aimed to examine the use of exploratory multivariate data analysis for evaluating the seminal quality of frogs, determining a possible hierarchy among frog species based on semen characteristics, and identifying the most effective clustering strategy. Data on the seminal quality of the butter frog (*Leptodactylus ocellatus,* native specie) were collected from seven samples, and for the bullfrog (*Lithobates catesbeianus,* exotic specie), five samples were used, all derived from a literature review. Four distinct clustering strategies were tested: Single Linkage, Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Complete Linkage, and Ward's method. Exploratory multivariate data analysis could effectively evaluate frog seminal quality, allowing the classification of frog species (both native and exotic) based on semen. Among the tested strategies, UPGMA emerged as the most suitable grouping method.
Keywords: Frog farming; Reproduction; Statistics.

Resumo
Apesar dos últimos anos o número de trabalhos de pesquisas na reprodução e qualidade do sêmen de rãs terem aumentado, as pesquisas com informações e técnicas estatísticas são escassos. O trabalho foi realizado com objetivo de verificar o uso da análise exploratória de dados multivariada na avaliação da qualidade seminal de rãs, possibilidade para hierarquia de espécies de rãs através do sêmen e qual estratégia de agrupamento seria a melhor. As informações necessárias sobre o sêmen de rã-manteiga (*Leptodactylus ocellatus,* espécie nativa) foram obtidas de sete amostras e sobre o sêmen de rã-touro (*Lithobates catesbeianus,* espécie exótica) foram obtidas de cinco amostras, todas as informações extraídas de revisão de literatura. As estratégias de agrupamentos foram quatro: Single Linkage (ligação simples), Unweighted Pair Group Method with Arithmetic Mean (não ponderado de agrupamento aos pares – UPGMA), Complete Linkage (ligação completa) e Ward's method (método Ward). O uso da análise exploratória de dados multivariada na avaliação da qualidade seminal de rãs foi possível e possibilitou a hierarquia de espécies de rãs (uma nativa e outra exótica) através do sêmen e a melhor estratégia de agrupamento foi UPGMA.
Palavras-chave: Estatística; Ranicultura; Reprodução.

Resumen
Aunque el número de estudios de investigación sobre la reproducción de ranas y la calidad del semen ha aumentado en los últimos años, la investigación que utiliza información y técnicas estadísticas es escasa. El trabajo se realizó con el objetivo de verificar el uso del análisis exploratorio de datos multivariados en la evaluación de la calidad seminal de las ranas, la posibilidad de jerarquía de especies de ranas a través del semen y qué estrategia de agrupamiento sería la mejor. La información necesaria sobre semen de rana mantequilla (Leptodactylus ocellatus, especie nativa) se obtuvo de siete muestras y sobre semen de rana toro (Lithobates catesbeianus, especie exótica) de cinco muestras, toda información extraída de la revisión de la literatura. Hubo cuatro estrategias de agrupación: vínculo único, método de grupo de pares no ponderados con media aritmética, vínculo completo y método de Ward. El uso del análisis exploratorio de datos multivariados para evaluar la calidad seminal de las ranas fue posible y permitió jerarquizar las especies de ranas (una nativa y otra exótica) a través del semen y la mejor estrategia de agrupamiento fue UPGMA.
Palabras clave: Estadística; Ranicultura; Reproducción.

1. Introduction

Despite the increasing volume of research on frog reproduction and semen quality in recent years (Ribeiro Filho et al., 1998; Pereira et al., 2017; Leal & Pereira 2021; Pereira, 2023), the availability of studies incorporating statistical methods and

analyses in this field remains limited (Pereira et al., 2012; Silva et al., 2017).

Multivariate statistical techniques refer to all statistical methods that analyze multiple measurements (variables) simultaneously for each individual or object under study. The variables should be random and interrelated, so their different effects cannot be significantly interpreted separately (Fernandes & Lima, 1991; Faria et al., 2012; Werncke et al., 2016; Castro et al., 2023).

Among the multivariate techniques, clustering allows for categorizing subjects or objects into exclusionary subgroups. It is an optimization approach aimed at generally maximizing the homogeneity of objects or individuals within groups and heterogeneity between groups (Waquil & Concha-Amin, 2015).

Clustering species based on external morphological traits has been the predominant approach; yet, in amphibian studies, categorizing species by the morphological similarities of their sperm has provided valuable information on enhancing phylogenetic classifications (Santos et al., 2015).

A dendrogram constructed using semen data from subpopulations of *Silurana tropicalis*, analyzed through computer-assisted sperm motility analysis (CASA), enabled the hierarchical classification of the species (Larroze et al., 2014).

Similarly, multivariate techniques were applied to a set of carcass data from Morada Nova sheep, reducing the dimensionality of the set of original variables, pinpointing the most discriminative variables, and quantifying the relationships between biometric and performance traits (Guedes et al., 2018).

Given the above explanations and the applicability of these statistical techniques, this study aims to explore the application of exploratory multivariate data analysis in assessing the seminal quality of frogs, determining a hierarchy among frog species based on semen characteristics, and identifying the most effective clustering strategy.

2. Methodology

The necessary information was sourced from a literature review. A study that evaluated seven semen samples from the butter frog (*Leptodactylus ocellatus*) (Pereira et al., 2013a) and another with five semen samples from the bullfrog (*Lithobates catesbeianus*) (Pereira et al., 2013b) were selected. The butter frog is indigenous to Brazil, whereas the bullfrog is considered an exotic species.

In the reviewed studies, each semen sample was denoted by a unique acronym (EXX), where 'E' represents the species, the first 'X' distinguishes between the butter frog (1) and the bullfrog (2), and the second 'X' sequentially numbers the samples.

The analyzed semen sample variables included volume (VE), color (CE), vigor (VI), motility (ME), and concentration (CCE). Spermatozoa, totaling 50 across samples, were categorized based on morphological traits: normal (SN), major defects (DMA), minor defects (DME), abnormal head (CABD), abnormal tail (CADE), broken tail (CAUF), coiled tail (CAUE), macrocephaly (MAE), microcephaly (MIE), isolated normal head (CIN), proximal droplets (GP), distal droplets (GD), and bent tail (CAUD).

After choosing the variables for processing cluster analysis, three procedures were carried out: data standardization, selection of a similarity coefficient, and determination of the clustering method.

Standardization was applied to both the semen sample data (columns) and the semen quality variables (rows). The Euclidean distance was chosen as the similarity coefficient. Four clustering approaches were evaluated: Single Linkage, Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Complete Linkage, and Ward's method, utilizing STATISTICA (2007) software for all analyses.

3. Results and Discussion

The hierarchical clustering diagram for semen samples from two frog species (bullfrog and butter frog) using four grouping strategies (Single Linkage, UPGMA, Complete Linkage, and Ward's Method) revealed that the UPGMA approach most effectively clustered the samples by species (Figure 1).

Figure 1 - Dendrogram for grouping semen samples from two frog species, employing four clustering strategies: Single Linkage, Unweighted Pair-Group Method with Arithmetic Mean (UPGMA), Complete Linkage, and Ward's method.
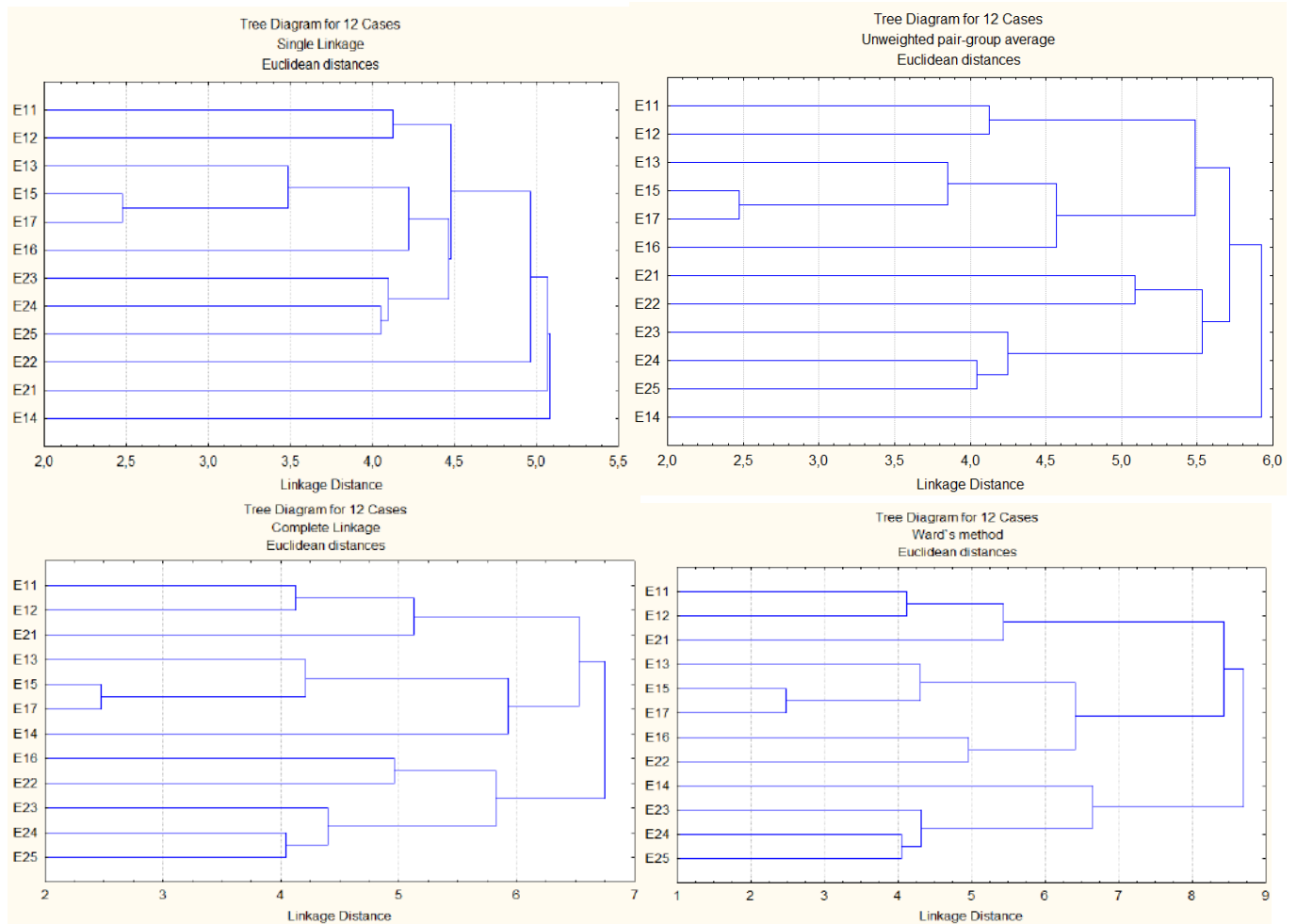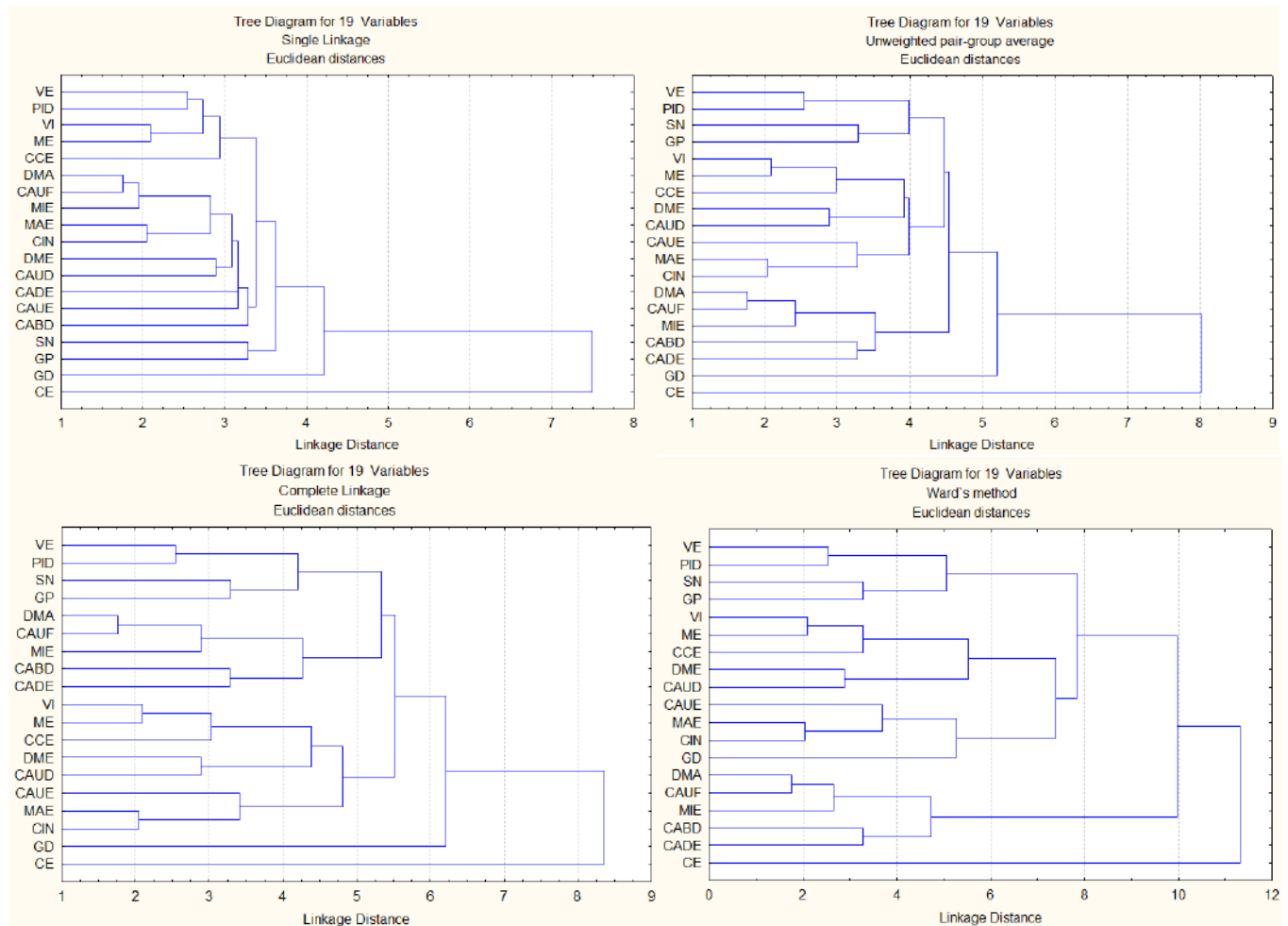

Source: Authors (2024).

By checking Figure 1 it is possible to verify the grouping of species through seminal quality, in addition to discriminating the sample that is different from the others.

The hierarchical clustering diagram for semen quality variables across the two frog species employing the same grouping strategies revealed that UPGMA was the most effective in organizing variable sets within the analyzed datasets (Figure 2).

Figure 2 - Dendrogram for grouping semen quality variables of two frog species, employing four clustering strategies: Single Linkage, Unweighted Pair-Group Method with Arithmetic Mean (UPGMA), Complete Linkage, and Ward's method.

Source: Authors (2024).

The Single Linkage method is characterized by the proximity between the nearest members of different groups. Conversely, the Complete Linkage method considers the distance between the furthest individuals of two groups, thereby addressing the chaining issue inherent to the Single Linkage method (Linden, 2009).

The UPGMA technique calculates the distance between two groups as the mean of all pairwise distances between members of each group, favoring the combination of groups with minimal internal variance or identical variances (Arriel et al., 2006).

Ward's method defines the distance between two clusters as the total sum of squares between the two groups considering all variables, ensuring that at each clustering step, the within-cluster sum of squares is minimized for all possible two-cluster combinations from the previous step. This approach tends to merge clusters with fewer observations and generally results in clusters with the same number of observations (Leal & Catalunha, 2023).

The UPGMA emerged as the most effective strategy in achieving the set goal of hierarchically clustering semen samples from two distinct frog species, namely the exotic bullfrog and the native butter frog, based on a variety of assessed parameters.

The dendrogram of variables served as an additional analysis to identify similarities among parameters, thereby enhancing the interpretability of the hierarchical results of the species. Notably, the most interesting outcome was seen for

'sperm color', which was distant from all other parameters.

This pronounced divergence of the 'sperm color' parameter from the others may stem from its acquisition method as delineated in the methodology section of the study, where semen was classified as '1' for transparent and '2' for cloudy (Pereira et al., 2013a; Pereira et al., 2013b). This distinction highlights the capability of this statistical tool to draw inferences that were previously unattainable.

Beyond clustering, the UPGMA method was also able to detect outlier samples, a critical aspect for informed decision-making. This statistical technique holds considerable potential for application in frog reproduction management, particularly in assessing the semen quality of frog specimens both intra- and inter-species.

Currently, there are no established standards for evaluating seminal quality in amphibians, including frogs, which are also livestock animals, unlike the well-defined standards for many other production species (CBRA, 2013). Thus, this statistical approach offers a promising avenue for seminal quality assessment. Nevertheless, further research and data accumulation are warranted to establish a routine classification framework.

4. Conclusion

The application of exploratory multivariate data analysis in assessing the seminal quality of frogs proved feasible and allowed for the hierarchical classification of frog species (one native and one exotic) based on semen characteristics.

The UPGMA strategy was identified as the most effective clustering method. This technique should be used more for seminal analysis in order to have greater validation.

Acknowledgments

References

Arriel, N. H. C., Di Mauro, A. O., Di Mauro, S. M. Z., Bakke, O. A., Unêda-Trevisoli, S. H., Costa, M. M., Capeloto, A. & Corrado, A.R. (2006). Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD. *Pesquisa Agropecuária Brasileira*, *41*(5), 801–809. 2006. https://doi.org/10.1590/S0100-204X2006000500012

Castro, H. V. M., Shibutani, L. J. S., Severo, R., Baia, A. D. B., Souza, E. B. de & Gama, M. A. S. da. (2024). Pathological and genetic variability of bacterial strains causing vegetable soft rot in west Pará region, Brazil. *Ciência Rural*, *54*(7), e20220221. https://doi.org/10.1590/0103-8478cr20220221

Colégio Brasileiro de Reprodução Animal. (2013). Manual para exame andrológico e avaliação de sêmen animal. (3a ed.), CBRA. http://cbra.org.br/br/publicacoes/manual-de-exame-andrologico/

Faria, P. N., Cecon P. R., Silva, A. R., Finger, F. L., Silva, F. F., Cruz, C. D. & Sávio F. L. (2012). Métodos de agrupamento em estudo de divergência genética de pimentas. *Horticultura Brasileira* 30: 428-432. https://www.scielo.br/j/hb/a/xZLJWnm77Lsfqfd6Q7LZqCp/?format=pdf&lang=pt

Fernandes, T. A. G. & Lima, J. E. (1991). Uso de análise multivariada para identificação de sistemas de produção. *Pesquisa Agropecuária Brasileira*, 26(10), 1823-1836.

Guedes, D. G. P., Ribeiro, M. N. & Carvalho, F. F. R. (2018). Multivariate techniques in the analysis of carcass traits of Morada Nova breed sheep. *Ciência Rural*, 48(9), e20170746.

Larroze, S., Pickford, D.B. & Holt, W. V. (2014) Validation of computer-assisted sperm-motility analysis in the amphibian *Silurana tropicalis*. *Reproduction, Fertility and Development*, http://dx.doi.org/10.1071/RD14015.

Leal, M. S. & Pereira, M. M. (2021). Ciclo anual reprodutivo de rãs-touro (*Lithobates catesbeianus*) no Estado do Rio de Janeiro: nual cycle of reproduction of bullfrogs in the state of Rio de Janeiro. *Revista Brasileira De Agropecuária Sustentável*, *11*(1), 14–21 https://doi.org/10.21206/rbas.v11i1.10433

Leal, W. E. & Catalunha, M. J. (2023). Aplicação do indicador de salubridade ambiental (ISA) em 208 municípios mineiros. Revista FT. https://doi.org/10.5281/zenodo.8256220

Linden, R. (2009) Técnicas de Agrupamento. *Revista de Sistemas de Informação da FSMA,* (4), 18-36. https://www.fsma.edu.br/si/edicao4/FSMA_SI_2009_2_Tutorial.pdf

Pereira, M. M. (2023). Growth curves of bullfrog (*Aquarana catesbeiana*) breeders in small frog farm. *Research, Society and Development,* 12(10), e80121043426. 10.33448/rsd-v12i10.43426.

Pereira, M. P., Ribeiro Filho, O. P., Zanuncio, J. C., Navarro, R. D., Seixas Filho, J. T. & Ribeiro, C. D. L. (2013a). Evaluation of the semen characteristic after induced spermiation in the bullfrog *Lithobates catesbeianus*. *Acta Scientiarum. Biological Science,* 35(3), 305-310.

Pereira, M. P., Ribeiro Filho, O. P., Zanuncio, J. C., Troni, A. R., Navarro, R. D., & Takamura. A. E. (2013b). Buserelin acetate-induced spermiation in *Leptodactylus ocellatus* and evaluation of semen characteristics. *Acta Scientiarum. Biological Science*, 35(1), 1-7.

Pereira, M. M., Ribeiro Filho, O. P. & Navarro, R. D. (2012). Importância da indução artificial na reprodução de rãs. *Revista Brasileira de Reprodução Animal*, 36, 100-104, http://www.cbra.org.br/pages/publicacoes/rbra/v36n2/pag100-104%20(RB364).pdf

Pereira, M. M., Mello, S. C. R. P., Seixas Filho, J. T., & Silva, W. N. (2017). Qualidade do sêmen de rã-touro em diferentes dosagens de hormônio, tempos de coleta e de resfriamento. Semioses, 12, 35-39. https://doi.org/10.15202/1981996x.2017v12n2p35

Ribeiro Filho, O. P., Lima, S. L., Andrade, D. R. & Seixas Filho, J. T. (1998). Estudo da desova de Rã-touro, *Rana catesbeiana*, mediante indução do acasalamento. *Revista Brasileira Zootecnia*, 27, 216-233. http://www.sbz.org.br/revista/artigos/2086.pdf

Santos, J. S., Introíni, G. O., Veiga-Menoncello, A. C. P., & Recco-Pimentel, S. M. (2015). Ultrastructure variation in the spermatozoa of Pseudopaludicola frogs (Amphibia, Anura, Leptodactylidae), with brief comments on its Phylogenetic relevance. *Journal of Morphology*, 276, 1495–1504.

Silva, W. N., Souza, M. S., Oliveira, E. C. P., Pereira, M. M., Seixas Filho, J. T. & Mello, S. C. R. P. (2017) Técnica de fertilização artificial de anfíbios anuros. *Semioses*, 12, 40-47. https://doi.org/10.15202/1981996x.2017v12n2p40

Waquil, P.A. & Concha-Amin, M. (2015). A construção de tipologias: uma aplicação das análises fatorial e de agrupamentos. In: Radomsky, G.F.W, Conterato, M.A., Schneider, S. Pesquisa em desenvolvimento rural: técnicas, base de dados e estatística aplicadas aos estudos rurais – volume 2. Editora da UFRGS, 147-164p.

Werncke, D., Gabbi, A. M., Abreu, A. S., Felipus, N. C., Machado, N. L., Cardoso, L. L., Schmid, F. A., Alessio, D. R. M., Fischer, V., & Thaler Neto, A. (2016). Qualidade do leite e perfil das propriedades leiteiras no sul de Santa Catarina: abordagem multivariada. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, *68*(2), 506–516.