

Seleção de características morfofuncionais de cavalos da raça Mangalarga Marchador por meio da análise de componentes principais

[Identification of morphofunctional traits in Mangalarga Marchador horse using principal component analysis]

C.T. Meira¹, I.G. Pereira^{2*}, M.M. Farah¹, A.V. Pires³, D.A. Garcia¹, V.A.R. Cruz¹

¹Doutorando – Faculdade de Ciências Agrárias e Veterinárias – Unesp – Jaboticabal, SP

²Escola de Veterinária – Universidade Federal de Minas Gerais – Belo Horizonte, MG

³Faculdade de Ciências Agrárias – UFVJM – Diamantina, MG

RESUMO

Dados de 14.288 animais da raça Mangalarga Marchador, nascidos de 1990 a 2005, foram utilizados para avaliar a redução da dimensionalidade do espaço multivariado para características morfofuncionais, por meio da análise de componentes principais. Foram consideradas as características: altura na cernelha, altura na garupa, comprimento da cabeça, comprimento do pescoço, comprimento do dorso, comprimento da garupa, comprimento da espádua, comprimento do corpo, largura da cabeça, largura das ancas, perímetro do tórax, perímetro da canela e a pontuação da marcha. Para tais características, obtiveram-se sete componentes principais, a partir da matriz de correlação, que apresentaram variância inferior a 0,7 (autovalor inferior a 0,7). Isso sugere sete variáveis para descarte, por apresentarem maiores coeficientes de ponderação, em valor absoluto, a partir do último componente principal. A razão para isso é que variáveis altamente correlacionadas com os componentes de menor variância representam variação praticamente insignificante. Com base nesses resultados, recomendam-se as seguintes características para serem mantidas em trabalhos futuros com esta base de dados: pontuação da marcha, altura na garupa, comprimento do dorso, comprimento da garupa, largura da cabeça e perímetro da canela.

Palavras-chave: análise multivariada, andamento, descarte de variáveis, equino, medidas lineares

ABSTRACT

Records from 14,288 animals of the Mangalarga Marchador breed, born from 1990 to 2005, were used to discard morphofunctional traits in a principal component analysis. The following traits were used: height at withers, height at croup, lengths of head, neck, back, croup, hip length and body, widths of head, hip width, thorax perimeter, cannon bone circumference and gait score. For the traits considered it was observed that 7 principal components showed variation lower than 0.7; suggesting that seven variables could be discarded. The reason is that when variable are highly correlated with the principal components of smaller variance, their variation is practically insignificant. Based on those results the recommendation is to maintain the following traits for future research with this database: gait score, height at croup, length of back, length of croup, width of head and cannon bone circumference.

Keywords: discard of variables, equine, gait, multivariate analysis, linear measures

INTRODUÇÃO

Em equinos, a caracterização morfológica está intrinsecamente relacionada ao seu valor econômico, pois as medidas morfométricas

podem influenciar no desempenho dos animais. Para o registro do Mangalarga Marchador na Associação Brasileira dos Criadores do Cavalo Mangalarga Marchador (ABCCMM), são avaliadas as medidas morfométricas e a pontuação da marcha, em que os animais devem obter as pontuações mínimas

Recebido em 4 de maio de 2013

Aceito em 13 de junho de 2013

*Autor para correspondência (corresponding author)

E-mail: idalmogp@vet.ufmg.br

de expressão/caracterização e andamento para serem registrados em definitivo. Porém, para fins de seleção, é possível identificar aquelas características que realmente contribuem para a discriminação dos indivíduos.

Para tanto, a análise de componentes principais (ACP) tem sido útil na elucidação das relações estruturais entre as medidas corporais dos animais e no descarte de variáveis, possibilitando, segundo Regazzi (2002), eliminar informações redundantes em decorrência da correlação com outras variáveis presentes na análise.

Segundo Morrison (2005), cada variável original está associada ao componente principal por meio de um vetor característico, cujo valor absoluto determina a importância da característica naquele componente principal. Dessa forma, é possível identificar quais variáveis originais são passíveis de descarte, o que se faz eliminando as características associadas aos maiores coeficientes nos últimos componentes, ou seja, naqueles que explicam percentuais muito reduzidos da variância total.

Assim, de acordo com Jolliffe (1972, 1973), o número de variáveis descartadas deve ser igual ao número de componentes cuja variância (autovalor) é inferior a 0,7. Com isso, é possível eliminar variáveis redundantes que, além de serem pouco informativas, promovem acréscimo no trabalho de avaliação e não apresentam informação adicional, o que não implica uma perda significativa de informação.

Métodos de análise multivariada para o acompanhamento de resultados com experimentos genéticos em morfologia, segundo Barbosa (1993), permitem definir variáveis adequadas para o estudo de variação genética, facilitando a compreensão dos resultados dos diversos sistemas de acasalamentos e a proposição de índices de seleção.

Objetivou-se com este estudo reduzir a dimensionalidade de um conjunto de características morfofuncionais de cavalos da raça Mangalarga Marchador, por meio da análise de componentes principais, eliminando as informações de variáveis redundantes.

MATERIAL E MÉTODOS

Os dados analisados são provenientes do Serviço de Registro Genealógico da ABCCMM, compostos de 12 características lineares e pontuação da marcha. Após consistência dos dados, eliminando-se mensurações incorretas e/ou informações incompletas, avaliou-se um total de 14.288 animais nascidos entre os anos de 1990 a 2005.

As características estudadas foram: altura na cernelha (distância vertical entre o processo espinhoso da quarta vértebra torácica e o solo), altura na garupa (distância vertical entre a tuberosidade sacral do ílio e o solo), comprimento da cabeça (distância entre o vértice da cabeça e a ponta do focinho), comprimento do pescoço (distância entre a nuca, osso atlas, e o terço médio do osso da escápula), comprimento do dorso (distância entre a base da cernelha e a tuberosidade sacral do ílio), comprimento da garupa (distância entre as partes cranial da tuberosidade ilíaca e a caudal da tuberosidade isquiática), comprimento da espádua (distância entre a articulação escapuloumeral e a cartilagem escapular), comprimento do corpo (distância entre a articulação escapuloumeral e a tuberosidade isquiática), largura da cabeça (distância entre as faces externas das arcadas orbitárias), largura das ancas (distância entre a tuberosidade coxal dos ossos ilíacos), perímetro do tórax (medida do diâmetro obtido sobre os processos espinhosos da oitava e nona vértebras torácicas), perímetro da canela (medida do diâmetro no terço médio do osso metacarpiano) e a pontuação da marcha.

Os animais foram submetidos a exame zootécnico a partir dos 36 meses de idade, do qual constou a mensuração e avaliação quanto à morfologia e andamento, de acordo com a determinação do Padrão Racial. Segundo esse padrão aprovado pelo conselho deliberativo técnico da ABCCMM e pelo Ministério da Agricultura Pecuária e Abastecimento, para fins de registro definitivo na associação, o animal deverá obter, no mínimo, 50% dos pontos de expressão/caracterização e de andamento.

O andamento marchado foi analisado visualmente e com o cavaleiro quanto aos quesitos: gesto, rendimento, diagrama, estilo e comodidade; sendo a pontuação dada como um

Seleção de características...

todo, e não especificamente para cada quesito. Porém, como houve mudanças nas proporções da pontuação da marcha e morfologia ao longo dos anos, foi feito um ajuste nas pontuações de marcha para colocá-las todas em uma única escala (0 a 100 pontos). Já as medidas morfométricas foram obtidas por meio de mensurações.

Para obtenção dos componentes principais utilizou-se o procedimento PRINCOMP do programa SAS (Statistical..., 2002), no qual o ponto de partida foi a matriz de correlação, em que as variáveis são padronizadas para média zero e variância igual a um. Optou-se pela utilização de uma matriz de correlação em vez de uma matriz de covariância para amenizar possíveis discrepâncias acentuadas entre as variâncias e permitir as comparações entre os autovetores em um componente. A solução, utilizando-se a matriz de correlação, é recomendada quando as variáveis são medidas em escalas muito diferentes entre si, pois essa matriz é equivalente à matriz das variáveis padronizadas (Johnson e Wichern, 1998), visto que as características lineares analisadas constam de avaliações métricas e a marcha por tabela de pontuação subjetiva.

A técnica de componentes principais, a partir da matriz de correlação, consiste em transformar um conjunto de p variáveis X_1, X_2, \dots, X_p em um novo conjunto Y_1, Y_2, \dots, Y_p , em que os Y 's apresentam as seguintes propriedades:

I. Cada componente principal (Y_i) é uma combinação linear das variáveis padronizadas (X_j), ou seja:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p = \sum_{j=1}^p a_{ij}X_j$$

onde a_{ij} são os autovetores, com $i = 1, 2, \dots, p$ e $\sum_{j=1}^p a_{ij}^2 = 1$

II. O primeiro componente principal, Y_1 , é tal que sua variância é máxima entre todas as combinações lineares de X . O segundo componente principal é não correlacionado com o primeiro e possui a segunda maior variância. Da mesma forma, definam-se os outros p componentes principais não correlacionados entre si, ou seja:

$$\text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_p)$$

III. A cada componente principal Y_i existem p autovalores (λ) ordenados de forma que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

IV. As combinações lineares formadas são não correlacionadas:

$$\text{Cov}(Y_1, Y_2) = \text{Cov}(Y_1, Y_3) = \dots = \text{Cov}(Y_{p-1}, Y_p) = 0$$

A importância relativa de um componente principal foi avaliada pela porcentagem de variância total que ele explica, ou seja, a porcentagem de seu autovalor em relação ao total dos autovalores de todos os componentes, que é dado por:

$$Y_i = \frac{\text{Var}(Y_i)}{\sum_{i=1}^p \text{Var}(Y_i)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100$$

O critério para descarte de variáveis utilizado foi baseado nas recomendações de Jolliffe (1972, 1973), que sugere que o número de variáveis descartadas deve ser igual ao número de componentes principais cuja variância (autovalor) é inferior a 0,7; e na sugestão de Regazzi (2002), o qual considera que a variável que apresentar o maior coeficiente em valor absoluto no componente principal de menor autovalor (menor variância) deverá ser menos importante para explicar a variância total e, portanto, passível de descarte. Assim, o processo de descarte consistiu em considerar o componente correspondente ao menor autovalor e rejeitar a variável associada ao maior coeficiente de ponderação (em valor absoluto). Então, o próximo menor componente foi avaliado. Esse processo continuou até que o componente associado ao autovalor inferior a 0,7 fosse considerado. A razão para isso é que variáveis altamente correlacionadas aos componentes principais de menor variância representam variação praticamente insignificante.

RESULTADOS E DISCUSSÃO

Os seis primeiros componentes principais explicaram 78,57% da variação total dos dados, sendo os primeiros componentes associados aos maiores autovalores, retendo, assim, maior variância dos dados.

Tabela 1. Componentes principais (CP), autovalores (λ_i), porcentagem da variância explicada pelos componentes (% VCP) e porcentagem acumulada das características morfofuncionais

Componentes principais	λ_i	% VCP	% VCP (acumulada)
CP ₁	5,1454	39,58	39,58
CP ₂	1,4218	10,94	50,52
CP ₃	1,1318	8,71	59,23
CP ₄	0,9040	6,95	66,18
CP ₅	0,8753	6,73	72,91
CP ₆	0,7358	5,66	78,57
CP ₇	0,6500	5,00	83,57
CP ₈	0,5375	4,13	87,70
CP ₉	0,4726	3,64	91,34
CP ₁₀	0,4211	3,24	94,58
CP ₁₁	0,3630	2,79	97,37
CP ₁₂	0,2713	2,09	99,46
CP ₁₃	0,0700	0,54	100,00

Conforme pode ser observado na Tabela 1, sete componentes apresentaram autovalores menores que 0,7. De acordo com o critério de Jolliffe (1972, 1973), sete variáveis podem ser descartadas, pois, além de o número de componentes que apresentaram autovalores menores que 0,7 ser sete, essas variáveis estão associadas a componentes de menor importância relativa e que explicam pouco da variabilidade dos dados. A baixa variabilidade de algumas características indica que elas contribuem pouco para a discriminação entre os indivíduos.

Tabela 2. Coeficientes de ponderação das características morfofuncionais de cavalos da raça Mangalarga Marchador com os componentes principais descartados em ordem de menor importância

Variáveis	Coeficientes						
	CP ₇	CP ₈	CP ₉	CP ₁₀	CP ₁₁	CP ₁₂	CP ₁₃
PM	-0,008441	-0,160406	0,049761	-0,006363	-0,005156	0,040904	0,025495
AC	-0,179291	0,039700	-0,241013	-0,038134	0,035852	0,226048	-0,742946
AG	-0,229578	0,021207	-0,285887	-0,176611	0,079722	0,302833	0,654815
CCa	-0,255405	-0,637959	0,181100	-0,102863	0,024918	-0,085216	-0,003611
CPe	0,280074	0,214078	0,626573	0,219704	0,012385	0,102363	0,071005
CD	-0,287399	0,318634	0,395462	0,073810	0,004559	0,110775	-0,046275
CG	-0,076314	0,422086	-0,001957	-0,418598	0,422286	0,093006	-0,025729
CE	-0,048383	0,263842	-0,332232	0,464722	-0,556973	-0,009461	0,073865
CC	0,002513	0,080660	-0,006578	-0,224559	-0,074982	-0,844358	0,016544
LC	0,491651	0,164384	-0,056075	-0,017110	0,105860	-0,101434	0,032224
LA	0,558600	-0,255412	0,045544	-0,357537	-0,334944	0,301489	-0,046025
PT	0,171550	-0,250558	-0,099013	0,575664	0,565953	-0,033006	0,028181
PC	-0,312172	-0,092650	0,387913	0,043400	-0,231678	0,039458	0,032562

PM: pontuação da marcha; AC: altura na cernelha; AG: altura na garupa; CCa: comprimento da cabeça; CPe: comprimento do pescoço; CD: comprimento do dorso; CG: comprimento da garupa; CE: comprimento da espádua; CC: comprimento do corpo; LC: largura da cabeça; LA: largura das ancas; PT: perímetro do tórax; PC: perímetro da canela.

Assim, as sete variáveis que apresentaram maiores coeficientes de ponderação, em valor absoluto, a partir do último componente principal em direção ao primeiro, foram descartadas, conforme apresentado na Tabela 2. Como o perímetro do tórax já havia sido descartado no componente principal 11 (CP₁₁), optou-se por descartar, no componente principal 10 (CP₁₀), a segunda característica associada ao maior coeficiente de ponderação. As variáveis sugeridas para descarte foram, respectivamente, em ordem de menor importância para explicar a variação total: altura na cernelha (AC), comprimento do corpo (CC), perímetro do tórax (PT), comprimento da espádua (CE), comprimento do pescoço (CPe), comprimento da cabeça (CCa) e largura das ancas (LA). Isso porque, além de estarem associadas a componentes que explicam muito pouco da variabilidade dos dados, apresentaram correlação linear simples significativas com as demais, fornecendo pouca ou nenhuma informação adicional e se caracterizando, assim, pela redundância (Tab. 3). Verifica-se, por exemplo, que a altura na cernelha (variável sugerida para descarte) é altamente correlacionada com a altura na garupa (0,9178); assim como o comprimento do corpo apresenta correlação alta com altura na garupa (0,7062). Esse comportamento pode ser observado para as demais características passíveis de descarte.

Seleção de características...

Pinto *et al.* (2005) avaliaram medidas morfométricas de potros da raça Mangalarga Marchador por meio da análise de componentes principais. Foram efetuadas 25 mensurações lineares e 11 mensurações angulares em potros e potras recém-nascidos, aos seis meses e aos 12 meses de idade. Os autores concluíram que a análise de componentes principais foi eficiente em reduzir o número de medidas morfométricas nas diferentes idades estudadas, pois não apresentaram variação importante para a discriminação dos animais, ou estão fortemente correlacionadas com alguma das outras variáveis, o que as tornam redundantes.

Os resultados de Yamaki *et al.* (2009) para características de produção de aves de três linhagens fêmeas de corte, mesmo sendo com espécie diferente do presente estudo, se assemelham. Dos doze componentes principais, sete apresentaram variância (autovalor) menor que 0,7, descartando, assim, sete variáveis que apresentaram redundância com as demais.

Já as características recomendadas para trabalhos futuros foram: pontuação da marcha (PM), altura na garupa (AG), comprimento do dorso (CD), comprimento da garupa (CG), largura da cabeça (LC) e perímetro da canela (PC). Pode-se observar, pela Tabela 3, que essas variáveis apresentaram menores correlações entre si.

Para que um animal possa ser registrado na ABCCMM, ele deverá obedecer a um padrão racial do qual constam avaliações do seu exterior a partir de mensurações de características morfométricas e ainda avaliação do seu andamento segundo tabela de pontuação. Levando-se em conta que algumas características apresentam mútuas correlações, a análise de componentes principais tem se mostrado uma ferramenta útil para concentrar informações disponíveis em um número reduzido de conjunto. Assim, essa análise foi usada no sentido de reduzir o número de mensurações (características) obrigatórias a serem avaliadas no momento de registro do animal. Porém, o descarte da característica altura na cernelha só será recomendado para trabalhos de pesquisas, por exemplo, pois essa característica, para a ABCCMM, é de importância para fins de registro, sendo que uma altura fora dos padrões ideais (medidas já pré-estabelecidas) é ponto de desclassificação para o registro genealógico definitivo dos animais. A estatística de componentes principais permite identificar as características a serem eliminadas, mas a escolha de qual ou quais características podem ser eliminadas cabe a uma decisão técnica. Os resultados dos componentes principais apenas indicam as características que são redundantes.

Tabela 3. Coeficientes de correlação de Pearson entre as características morfofuncionais de cavalos da raça Mangalarga Marchador

	PM	AC	AG	CCa	CPe	CD	CG	CE	CC	LC	LA	PT	PC
PM	1,000												
AC	0,082	1,000											
AG	0,045	0,917	1,000										
CCa	-0,005	0,434	0,424	1,000									
CPe	0,006	0,515	0,434	0,268	1,000								
CD	0,066	0,130	0,196	0,210	-0,152	1,000							
CG	0,146	0,359	0,324	0,315	0,304	0,005	1,000						
CE	0,132	0,537	0,469	0,444	0,349	0,142	0,560	1,000					
CC	0,133	0,726	0,706	0,360	0,465	0,179	0,442	0,525	1,000				
LC	0,011	0,239	0,229	0,238	-0,005	0,318	0,004	0,154	0,137	1,000			
LA	0,105	0,446	0,450	0,306	0,289	0,168	0,401	0,451	0,583	0,174	1,000		
PT	0,125	0,548	0,514	0,329	0,301	0,162	0,428	0,515	0,588	0,139	0,586	1,000	
PC	0,079	0,421	0,379	0,160	0,168	0,019	0,352	0,308	0,414	0,115	0,324	0,436	1,000

PM: pontuação da marcha; AC: altura na cernelha; AG: altura na garupa; CCa: comprimento da cabeça; CPe: comprimento do pescoço; CD: comprimento do dorso; CG: comprimento da garupa; CE: comprimento da espádua; CC: comprimento do corpo; LC: largura da cabeça; LA: largura das ancas; PT: perímetro do tórax; PC: perímetro da canela.

Barbosa (1993) utilizou três métodos de análise multivariada – dentre eles a análise de componentes principais – para avaliar a

importância das medidas lineares entre animais campeões e não campeões da raça Mangalarga Marchador. As variáveis puderam ser reduzidas

de doze para sete, sendo semelhante ao resultado encontrado no presente estudo. No caso de machos não campeões, eliminaram-se as seguintes variáveis: altura na cernelha, comprimento do corpo, comprimento da garupa, perímetro do tórax e comprimento da espádua. Para machos campeões, as variáveis descartadas foram: altura na cernelha, comprimento da garupa, comprimento da espádua, largura das ancas e perímetro da canela.

A aplicação dessa análise na espécie equina tem se mostrado eficiente em avaliar características que possam realmente discriminar os animais e usá-las em programas de seleção, o que também foi constatado por outros autores (Barbosa, 1993; Kashiwamura et al., 2001; Miserani et al., 2002; Pinto et al., 2005; Santos, 2006). Entretanto, para a escolha dos critérios de seleção a serem praticados na raça, deve-se estimar componentes de (co)variâncias genéticas para as características selecionadas. Ao conhecer a magnitude desses componentes pode-se, então, estabelecer os critérios de seleção de acordo com os objetivos da seleção. A análise de componentes principais proporcionou a identificação das características que apresentam maiores variações, que, portanto, podem responder melhor à seleção.

CONCLUSÕES

Tendo em vista os resultados obtidos, a análise de componentes principais se mostrou efetiva e permitiu o descarte de sete variáveis que apresentaram baixa variabilidade ou foram redundantes por estarem correlacionadas com as de maior importância. Assim, o menor número de variáveis necessárias para explicar a variação total resulta em economia de tempo e de recursos em futuros trabalhos que utilizarão essa mesma base de dados, sem perda significativa de informação.

AGRADECIMENTOS

À Associação Brasileira dos Criadores do Cavalo Mangalarga Marchador (ABCCMM) e a FAPEMIG pelo apoio financeiro.

REFERÊNCIAS

- BARBOSA, C.G. *Estudo morfométrico na raça Mangalarga Marchador. Uma abordagem multivariada*. 1993. 76f. Dissertação (Mestrado em Zootecnia) - Universidade Federal de Minas Gerais, Belo Horizonte.
- JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. Englewood Cliffs: Prentice Hall, 1998. 816p.
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. I: Artificial data. *Appl. Stat.*, v.21, p.160-173, 1972.
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. II: Real data. *Appl. Stat.*, v.22, p.21-31, 1973.
- KASHIWAMURA, F.; AVGAANDORJ, A.; FURUMURA, K. Relationships among body size, conformation, and racing performance in Banei Draft racehorses. *J. Equine Sci.*, v.12, p.1-7, 2001.
- MISERANI, M.G.; McMANUS, C.; SANTOS, S.A. et al. Avaliação de fatores que influem nas medidas lineares do cavalo Pantaneiro. *Rev. Bras. Zootec.*, v.31, p.335-341, 2002.
- MORRISON, D.F. *Multivariate Statistical Methods*. Australia: Brooks/Cole Thomson Learning, 2005. 469p.
- PINTO, L.F.B.; ALMEIDA, F.Q.; AZEVEDO, P.C.N. et al. Análise multivariada das medidas morfométricas de potros da raça Mangalarga Marchador: Análise Fatorial. *Rev. Bras. Zootec.*, v.34, p.613-626, 2005.
- REGAZZI, A.J. *Análise multivariada: notas de aula*. Viçosa: UFV, 2002.
- STATISTICAL Analysis System - SAS, 2002. *SAS Software: versão 9.0*. SAS Institute Inc., Cary, NC, USA.
- SANTOS, L.M. *Morfologia e genética do cavalo Campolina*. 2006. 48f. Dissertação (Mestrado em Genética) - Universidade Federal de Minas Gerais, Belo Horizonte.
- YAMAKI, M.; MENEZES, G.R.O.; PAIVA, A.L.C. et al. Estudo de características de produção de matrizes de corte por meio da análise de componentes principais. *Arq. Bras. Med. Vet. Zootec.*, v.61, p.227-231, 2009.