

ANÁLISE DE COMPONENTES PRINCIPAIS DE VALORES GENÉTICOS PREDITOS PARA CARACTERÍSTICAS SELECIONADAS NA RAÇA NELORE

**ISADORA DOS SANTOS VICENTE¹; ARIONE AUGUSTI BOLIGON²; FABIO
RICARDO PABLOS DE SOUZA³; RICARDO ZAMBARDA VAZ⁴**

¹Universidade Federal de Pelotas – isadoravicente@msn.com

²Universidade Federal de Pelotas – arioneboligon@yahoo.com.br

³Universidade Federal de Pelotas – fabiopablos@hotmail.com

⁴Universidade Federal de Pelotas – rzvaz@terra.com.br

1. INTRODUÇÃO

As análises multivariadas são um conjunto de métodos estatísticos que permitem avaliar simultaneamente medidas múltiplas de cada indivíduo. A análise de componentes principais foi introduzida por PEARSON (1901) e fundamentada por HOTELLING (1933), sendo atualmente aplicada em diversas áreas, auxiliando os pesquisadores na manipulação e interpretação dos conjuntos de dados. A análise de componentes principais é um método multivariado geralmente utilizado para se explorar a variabilidade das características e a estrutura de um conjunto de dados. Com a aplicação dessa metodologia, podem-se condensar as informações contidas em todas as variáveis originais em um número menor de variáveis latentes ortogonais, independentes entre si, denominadas componentes principais, com menor perda possível de informação.

Na avaliação genética é comum a disponibilidade de um grande número de características, muitas das quais podem mostrar-se redundantes ou altamente correlacionadas. Quando o número de características mensuradas é grande, há possibilidade de algumas serem de pouca contribuição para a discriminação dos indivíduos avaliados. Dessa forma, o resultado seria um acréscimo considerável de trabalho para mensurar grande número de variáveis sem melhoria na precisão, além de tornar mais complexa a análise e a interpretação dos resultados.

Em programas de melhoramento de bovinos de corte, os objetivos específicos de seleção em geral contemplam a redução na idade ao abate, melhorias no rendimento e acabamento de carcaça dos animais, aumento da eficiência produtiva e reprodutiva das fêmeas, sem aumentar suas exigências nutricionais/custo de manutenção, entre outros. Para atingir tais objetivos, esses programas disponibilizam avaliações genéticas para várias características. Entretanto, a seleção com base em valores genéticos de várias medidas de interesse tende a reduzir o ganho genético para cada uma delas individualmente, se comparada à seleção praticada para poucas características. Assim, pode-se fazer uso da técnica de componentes principais como objetivo de resumir a informação contida no complexo de variáveis originais, ou seja, nos valores genéticos, eliminando as informações redundantes (KHATTREE; NAIK, 2000).

O presente estudo foi desenvolvido com o objetivo de se verificar a importância de valores genéticos preditos para características mensuradas em bovinos da raça Nelore, com a utilização de componentes principais, além de estabelecer critérios para eliminação de variáveis redundantes, com reduzida perda de informações.

2. METODOLOGIA

As informações utilizadas no presente trabalho fazem parte da base de dados denominada de Aliança Nelore, com medidas provenientes de 600.132 animais, distribuídos em 263 diferentes fazendas e nascidos entre os anos de 1984 e 2010, sendo mantidos em pastagens tropicais. Foram avaliados o ganho em peso do nascimento a desmama (GND) e pós desmama (GDS), escores de conformação, precocidade e musculatura a desmama (CD, PD e MD, respectivamente) e ao sobreano (CS, PS e MS, respectivamente) e perímetro escrotal (PE).

Os animais foram pesados ao nascimento, desmama (em torno de 7 meses de idade) e ao sobreano (em torno de 18 meses de idade). Além das pesagens, os animais foram avaliados visualmente para escores de conformação, precocidade e musculatura a desmama e ao sobreano, quando também é obtida a medida do perímetro escrotal, através de uma fita métrica adequada. As notas dos escores são dadas a cada animal individualmente, avaliando-se o animal em relação ao grupo de contemporâneos ao qual pertence, sendo atribuídas por técnicos treinados. As notas variam de um a cinco, sendo cinco a melhor expressão da característica.

As predições dos valores genéticos (VG) foram obtidas em análises uni-características, utilizando um modelo animal linear (para GND, GDS e PE) e threshold (para os escores visuais) em abordagem bayesiana, sendo realizadas com o auxílio de programas computacionais desenvolvidos por MISZTAL et al. (2002). As análises consistiram em cadeia única de 800.000 ciclos, com um período de burn-in de 200.000 ciclos, retirando-se amostras a cada 20 iterações.

Os VG das características consideradas foram padronizados visando evitar inconsistências nas soluções da técnica multivariada devido a diferentes escalas ou magnitudes das variáveis. Essa padronização (resultando em média nula e variância unitária) foi obtida como: $z_i = (x_i - \bar{x}_i) / s_i$, em que: z_i é o valor padronizado da variável x_i , \bar{x}_i é a média da característica i e s_i é o respectivo desvio padrão. Os VG padronizados dos animais para todas as características que compõem os índices de seleção (VGGND, VGCD, VGPD, VGMD, VGGDS, VGCS, VGPS, VGMS e VGPE) foram utilizados em uma análise de componentes principais.

A escolha dos componentes principais que explicaram a maior parte da variação do conjunto de dados original foi determinada por aqueles com autovalores superiores à unidade, segundo o critério de Kaiser (KAISER, 1958). As análises foram processadas no software R (2014).

3. RESULTADOS E DISCUSSÃO

Em análises de componentes principais, segundo o critério de Kaiser (KAISER, 1958) deve-se estudar apenas a variação contida nos autovetores (componentes principais) que possuem autovalores acima da unidade, pois estes contêm informações relevantes sobre as variáveis. Os primeiros componentes respondem pela maior parte da variância, de forma que, no presente estudo, o máximo de informação, em termos de variação total, está contido nos primeiros três componentes (Tabela 1), os quais apresentaram autovalores maiores do que 1. Dessa forma, podemos simplificar o processo de seleção dos animais ao considerar um número menor de variáveis, sem uma perda significativa de informações.

Tabela 1: Variância explicada e cumulativa dos componentes principais (CP).

CP	Variância explicada (%)	Variância cumulativa (%)
1	48,06	48,06
2	18,03	66,09
3	12,97	79,06
4	8,49	87,55
5	4,51	92,06
6	2,84	94,9
7	2,21	97,11
8	1,78	98,88
9	1,11	100,00

No presente estudo, os três primeiros componentes principais foram responsáveis por 79,06% da variação total existente nos valores genéticos preditos para as características estudadas, enquanto que os demais componentes explicaram isoladamente pouca parte dessa variação. O primeiro componente principal (CP1), responsável por 48,06% da variação total, apresentou alto poder discriminatório para os valores genéticos dos escores de precocidade e musculatura (Tabela 2).

Tabela 2: Correlações entre os valores genéticos das características que compõem os índices de seleção com os três primeiros componentes principais (CP).

Características ¹	CP1	CP2	CP3
VGGND	0,73	0,54	0,21
VGCD	0,62	0,58	0,39
VGPD	0,79	0,32	-0,38
VGMD	0,78	0,37	-0,37
VGGDS	0,55	-0,57	0,38
VGCS	0,65	-0,15	0,63
VGPS	0,76	-0,44	-0,24
VGMS	0,77	-0,42	-0,26
VGPE	0,54	-0,22	-0,10

¹VGGND = valor genético do ganho do nascimento a desmama, VGCD = valor genético da conformação a desmama, VGPD = valor genético da precocidade a desmama, VGMD = valor genético da musculatura a desmama, VGGDS = valor genético do ganho da desmama ao sobreano, VGCS = valor genético da conformação ao sobreano, VGPS = valor genético da precocidade ao sobreano, VGMS = valor genético da musculatura ao sobreano, VGPE = valor genético do perímetro escrotal. Valores grifados: maior correlação dentro de cada componente principal.

O segundo CP, explicando 18,03% da variação total, foi associado de maneira positiva com as características a desmama e, de maneira negativa com as características ao sobreano. Esse resultado indica o agrupamento dos animais ao redor de dois extremos: um grupo com boas características genéticas a desmama e outro grupo com boas características genéticas no pós-desmama. Foram observadas diferenças em relação ao desempenho a desmama *versus* sobreano, o que pode estar evidenciando um contraste entre animais com crescimento diferenciado. Este segundo CP poderia ser utilizado na tentativa de contrastar animais com maior ou menor precocidade de crescimento ou taxa de maturação. O terceiro CP foi responsável por 12,97% da variação total, apresentando maior poder discriminante para VGCS, com correlação de 0,63. VAL et al. (2008) utilizando

valores genéticos preditos para peso aos 120 e 210 dias (efeito direto e materno), peso e perímetro escrotal aos 365 e 450 dias (efeito direto) e idade ao primeiro parto relataram que dois componentes principais foram suficientes para capturar a maior variação nos valores genéticos. Além disso, o primeiro componente principal estava mais associado com características de peso e perímetro escrotal e, o segundo, as características maternas de animais da raça Nelore.

4. CONCLUSÕES

O primeiro componente principal apresentou associação positiva com os valores genéticos de todas as características estudadas, mas de maior magnitude com os escores visuais de precocidade e musculatura. Por outro lado, a seleção baseada no segundo componente deve proporcionar melhorias genéticas para as características obtidas na desmama.

Análise de componentes principais de valores genéticos incluídos nos índices de seleção de bovinos da raça Nelore poderia auxiliar a avaliação genética de características de interesse econômico.

5. REFERÊNCIAS BIBLIOGRÁFICAS

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v.24, p.417-441, 1933.

KAISER, H.F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v.23, p.187-200, 1958.

KHATTREE, R.; NAIK, D.N. Multivariate data reduction and discrimination with SAS software. Cary: SAS Institute, 2000. 574p.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D.H. BLUPF90 and related programs (BGF90). **Proc. 7th World Congress on Genetics Applied to Livestock Production**. Montpellier, France. Communication No 28-07, 2002.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, v.2, p.559-572, 1901.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Online. 2014. Disponível em: <http://www.R-project.org/>

VAL, J.E. ; FERRAUDO, A.S.; BEZERRA, L.A.F.; CORRADO, M.P.; LÔBO, R.B.; FREITAS, M.A.R.; PANETO, J.C.C. Alternativas para seleção de touros da raça Nelore considerando características múltiplas de importância econômica. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.60, n.3, p.705-712, 2008.