

USO DE ANÁLISE ESTATÍSTICA MULTIVARIADA PARA TIPIFICAÇÃO DE PRODUTORES DE LEITE DE MINAS GERAIS

Use of multivariate statistical analysis for Minas Gerais' milk producers characterization

Armando Chinelatto Neto¹, Gilmar Pinheiro Cunha Castro², João Eustáquio de Lima³

RESUMO

O presente trabalho teve como objetivo principal identificar e caracterizar sistemas de produção de leite por meio de métodos de análise estatística multivariada. Foram utilizadas a análise fatorial, cluster, discriminante e correlação canônica. A escala e custeio da produção foram o principal critério de classificação entre os produtores. Combinando-se os resultados obtidos com a análise de cluster e a análise de correlação canônica, o segundo critério de classificação agrega a área de pastagens e os cuidados com a sanidade do rebanho. Foram identificados três grupos de produtores, destacando-se o primeiro porque agrega cerca de 90% do total analisado e é composto por produtores relativamente menores. A análise discriminante e as médias entre indicadores financeiros confirmam esta classificação.

Palavras-chave: análise multivariada, pecuária leiteira, Minas Gerais.

ABSTRACT

The present work had as its main objective to identify and characterize milk production systems through methods of multivariate statistical analysis. Factorial analysis, cluster, discriminant and canonical correlation were utilized. The production gradation and defrayal were the main classification criterium among the producers. Combining the results obtained with the cluster analysis and the canonical correlation analysis, the second classification analysis aggregates the grazing lands and the care with the flock health. Three producers groups were identified, among which the first one is prominent for aggregating about 90% of the total analyzed and is composed by relatively smaller producers. The discriminant analysis and the measuring among financial indicators confirm such classification.

Key words: multivariate analysis, milk cattle raising, Minas Gerais.

1. INTRODUÇÃO

O complexo agroindustrial do leite passou por grandes transformações, caracterizadas pela saída do governo como agente determinante do preço, a abertura comercial contribuindo para criar uma nova situação de concorrência para os produtores nacionais, e a estabilidade econômica resultante do Plano Real.

Segundo Gomes (1996), a implantação do sistema de preços livres representou forte incentivo à modernização da pecuária leiteira nacional, e o novo padrão de concorrência fez com que o produtor ajustasse sua atividade à nova realidade.

Face a este novo ambiente econômico, é necessário considerar a importância social e econômica da cadeia do leite. Segundo dados do Censo Agropecuário de 1996, o valor da produção de leite corresponde a 9,68% do PIB agropecuário nacional e constitui importante elo com

outros setores da economia. No âmbito social, a cadeia do leite gera 3,5 vezes mais empregos que a construção civil, 3,02 mais que a siderurgia e 4,77 mais que a indústria automobilística (OLIVEIRA, 1998).

Diante das transformações ocorridas na atividade leiteira e de sua relevância no complexo agroindustrial, é importante o levantamento e a caracterização dos produtores para a seleção de problemas de pesquisa e para a definição de políticas para o setor. De acordo com Fernandes & Lima (1991), para conhecer a realidade das propriedades rurais e encontrar subsídios para gerar e transferir tecnologias compatíveis com esta realidade, torna-se necessário ter conhecimento do perfil dessas propriedades. Portanto, esse perfil será importante para a adaptação de tecnologias já existentes, bem como a geração e transferência de novas tecnologias compatíveis com as condições encontradas.

¹ Doutorando em Economia Aplicada – Universidade Federal de Viçosa – Departamento de Economia Rural – 36.571-000 – Viçosa, MG – achinelatto@yahoo.com.br

² Doutorando em Economia Aplicada – Universidade Federal de Viçosa – Departamento de Economia Rural – 36.571-000 – Viçosa, MG – gcastro@ufv.br

³ Professor Titular, Ph. D – Universidade Federal de Viçosa – Departamento de Economia Rural – 36.571-000 – Viçosa, MG – jelima@ufv.br

Recebido em 12/04/04 e aprovado em 24/09/04

Neste contexto, objetivou-se com este trabalho identificar e caracterizar os produtores de leite de Minas Gerais utilizando-se métodos de análise estatística multivariada.

2. METODOLOGIA

A produção de leite envolve a conjugação de várias dimensões, sendo necessário a utilização de técnicas que permitam a análise simultânea de um conjunto de variáveis. Portanto, serão utilizados os métodos de análise fatorial, análise de cluster, análise discriminante e análise de correlação canônica. A análise fatorial permite descrever um conjunto de variáveis com um número menor de fatores. Isto é possível quando as variáveis apresentam certo grau de correlação entre si, de modo que grande parte da sua variância possa ser explicada por fatores comuns a todas elas e não pelos fatores específicos a cada uma.

De acordo com Manly (1994), pode-se sistematizar o método de análise fatorial por meio de uma variável Z padronizada como mostra a equação (1):

$$Z_i = A_{i1}F_1 + A_{i2}F_2 + \dots + A_{ik}F_k + d_i U_i \quad (1)$$

Em que:

A_i = Carga fatorial;

F_k = Fator comum;

U_i = Fator único (d é a sua carga fatorial);

k = Número de fatores que serão considerados, sendo menor ou igual a N ; e

N = Número de variáveis.

Inicialmente as variáveis são padronizadas⁴, de modo que tenham média zero e desvio padrão 1, evitando que ordens de grandeza diferentes prejudiquem a análise.

É construída a matriz de correlação entre as variáveis e são calculados seus autovalores (“eigenvalues”) que representam quanto da variância da amostra pode ser atribuída a cada fator, definindo-se quantos fatores serão extraídos. Geralmente, são escolhidos os fatores cujos autovalores são maiores que a unidade, pois autovalores menores explicam menos que a variância de uma das variáveis originais.

As cargas fatoriais estimadas, chamadas de “factor loadings”, são diretamente relacionadas com os coeficientes de correlação entre o fator e a variável original. O somatório do quadrado das cargas fatoriais fornece uma

estimativa da comunalidade, ou seja, quanto da variância de cada variável pode ser explicada pelos fatores comuns. Os autovalores representam quanto da variância da amostra é explicada pelos fatores individualmente.

É comum obter uma matriz de difícil interpretação, na qual não é possível identificar quais variáveis são mais importantes para cada fator. Diante deste problema, processa-se a rotação da matriz de cargas fatoriais, associando de maneira mais nítida um número de variáveis a cada fator. Entre os métodos de rotação de fatores destaca-se o VARIMAX, pois permite que os fatores sejam ortogonais (não correlacionados) e aproxima as cargas fatoriais a valores próximos a zero ou um. Porém, outros métodos podem ser eventualmente utilizados para tornar mais nítida a associação entre os fatores e as variáveis originais.

Os escores fatoriais (“factor scores”) são os valores do produto da matriz de coeficientes de cada fator pela matriz transposta das variáveis padronizadas para cada unidade da amostra, sendo utilizados para classificar a unidade quanto a característica captada pelo fator.

Serão usadas as seguintes variáveis para determinar os fatores de diferenciação dos produtores:

PLE = Produção de leite (litros por dia);

VLA = Número de vacas em lactação (n°);

GMO = Gastos totais com mão-de-obra familiar e contratada (R\$);

QCO = Quantidade de concentrado (kg);

GMI = Gastos com minerais (R\$);

GEC = Gastos com energia e combustíveis (R\$);

GME = Gastos com medicamentos (R\$);

QVE = Quantidade de vermífugo (ml);

APN = Área com pastagens naturais (hectares);

APF = Área com pastagens formadas (hectares);

ACA = Área com capineiras (hectares); e

ACM = Área com cana-de-açúcar forrageira e milho para silagem (hectare).

De posse dos escores fatoriais, a análise de *cluster* permite classificar o conjunto de unidades ou observações em grupos homogêneos. De acordo com Duran & Odell (1974), citados por Souza (2000), este método permite determinar, a partir das características das unidades,

⁴ A padronização consiste em expressar as variáveis em termos de unidades de desvio padrão. A operacionalização consiste em subtrair a média e dividir pelo desvio padrão.

subconjuntos em que cada unidade pertença a somente um subconjunto e que as unidades agrupadas em um mesmo subconjunto sejam similares e aquelas pertencentes a grupos distintos sejam diferentes.

Monta-se uma matriz de dados com os valores dos “factor scores” que são capazes de captar as características de interesse. Para verificar se a unidade A é mais parecida com B do que com C são construídas estimativas de distância, podendo ser obtidas usando-se a distância euclidiana⁵.

A escolha do método para o agrupamento exige o conhecimento de suas propriedades e dos objetivos da pesquisa. De acordo com Hair Júnior (1995), citado por Souza (2000), o critério essencial é a maximização das diferenças entre os grupos em relação à variação dentro dos grupos.

Entre as técnicas utilizadas, as mais comuns são as hierárquicas aglomerativas, nas quais classificam-se as unidades com sucessivas fusões em grupos, reduzindo a um único grupo ao final. Monta-se a matriz de distância entre as unidades, e inicia-se as fusões com base nas proximidades. Os resultados são resumidos em um dendrograma, que corresponde a um gráfico bidimensional combinando a ocorrência da fusão com a estimativa de distância das unidades agrupadas.

Diante das diferentes técnicas existentes, determinando diferentes medidas de distância e critérios de agrupamento, podem ser empregadas e comparadas diversas técnicas para a validação dos *clusters* obtidos. As distâncias podem ser obtidas pelo quadrado das distâncias euclidianas, potencializando as diferenças, ou como foi realizado por Hoffmann (1992), atribuindo-se ponderações aos escores fatoriais com base na contribuição destes para explicar a variância. Quanto aos métodos de agrupamento existentes, Souza (2000) utilizou os métodos de ligação simples, ligação completa, ligação média, e outros, com o intuito de optar pelo que melhor se ajustou aos seus objetivos. O que se sugere é a natureza exploratória dos métodos, em que se optou pelas distâncias euclidianas, o que ofereceu melhor diferenciação entre as unidades analisadas.

Após agrupadas as observações em grupos homogêneos, foi utilizada a análise discriminante para a análise desses grupos. De acordo com Züge & Carvalho

Neto (1999), trata-se de uma técnica estatística multivariada que tem o objetivo de confirmar ou verificar uma classificação feita *a priori*.

Para Haddad (1989), essa técnica testa a significância dessa classificação prévia e determina quais variáveis tem o poder de distinguir em que grupo devem estar as unidades do estudo. Consiste na criação de uma ou mais variáveis Z, sendo combinações lineares das variáveis originais, da seguinte forma:

$$Z_i = a_1 X_{1i} + a_2 X_{2i} + \dots + a_n X_{ni} \quad (2)$$

Os valores de Z_i são os escores da variável Z, os a_j são os coeficientes ponderados calculados pelo método e os X_{ji} são as variáveis discriminadoras. Espera-se que os escores computados aos elementos de um mesmo grupo sejam semelhantes e diferentes entre os grupos.

As variáveis utilizadas para fazer a discriminação dos grupos foram:

ATO = Área total (ha);

KFI = Capital fixo imobilizado (R\$);

GCO = Gastos com concentrado (R\$); e

PDT = Produtividade (litros/total de vacas).

Como é descrito por Haddad (1989) e Valentin (2000), a função discriminante é estimada de forma a maximizar a separação ou diferença entre os grupos e a homogeneidade interna de cada grupo⁶, resolvendo a seguinte equação:

$$|W^{-1}B - I\lambda| = 0 \quad (3)$$

Em que:

W = Matriz da dispersão entre os grupos;

B = Matriz da dispersão dentro dos grupos; e

λ = Os autovalores ou raízes características de $W^{-1}B$.

O número de funções discriminantes, considerando p variáveis e g grupos, é o mínimo entre p e $g-1$. São extraídas em ordem decrescente de importância, ou seja, a

⁵ A distância euclidiana entre duas unidades é a extração da raiz quadrada da soma dos quadrados das diferenças entre os diferentes escores fatoriais das unidades.

⁶ As hipóteses básicas do método de análise discriminante são: a) Os grupos de variáveis discriminantes tem distribuição multivariada normal; b) As matrizes de covariâncias dos grupos são iguais; e c) Os grupos diferem quanto as médias.

primeira função extrai o máximo de variância entre os grupos, a segunda função extrai a maior variância remanescente, e assim por diante. Para cada função discriminante existe um autovalor representando a porção da variância total entre os grupos que é explicada pela função.

Em seguida, é utilizada a análise de correlação canônica para associar dois grupos de variáveis. O primeiro é associado à alimentação do rebanho, e o segundo é associado aos cuidados com a sanidade do rebanho. São consideradas as variáveis relacionadas na Tabela 1.

Estes dois grupos de variáveis foram analisados pelo método de Correlação Canônica. De acordo com Bernstein et al. (1988), Dillon & Goldstein (1984) e Manly (1994), na análise de correlação canônica são extraídas combinações lineares de dois conjuntos de medidas das mesmas unidades de estudo. São pares canônicos de modo que o primeiro par responda pela maior parcela de variância em cada conjunto de medidas e pela maior correlação entre os dois conjuntos. O segundo par responderá pela segunda maior parcela de variância e pela segunda maior correlação, independentes do primeiro par; e assim por diante.

De acordo com Gomes (1986), as variáveis geradas pela primeira correlação canônica podem ser expressas da seguinte forma:

$$Y_{11} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (4)$$

$$Y_{12} = b_{11}X_{p+1} + b_{12}X_{p+2} + \dots + b_{1n}X_n \quad (5)$$

Em que Y_{11} é a combinação linear do primeiro grupo de variáveis e Y_{12} a combinação linear do segundo grupo. Os coeficientes a 's e b 's são determinados de modo que seja obtida a máxima

correlação entre Y_{11} e Y_{12} . Analogamente são gerados outros pares canônicos.

Foram usados neste trabalho os dados de 348 produtores de leite que fizeram parte da amostra de produtores entrevistados por SEBRAE-MG & FAEMG (1996), para elaboração do Diagnóstico da Pecuária Leiteira de Minas Gerais em 1995/96. Devido a existência de produtores que não utilizaram todos os fatores de produção, a amostra foi reduzida para 294.

3. RESULTADOS E DISCUSSÃO

Os resultados da análise fatorial após a rotação pelo método VARIMAX estão descritos na Tabela 2. Foram destacados com um asterisco os valores iguais ou superiores a 0,6 tornando possível a identificação de três fatores.

O primeiro fator representa o nível de escala/custeio da produção e os cuidados com a sanidade do rebanho. Está positivamente correlacionado com a produção diária de leite (PLE), número de vacas em lactação (VLA), quantidade de concentrado (QCO), gastos totais com mão-de-obra (GMO), gastos com minerais (GMI), gastos com energia e combustíveis (GEC), gastos com medicamentos (GME) e a quantidade de vermífugo utilizada (QVE).

O segundo fator representa o fornecimento de volumoso, comum na época de secas. Está positivamente correlacionado com a área de capineira (ACA) e a área destinada ao plantio da cana-de-açúcar e milho para silagem (ACM).

No terceiro fator predominam a área com pastagens naturais (APN) e pastagens formadas (APF), representando o fornecimento de alimentação volumosa comum na época das águas, caracterizada pelo uso de pastagens.

TABELA 1 – Grupos de variáveis utilizadas na análise de correlação canônica

Grupo 1	Grupo 2
Área com pastagem formada (ha)	Gastos com medicamentos (R\$)
Área com pastagem natural (há)	Quantidade de vermífugo (ml)
Área com capineira (ha)	Quantidade de doses de vacina contra aftosa
Área com forrageira e silagem (ha)	Quantidade de doses de vacina contra brucelose

Fonte: dados da pesquisa.

TABELA 2 – Cargas fatoriais após a rotação ortogonal e a comunalidade

Variáveis	Fator 1	Fator 2	Fator 3	Comunalidade
PLE	* 0,685	0,606	0,208	0,880
VLA	* 0,730	0,386	0,428	0,865
GMO	* 0,686	0,581	0,223	0,858
QCO	* 0,835	0,149	0,106	0,730
GMI	* 0,851	0,216	0,076	0,776
GEC	* 0,627	0,597	0,153	0,773
GME	* 0,700	0,318	0,366	0,725
QVE	* 0,867	-0,029	0,109	0,765
APN	-0,014	0,128	* 0,875	0,783
APF	0,401	0,071	* 0,619	0,549
ACA	-0,032	* 0,848	0,149	0,743
ACM	0,569	* 0,600	-0,044	0,686

Fonte: dados da pesquisa.

Os três fatores são capazes de diferenciar os produtores pois explicam 76,1% da variância total dos dados. Os percentuais individuais são de 41,8%, 20,7% e 13,6%, respectivamente. Dada a maior contribuição do primeiro fator para explicar a variância total dos dados e sua correlação com oito variáveis, pode-se concluir que a escala e o nível de custeio da produção é o melhor critério de classificação das propriedades.

A comunalidade mede a contribuição dos fatores para explicar a variância total de cada variável. Cada comunalidade é igual à soma dos quadrados das cargas fatoriais, sendo a maioria superior a 0,7. A área com pastagem formada e a área de cana-de-açúcar e milho são exceções, apresentando comunalidades de 0,549 e 0,686, respectivamente.

O teste de esfericidade de Bartlett testa a hipótese da matriz de correlações populacionais simples entre as variáveis ser igual a matriz identidade. Espera-se rejeitar esta hipótese, pois é necessário existir correlações entre grupos de variáveis para que façam parte do mesmo fator. De acordo com o teste de Bartlett, a matriz de correlações populacionais é diferente da matriz identidade a 1% de significância.

O teste de KMO é uma medida de adequabilidade da análise fatorial. Considera a razão entre os coeficientes de correlação simples e a soma dos coeficientes de correlações parciais e simples, tendo o objetivo de indicar a necessidade de eliminação de variáveis. Foi encontrado o valor de 0,912, sendo que quanto mais próximo de 1 melhor a adequabilidade da análise.

Obtidos os escores fatoriais para cada um dos 294 produtores de leite, com relação a cada fator, realizou-se a análise de cluster. Foi utilizado o método de ligação simples

para a identificação de grupos homogêneos de produtores. O resultado foi a classificação de 5 grupos, agregando 277 produtores no primeiro grupo, 8 produtores no segundo grupo e 7 no terceiro. Dois grupos foram formados por um produtor em cada.

A análise das variáveis e dos escores que constituíram a análise fatorial dos produtores mostrou a maior importância do primeiro fator para fazer a distinção dos grupos. Como esse fator está mais associado à escala e nível de custeio da produção, os resultados determinam a existência de 277 pequenos produtores, ou menores em comparação aos 15 proprietários restantes. Portanto, trata-se de 94,2% de pequenos produtores, ou seja, a grande maioria.

A análise discriminante foi utilizada para testar a classificação dos três primeiros grupos. Foram desconsiderados os dois últimos grupos porque cada um é formado por somente uma propriedade. Portanto, testou-se a classificação de 292 produtores.

Foram utilizadas as variáveis área total utilizada (ATO), o capital fixo imobilizado (KFI), os gastos com concentrado (GCO) e a produtividade medida em litros por vacas (PDT). Essas variáveis podem ser consideradas como *proxys* para uso de terra, nível de investimento ou capital, custeio da alimentação e tecnologia.

De acordo com o teste do λ de Wilks, duas funções são estatisticamente relevantes para separar os grupos ao nível de 1% de significância. De acordo com as raízes características, a primeira função explica 84,5% da variância dos dados, e a segunda explica 15,5%. As correlações entre os escores discriminantes e as variáveis estão descritas na Tabela 3:

TABELA 3 – Coeficientes de correlação ente os escores discriminantes e as variáveis

Variáveis	Função 1	Função 2
ATO	0,920 **	-0,391
KFI	0,748 **	0,519 *
GCO	0,281	0,832 **
PDT	0,045	0,491 *

Fonte: dados da pesquisa.

** correlações fortes

* correlações moderadas

O primeiro escore discriminante correlaciona-se com o uso de terra cultivada com o objetivo de alimentar o rebanho (ATO) e o capital ou nível de investimento (KFI). O segundo escore discriminante correlaciona-se com os gastos com a alimentação (GCO), e moderadamente com capital e a produtividade (PDT).

De maneira geral, todas as variáveis utilizadas foram relevantes para a classificação dos produtores. O capital fixo imobilizado (variável importante nas duas funções discriminantes), área total (variável mais importante da função discriminante de maior raiz característica) e o gasto com concentrado (variável mais importante da segunda função discriminante) têm maior capacidade de diferenciação dos grupos.

A tecnologia, medida pela produtividade parcial, pode ser considerada como relativamente constante ou, por outro lado, heterogênea dentro dos grupos, pois não foi importante para discriminar os produtores. Como o principal critério de formação dos grupos de produtores está associado à escala de produção, a análise dos coeficientes de correlação simples entre a produtividade e variáveis associadas a este critério explica os resultados encontrados. Obteve-se os coeficientes de correlação simples de 0,53 e 0,16 entre a produtividade e a produção total, e o capital fixo imobilizado.

Quanto às estatísticas de classificação, confirmou-se 94,5% da classificação previamente feita pela análise de *cluster*. No primeiro grupo foram 263 classificações corretas, 9 produtores foram considerados como pertencentes ao grupo dois e 2 produtores foram considerados como pertencentes ao grupo três (96% de previsões corretas). No segundo grupo foram 5 classificações corretas e 3 produtores foram considerados como pertencentes ao grupo um (62,5% de acerto). No grupo três foram 5 classificações corretas e 2 produtores foram considerados como pertencentes ao grupo um. Confirmou-se a existência de um grande grupo formado pelos pequenos produtores, agregando cerca de 277 produtores de acordo com a análise de *cluster* e 265 produtores de acordo com a análise discriminante.

Quanto a análise de correlação canônica, procurou-se associar variáveis que representam os gastos com a sanidade do rebanho com a área destinada a alimentação constituída por pastagens e silagem. O primeiro grupo é formado pela área de pastagens naturais (APN), área de pastagens formadas (APF), área de capineiras (ACA) e área plantada com cana-de-açúcar e milho. O segundo grupo é composto pelos gastos com medicamentos (GME), número de doses de vacinas contra aftosa (QVA), número de doses de vacina contra brucelose (QVB) e quantidade de vermífugo aplicada (QVE). Foi encontrado o coeficiente de correlação de 0,818 entre o primeiro par canônico, com nível de significância de 1%.

Os coeficientes de correlação entre as variáveis originais e as variáveis canônicas estão descritas na Tabela 4:

A primeira variável canônica do segundo grupo correlaciona-se com todas as variáveis relativas aos cuidados com a sanidade do rebanho. A primeira variável canônica do primeiro grupo está correlacionada com a área destinada à pastagens, destacando-se as pastagens formadas. O primeiro par canônico evidencia a relação entre a alimentação via pastagens, principalmente naturais, e os cuidados com a sanidade do rebanho.

Note que a primeira variável canônica é formada pelas variáveis que estão associadas ao segundo e terceiro fatores obtidos pela análise fatorial. Como esses dois fatores são responsáveis pela explicação de 34,3% da variância total (20,7% e 13,6%, respectivamente) e foi encontrado um coeficiente de correlação canônica de 0,818 entre o primeiro par canônico, conclui-se que os cuidados com a sanidade do rebanho e a alimentação por meio de pastagens podem ser um segundo critério de separação entre os grupos de produtores.

Foram calculadas as médias da renda bruta (produção anual multiplicada pelo preço), margem bruta operacional efetiva (renda bruta menos os custos operacionais efetivos, ou seja, receitas menos desembolsos) e produção diária em litros. Os resultados para os três grupos estão descritos na Tabela 5:

TABELA 4 – Coeficientes de correlação entre as variáveis originais e canônicas

Primeiro grupo de variáveis		Segundo grupo de variáveis	
APN	* 0,548	GME	* 0,636
APF	** 0,950	QVA	** 0,978
ACA	0,194	QVB	* 0,451
ACM	0,394	QVE	* 0,537

Fonte: dados da pesquisa

* Correlação moderada

** Alta correlação

TABELA 5 – Médias da renda bruta, margem bruta operacional e produção dos grupos

Grupos	Renda bruta (R\$/ano)	Margem bruta operacional (R\$/ano)	Produção (litros/dia)
1	14.034,20	3.144,08	147,00
2	53.137,59	14.700,89	587,38
3	82.455,67	16.859,11	807,21

Fonte: dados da pesquisa

A relação entre os valores médios demonstra a diferença de escala entre os produtores. O primeiro grupo possui menor renda bruta anual, menor margem bruta operacional e menor produção diária. O segundo grupo possui valores intermediários e o terceiro possui os maiores valores.

Percebe-se a proximidade dos valores encontrados para margem bruta entre o segundo e terceiro grupos, podendo estar associados aos limites de ganho com a escala de produção. Entre o segundo e o primeiro grupo, verifica-se que o ganho relativo de margem bruta operacional é maior que o ganho relativo de renda bruta. Porém, a comparação entre o terceiro e segundo grupo revela que, neste nível, há maior ganho relativo de renda bruta do que ganho relativo de margem bruta.

4. CONCLUSÕES

Análises preliminares, com base na extração de fatores e a formação de grupos, revelou que as variáveis mais importantes para a diferenciação dos produtores estão associadas à escala e custeio da produção. Em seguida o uso de pastagens, podendo ser associado ao modelo extensivo de produção comum dos períodos de águas; e por último ao uso de volumoso ou tratamento no coxo. Os três fatores extraídos foram capazes de diferenciar os produtores, haja vista a porcentagem que explicam da variância total dos dados originais.

Confirmou-se a diferenciação dos grupos por meio da análise discriminante, na qual as variáveis mais importantes foram o nível de exploração de terra, capital e gastos com alimentação. A tecnologia é relativamente constante, ou heterogênea dentro dos grupos, pois a produtividade não foi variável importante para diferenciar os produtores.

Com base na combinação dos resultados obtidos pela análise fatorial e análise de correlação canônica, pode-se estabelecer um critério de classificação agregando a área cultivada com vistas à alimentação (destacando-se a área constituída por pastagens) e os cuidados com a sanidade do rebanho.

Portanto, são estabelecidos dois critérios de diferenciação entre os produtores: 1) Escala e custeio da produção; 2) Cuidados com a sanidade do rebanho e área destinada à pastagens.

Foram estabelecidos três grupos de produtores. Destaca-se o primeiro grupo porque agrega cerca de 90% do total analisado e é composto por pequenos produtores.

A análise dos indicadores financeiros utilizados confirma esta classificação entre os produtores. Nota-se a proximidade entre os valores obtidos nos dois últimos grupos, formados pelos maiores produtores, podendo significar os limites de ganho com a escala de produção.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BERNSTEIN, I. H.; TENG, G. K.; GARBIN, C. P. **Applied multivariate analysis**. New York: Spriger-Verlag, 1988.
- DILLON, W. R.; GOLDSTEIN, M. **Multivariate analysis: methods and applications**. New York: J. Wiley & Sons, 1984.
- FERNANDES, T. A. G.; LIMA, J. E. Uso de análise multivariada para identificação de sistemas de produção. **Pesquisa Agropecuária Brasileira**, Brasília, v. 26, n. 10, p. 1823-1836, out. 1991.
- GOMES, S. T. **Condicionantes da modernização do pequeno agricultor**. São Paulo: IPE, 1986. 181 p.
- GOMES, S. T. **A economia do leite**. Coronel Pacheco: EMBRAPA-Gado de Leite, 1996. 98 p.
- HADDAD, P. R. **Economia regional: teorias e métodos de análises**. Fortaleza: Banco do Nordeste do Brasil, 1989. 694 p.
- HOFFMANN, R. A dinâmica da modernização da agricultura em 157 microrregiões homogêneas do Brasil. **Revista de Economia e Sociologia Rural**, Brasília, v. 30, n. 4, p. 271-290, out./dez. 1992.
- MANLY, B. F. J. **Multivariate statistical methods**. 2. ed. New York: Chapman & Hall, 1994. 215 p.
- OLIVEIRA, S. Agricultura é a melhor saída para o desemprego. **Diário do Comércio**, Belo Horizonte, p. 11, 25 mar. 1998.
- SEBRAE-MG; FAEMG. **Diagnóstico da pecuária leiteira de Minas Gerais: relatório de pesquisa**. Belo Horizonte, 1996. 2 v.
- SOUZA, P. M. **Modernização e mudanças estruturais na agricultura brasileira, 1970 a 1995**. 2000. 287 f. Tese (Doutorado em Economia Rural) - Universidade Federal de Viçosa, Viçosa, 2000.
- VALENTIN, J. L. **Ecologia numérica: uma introdução à análise multivariada de dados ecológicos**. Rio de Janeiro: Interciência, 2000. 117 p.
- ZÜGE, M.; CARVALHO NETO, A. Utilização de métodos estatísticos multivariados na avaliação do desempenho empresarial. **Revista Paranaense de Desenvolvimento**, Curitiba, n. 97, p. 101-112, set./dez. 1999.