

PROPOSTA DE APLICAÇÃO DE MODELO DE ANÁLISE DISCRIMINANTE PARA CLASSIFICAÇÃO DO LEITE PELA QUALIDADE

CARLA ADRIANA PIZARRO SCHMIDT^{1*}, GENILSO GOMES DE PROENÇA²;
TÁSSIO DE MORAIS GARCIA³; JOSÉ AIRTON AZEVEDO DOS SANTOS⁴; CELEIDE PEREIRA⁵

¹Engenheira Agrônoma Dr^a. Agronomia, Professora da UTFPR, Medianeira-PR, carlaschmidt@utfpr.edu.br

²Técnico Agrícola, Tecnólogo em Agronegócios, Mestrando do programa de Tecnologias Computacionais Para o Agronegócio, UTFPR, Campus Medianeira, genilsogp@gmail.com

⁵Aluno do curso de Engenharia de Produção, UTFPR, Medianeira-PR, tassiomgarcia@gmail.com

⁴Dr. Engenharia Elétrica, Professor da UTFPR, Medianeira-PR, airton@utfpr.edu.br

⁵Dr^a em Ciência dos Alimentos, Professora da UTFPR, Medianeira-PR, celeidep@yahoo.com.br

Apresentado no

Congresso Técnico Científico da Engenharia e da Agronomia – CONTECC'2016
29 de agosto a 1 de setembro de 2016 – Foz do Iguaçu, Brasil

RESUMO: A qualidade do leite vem se mostrando um fator de preocupação, diversas alterações na cadeia produtiva foram adotadas, com vistas a busca por mais qualidade, implantação de instruções normativas e uso de bonificações ao produtor. Pontos como a nutrição animal, redução da incidência de mastites e higiene na ordenha são os principais e serão abordados ao longo desse estudo. O objetivo do trabalho foi a avaliação da possibilidade de aplicação da análise discriminante na separação de níveis que qualidade do leite e a classificação do leite com base nesses níveis para bonificação do conjunto de atributos, incentivando a busca pela qualidade como um todo dentro do sistema produtivo. O método mostrou-se promissor e a ampliação do banco de dados e a aplicação de alguns outros tratamentos estatísticos podem transformar o método em uma ferramenta adequada de avaliação.

PALAVRAS-CHAVE: Mineração de Dados; Regras de Classificação; Bonificação por Qualidade.

DISCRIMINANT ANALYSIS MODEL PROPOSED TO APPLICATION ON MILK CLASSIFICATION BY QUALITY

ABSTRACT: The milk quality has proved a factor of concern; a number of changes in the production chain were adopted, in order to search for more quality, implementation of normative instructions and use subsidies to producers. Points, such as animal nutrition, reducing the mastitis incidence and milking hygiene are major and will be addressed throughout this study. The objective was to evaluate the possibility of discriminant analysis application to levels of milk quality separate and does the milk classification based on these levels for bonus the set of attributes, encouraging the search for quality as a whole within the productive system. The method was promising and expanding this database and applied any others, statistical treatments can make the method an appropriate assessment tool.

KEYWORDS: Data Mining; Classification Rules; Bonus for Quality.

INTRODUÇÃO

Fatores como o investimento em tecnologia e pesquisa segundo Jank, Nassar & Tachinardi (2004-2005), são os principais responsáveis pelo aumento de produtividade e pela posição de destaque do agronegócio brasileiro. Corrêa, Veloso & Barczsz (2010), afirmam ser o setor lácteo um dos sistemas agroindustriais brasileiros mais relevantes para a economia do nosso País. O pagamento do leite valorizando sólidos e qualidade, de acordo com Carvalho (2010), foi um passo relevante na modernização da gestão da cadeia produtiva, mas, o Brasil ainda se encontra muito aquém do esperado, quando comparado a outros países mais desenvolvidos.

A avaliação do leite recebido pelos laticínios é realizada com base em diversas análises, mas cada uma delas é utilizada individualmente agregando valor em bonificação quando dentro dos limites estabelecidos. O presente estudo objetivou a apresentação de uma proposta de aplicação da análise discriminante na avaliação do conjunto formado pelas três principais análises, na identificação de três

níveis de qualidade do leite, denominadas superior, intermediária e inferior, valorizando os produtores que chegarem mais próximo das metas desse conjunto e penalizando os que se distanciarem, com base na análise estatística multivariada.

MATERIAIS E MÉTODOS

Para iniciar o desenvolvimento do estudo, escolheu-se três análises, CCS (Contagem de Células Somáticas) e CBT (Contagem de Células Bacterianas Totais) dados em UFC (Unidades Formadoras de Colônias) em cada mL de leite e o EST (Extrato Seco Total), dado em porcentagem, dentre todas as possíveis análises de qualidade do leite, sabendo-se que estas são análises previstas na legislação do setor, mais especificamente na Instrução Normativa 62 (Brasil, 2011), e também utilizadas na prática para avaliação da qualidade do leite e realização dos cálculos de bonificação pelos laticínios. Elas representam respectivamente a sanidade do animal pela presença ou não de mastite, a higiene na ordenha e a quantidade e qualidade da alimentação dos animais, por meio da avaliação do percentual de sólidos presentes no produto. Os limites utilizados para realização do estudo, escolhidos com base nos limites preconizados pela empresa que forneceu acesso a seu banco de dados, encontram-se descritos na Tabela 1.

Tabela 1. Valores mínimos e máximos utilizados pela agroindústria para avaliação da qualidade do leite entregue pelos produtores para fins de cálculo das bonificações.

Níveis de Qualidade	CCS (UFC/mL)		CBT (UFC/mL)		EST (%)	
	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo
Inferior	601.000	Sem limite	601.000	Sem limite	Sem limite	11,60
Intermediária	301.000	600.000	301.000	600.000	11,61	11,80
Superior	Sem limite	300.000	Sem limite	268.000	11,81	Sem limite

Realizou-se a coleta de dados reais dessas três análises, para composição de um banco de dados com 50 repetições para cada uma das 3 classes distintas de produto, nomeadas nesse estudo como: superior, intermediária e inferior, de forma que se pudesse obter um conjunto de dados reais, para cada nível de qualidade, para os três atributos do produto acompanhados CCS, CBT e EST. Os dados não foram normalizados e os valores discrepantes não foram removidos para construção dos modelos.

Com o auxílio do *software* livre de mineração de dados Tanagra, o conjunto de 450 dados obtido dessa coleta, foi submetido a realização de uma análise discriminante linear, obtendo-se as funções de classificação. Aplicou-se ainda uma análise discriminante canônica, para obtenção dos centroides e dos modelos para o cálculo das coordenadas do primeiro (x) e segundo (y) fatores, dentro do gráfico de correlação canônica. Utilizou-se ainda o *software* Action Stats[®] 3.1 para obtenção dos valores mínimos e máximos encontrados para os níveis de qualidade na amostra de 450 dados utilizados para a construção dos modelos. Após a obtenção dos resultados para as amostras dos grupos distintos, três amostras de análises individuais de três produtores foram posicionadas, indicando a classificação destas em relação às amostras utilizadas no treinamento do modelo.

RESULTADOS E DISCUSSÃO

Os resultados da análise de mínimos e máximos encontrados na amostra de 450 dados utilizados para a aplicação da análise discriminante, encontram-se apresentados na Tabela 2. Cabe destacar que o limite legal mínimo de acordo com a IN 62 (Brasil, 2011) para EST, seria de 8,4%, observou-se que na amostra de dados reais escolhida o valor mínimo obtido chegou exatamente a esse limite. Já os valores de CCS e CBT máximos encontrados foram bem elevados, mas cabe destacar que podem ocorrer resultados reais ainda superiores, dessa forma esse trabalho prévio não esgota as possibilidades e pode ser ampliado, com vistas a realização de transformação nos dados, ampliação do banco de dados de treinamento do modelo, até a obtenção de resultados que possam cobrir melhor as várias possibilidades. Os modelos obtidos deverão ainda ser adequados, levando-se em conta as futuras alterações previstas de redução dos limites legais, conforme a IN 62 (Brasil, 2011).

Mesmo diante dessas limitações a aplicação da análise discriminante mostrou-se útil na construção dos grupos e os modelos se apresentaram capazes de discriminar os diferentes níveis de qualidade, com base no conjunto de análises de qualidade composto pelas três análises escolhidas. O valor de Wilks' Lambda para a análise multivariada como um todo foi 0,1565 ou seja próximo de zero indicando que os centroides encontrados apresentam capacidade de separar os grupos e classificar

amostras utilizadas na construção do modelo de forma eficiente. Os testes de Bartlett e Rao também apresentados pelo Tanagra indicaram também um bom desempenho do modelo como um todo e a taxa de erro apresentada foi de 11,33%.

Tabela 2. Valores mínimos e máximos dos valores reais das análises de avaliação da qualidade do leite entregue pelos produtores ao laticínio estudado.

Níveis de Qualidade	CCS (UFC/mL)		CBT (UFC/mL)		EST (%)	
	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo
Inferior	614.000	3.757.000	631.000	6.835.000	8,40	11,59
Intermediária	303.000	596.000	301.000	592.000	11,62	11,8
Superior	16.000	300.000	7.000	268.000	11,81	15,68

Gonçalves, Dias & Muniz (2008), explicam que a análise discriminante permite a identificação dos fatores que apresentam maior capacidade de predição em relação ao fenômeno em estudo, nesse trabalho os resultados individuais de cada uma das variáveis utilizadas para construção do modelo pela análise discriminante, podem ser visualizados na Figura 1, sendo que todas elas foram estatisticamente significativas e devem ser mantidas no modelo ($p < 0,05$). Os valores de Wilks' Lambda foram também todos próximos de zero o que indica a boa qualidade do modelo que utiliza as três variáveis. Os coeficientes de cada uma nas funções discriminantes também estão listados na Figura 1, os modelos de classificação sugeridos pelo software foram apresentados nas Equações 1, 2 e 3.

Figura 1. Resultados da Análise Discriminante Linear, valores das funções de classificação e avaliação estatística de cada um dos atributos de qualidade do leite utilizados para a classificação.

Atributos	Funções de Classificação			Avaliação Estatística			
	Inferior	Intermediária	Superior	Wilks L.	Partial L.	F(2,145)	p-value
CCS	0,000008	0,0000001	-0,000002	0,267876	0,584105	51,62158	0,000000
CBT	0,000001	-0,000002	-0,000003	0,244845	0,639048	40,95005	0,000000
EST	43,449808	46,074644	49,241362	0,234131	0,668291	35,98573	0,000000
Constante	-251,422339	-271,244531	-308,662818	-			

$$\text{Inferior} = 0,000008 \text{ CCS} + 0,000001 \text{ CBT} + 43,449808 \text{ EST} - 251,422339 \quad (1)$$

$$\text{Intermediária} = 0,0000001 \text{ CCS} - 0,000002 \text{ CBT} + 46,074644 \text{ EST} - 271,244531 \quad (2)$$

$$\text{Superior} = 0,000002 \text{ CCS} - 0,000003 \text{ CBT} + 49,241362 \text{ EST} - 308,662818 \quad (3)$$

Calculou-se os valores limites dos níveis de qualidade pelos modelos, foram obtidos os valores inferiores, superiores e médios, com base nos valores do banco de dados de treinamento, valores extremos foram desconsiderados e os resultados encontram-se na Tabela 3.

Tabela 3. Valores máximos, mínimos e médios propostos, resultantes da aplicação dos modelos para cada um dos três níveis de qualidade do leite, com base nos valores das análises reais utilizadas.

Níveis de Qualidade	Valores resultantes do modelo		
	Mínimo	Máximo	Média
Inferior	109,5401	263,4519	186,4960
Intermediária	263,5711	271,3119	267,4415
Superior	272,8887	463,2377	368,0632

De acordo com Pizzol (2004), a análise discriminante, além de separar os grupos, também pode ser aplicada para classificar indivíduos dentro dos grupos estudados. Dessa forma, se uma amostra aleatória de análise de leite de um produtor for coletada pode-se identificar a qual grupo pertence pelos modelos. A Tabela 4 apresenta resultados reais de análise de leite de 3 produtores e sua respectiva classificação, de acordo com os modelos construídos nesse trabalho, a classificação deve ser feita com base no modelo que resulta no maior valor calculado.

Utilizando a técnica atualmente empregada pelo laticínio pelos limites descritos na Tabela 1, pode-se observar que o produtor 1 teria as análises de CCS e CBT classificadas como nível intermediário e EST como superior e aqui em conjunto ele recebeu a classificação de superior, cabe destacar que os valores de CCS e CBT desse produtor não estão muito distantes do limite de 300.000 necessário para classificação individual superior. Em relação ao produtor 2, seria classificado

individualmente para CCS e CBT como inferior e EST como superior, mas muito próximo do limite de intermediário, dessa forma a classificação do produtor como um todo em inferior mostrou-se próxima da classificação realizada atualmente. No caso do produtor 3 percebeu-se que para CBT ele seria realmente classificado como intermediário sendo que para CCS e EST estaria na classificação inferior, ele foi classificado no nível de qualidade intermediária com base no modelo multivariado.

Tabela 4. Dados reais de análise de leite de três produtores de leite, para os atributos estudados nesse artigo, valores de.

Análises realizadas	Produtores escolhidos		
	Produtor 1	Produtor 2	Produtor 3
CCS (UFC/mL)	484.000	984.000	885.000
CBT (UFC/mL)	459.000	2.947.000	549.000
EST (%)	12,38	12,61	11,37
Classificação:	Valores calculados pelo modelo		
Inferior	290,8172	307,2987	250,2310
Intermediária	298,2899	303,9611	251,6147
Superior	300,5362	305,3978	251,3345
Classificação	Superior	Inferior	Intermediária

Os resultados apresentados parecem indicar que o modelo aqui apresentado estaria sendo brando, penalizando pouco produtores com leites de qualidade inferior, cabe destacar que a presença de valores bastante elevados de CBT e CCS presentes no banco de dados utilizado podem estar distorcendo os padrões, por levar o centroide do grupo inferior muito para baixo, uma análise mais criteriosa dos limites ainda precisa ser realizada antes da aplicação prática do modelo, mas a ilustração da possibilidade da aplicação desse método não é comprometida por essa observação. Em futuros estudo deve-se ainda avaliar a possibilidade de remoção desses outliers e a normalização dos dados com vista a melhorias na classificação. Os resultados da análise de correlação canônica mostraram significância para os dois fatores resultantes da análise sendo que o fator 1 foi mais significativo que o 2 e os valores de correlação foram mais elevados (Figura 2).

Figura 2. Resultados da Matriz de Correlação Canônica.

Fatores	Fator n1			Fator n2		
	Total	Within	Between	Total	Within	Between
CCS	-0,863971	-0,591443	-0,996748	-0,175396	-0,264059	-0,080587
CBT	-0,798490	-0,497568	-0,986985	-0,326677	-0,447681	-0,160812
EST	0,756768	0,465481	0,946226	-0,649670	-0,878819	-0,323507

As equações para obtenção das coordenadas relativas a cada ponto, que representam as amostras utilizadas na construção do modelo, bem como qualquer conjunto de dados amostrado, encontram-se ilustradas no gráfico da Figura 3.

Figura 3. Gráfico ilustrativo dos pontos obtidos com base no conjunto das coordenadas, calculadas pela análise de correlação canônica, para os três níveis de qualidade do produto utilizadas (inferior, intermediária e superior).

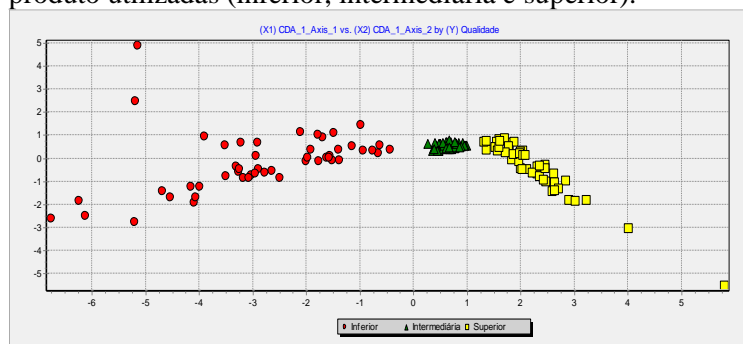


Figura 4. Coordenadas dos centroides calculados para os três níveis de qualidade do leite.

Qualidade	Fator n1	Fator n2
Inferior	-2,848477	-0,164312
Intermediária	0,638445	0,528875
Superior	2,210030	-0,364563
Correlação Canônica	0,820136	0,130078

Com os valores apresentados pelo software as Equações 4 e 5, foram construídas. O conjunto de coordenadas relativas aos centroides dos três níveis de qualidade do leite que formam os três diferentes grupos apresentados na Figura 3 podem ser visualizados na Figura 4.

$$\text{Fator 1 (x)} = -0,0000019 \text{ CCS} - 0,0000008 \text{ CBT} + 1,0797917 \text{ EST} - 10,81554947 \quad (4)$$

$$\text{Fator 2 (y)} = -0,0000008 \text{ CCS} - 0,0000005 \text{ CBT} - 1,6450321 \text{ EST} + 20,43322569 \quad (5)$$

Calculando-se as coordenadas, para os três produtores utilizados como exemplo nesse trabalho, elaborou-se a Tabela 5. Realmente as coordenadas calculadas pelo programa ficaram próximas dos centroides de classificação do produto da mesma forma que calculado pelas equações discriminantes o que confirmou os resultados de classificação do leite desses produtores em superior, inferior e intermediária respectivamente.

Tabela 5. Resultados das coordenadas calculadas para a qualidade do leite dos três produtores utilizados como exemplo nesse estudo.

	Produtor 1		Produtor 2		Produtor 3	
	X	Y	X	Y	X	Y
Valores das coordenadas	1,265472	-4,03377	-1,42657	-9,65612	-0,65901	-5,62528

CONCLUSÕES

A metodologia proposta mostrou-se útil para a classificação do leite com base no conjunto dos três diferentes atributos de qualidade, nos três níveis de qualidade propostos, inferior, intermediário e superior. Mais estudos podem ser realizados com vistas a avaliar as melhorias que ocorreriam no modelo por meio do aumento do tamanho do banco de dados, pela normalização dos dados por transformações e pela remoção de valores extremos, que podem estar interferindo de forma tendenciosa nos modelos apresentados.

Sugere-se que as agroindústrias pensem na adoção dessa ferramenta e invistam no desenvolvimento de modelos multivariados de avaliação, pois estes são capazes de avaliar a qualidade como um todo, bonificando a boa conduta. De nada adianta o produtor alimentar bem os animais e não cuidar da saúde ou fazer tudo correto e não ter higiene no momento da ordenha, qualidade é um conjunto e deveria ser tratado como tal e é o que está sendo proposto neste estudo.

AGRADECIMENTOS

Ao laticínio, que forneceu o acesso aos dados reais de análise de qualidade do leite e a Universidade Tecnológica Federal do Paraná que motivou e possibilitou o desenvolvimento desse projeto de extensão intitulado Interação Universidade e Empresa: Avaliação Estatística de Dados de Acompanhamento de Processos Produtivos, onde a ideia de realização dessa análise surgiu.

REFERÊNCIAS

- Brasil. Ministério da Agricultura Pecuária e Abastecimento. Instrução Normativa No 62. Aprovada em 29 de dezembro de 2011. Dispõe sobre os regulamentos técnicos de produção, identidade e qualidade de leite. Diário Oficial [da] República Federativa do Brasil, Brasília, 2011.
- Corrêa, C. C.; Veloso; A. F.; Barczsz; S. S. A Logística de Coleta e Distribuição do Leite como Diferencial Competitivo para os Pequenos Processadores de Leite. In: Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural, 48. Campo Grande, 25 a 28 de julho de 2010.
- Goncalves, C. A.; Dias, A. T.; Muniz, R. M. Análise discriminante das relações entre fatores estratégicos, indústria e desempenho em organizações brasileiras atuantes na indústria manufatureira. Revista de Administração Contemporânea, Curitiba, v.12, n.2, p.287-311, 2008.
- Jank, M. S.; Nassar, A. M.; Tachinardi, M. H. Revista USP, São Paulo, n.64, p. 14-27, dezembro/fevereiro 2004-2005.
- Pizzol, S. J. S. Combinação de grupos focais e análise discriminante: um método para tipificação de sistemas de produção agropecuária. Revista de Economia e Sociologia Rural, Brasília, v.42, n.3, p.451-468, 2004.