

Seleção de variáveis de desempenho de suínos por meio da análise de componentes principais

[Identification of informative performance traits in swine using principal component analysis]

L. Barbosa¹, P.S. Lopes^{2*}, A.J. Regazzi³, S.E.F. Guimarães², R.A. Torres²

¹Estudante de Pós-Graduação – UFV

²Departamento de Zootecnia – UFV – Viçosa, MG

Av. P.H. Rolfs, s/n

36571-000 – Viçosa, MG

³Departamento de Informática – UFV – Viçosa, MG

RESUMO

Dados de 435 animais de uma população F2 de suínos foram utilizados para avaliar a possibilidade de redução da dimensão do espaço multivariado, por meio da técnica de componentes principais. Foram avaliadas as seguintes características de desempenho: tamanho da leitegada ao nascer (TLN), tamanho da leitegada à desmama (TLD), número de tetos (NT), pesos ao nascer (PN), aos 21 (P21), aos 42 (P42), aos 63 (P63) e aos 77 (P77) dias de idade e ganho médio de peso diário (GPD), consumo de ração (CR) e conversão alimentar (CA) dos 77 aos 105 dias de idade. Seis componentes principais, obtidos a partir da matriz de correlação, apresentaram variância inferior a 0,7 (autovalores inferiores a 0,7), o que sugere seis variáveis para descarte, por apresentarem maiores coeficientes, em valor absoluto, a partir do último componente principal. As variáveis descartadas apresentaram correlação linear simples significativa com as demais. Em razão do grande número de variáveis redundantes, 54,5% delas podem ser eliminadas, sendo recomendada a avaliação de apenas TLN, NT, P77, CR e CA, sem que haja perda considerável da informação.

Palavras-chave: suíno, análise multivariada, correlação, crescimento, descarte, desempenho, tamanho da leitegada

ABSTRACT

Using principal component analysis, records of 435 animals from an F2 swine population were used to identify independent and informative variables of economically important performance. The following performance traits were recorded: litter size at birth (BL), litter size at weaning (WL), teat number (TN), birth weight (BW), weight at 21 (W21), 42 (W42), 63 (W63) and 77 (W77) days of age, average daily gain (ADG), feed intake (FI) and feed:gain ratio (FGR) from 77 to 105 days of age. Six principal components expressed variation lower than 0.7 (eigen values lower than 0.7) suggesting that six variables could be discarded with little information loss. The discarded variables present significant simple linear correlation with the retained variables. Retaining variables BL, TN, W77, FI and FGR and eliminating all the rest would retain most of the relevant information in the original data set.

Keywords: swine, multivariate analysis, correlation, discard, growth, litter size, performance

Recebido para publicação em 12 de maio de 2004

Recebido para publicação, após modificações, em 10 de fevereiro de 2005

*Autor para correspondência (corresponding author)

E-mail: leandro.b@vicosa.ufv.br

INTRODUÇÃO

A grande maioria das características de importância econômica submetidas à seleção é de natureza quantitativa. Tais características são relacionadas em magnitude e sentido variáveis, de tal forma que a seleção em uma provoca mudanças em outras, razão pela qual, se não forem consideradas, podem levar a erros na avaliação dos indivíduos (Sakaguti, 1994). Entretanto, dentro de um conjunto de variáveis, algumas provêm informações redundantes em razão de suas correlações e podem ser de difícil mensuração ou de custos elevados. Em virtude dos altos custos dos programas de melhoramento, a recomendação geral dos pesquisadores tem sido medir o máximo possível de características, o que tem tornado, por vezes, tais programas ainda mais onerosos e com considerável acréscimo de trabalho. O problema tem sido definir as características que deveriam ser medidas em programas de melhoramento. Para contorná-lo, uma opção seria definir, por meio de técnicas de análise multivariada, as características quantitativas ou o conjunto delas que seriam responsáveis pela maior parte da variação total observada, ou seja, avaliar a informação adicional provida por algumas medidas quando outras já estiverem disponíveis. Segundo Liberato (1995), desde que, na prática, não haja limite para o número de medidas que possam ser obtidas em qualquer problema, é relevante examinar se algumas não estão em excesso na presença de outras.

A técnica de componentes principais, um dos procedimentos multivariados, visa, como um dos propósitos na pesquisa animal, resumir um grande conjunto de características em outro menor, de sentido biológico, além de examinar as correlações entre as características estudadas, avaliar a importância de cada caráter e promover a eliminação daquelas que contribuem pouco.

O trabalho objetivou reduzir a dimensão de um conjunto original de variáveis medidas em suínos, com a menor perda de informação possível, eliminando as informações redundantes existentes em decorrência da correlação entre variáveis e descartando as de pouca contribuição para a variação total.

MATERIAL E MÉTODOS

Os dados são provenientes de uma geração F2 de suínos, desenvolvida no Departamento de Zootecnia da Universidade Federal de Viçosa, em Viçosa, MG, obtidos no período de novembro de 1998 a julho de 2001.

Foram formadas, inicialmente, duas famílias provenientes do cruzamento de dois varrões da raça nativa brasileira (Piau) com 18 fêmeas originadas de linhagem desenvolvida na UFV, pelo acasalamento de animais das raças Landrace, Large White e Pietran. A geração F1 nasceu nos meses de março a maio de 1999 e era formada por 106 fêmeas e 134 machos. Foram selecionados 11 varrões F1 provenientes de diferentes leitegadas, os quais foram acasalados (monta natural) com 54 fêmeas, evitando-se o acasalamento de parentes. Os acasalamentos ocorreram entre os meses de fevereiro e outubro de 2000, para produção da geração F2, que nasceu entre junho de 2000 e fevereiro de 2001. Assim, foram obtidos 844 animais F2, divididos em cinco lotes: lote 1, composto por animais nascidos entre 20/06/00 e 03/07/00; lote 2, animais nascidos entre 03/08/00 e 23/08/00; lote 3, animais nascidos entre 16/09/00 e 01/11/00; lote 4, animais nascidos entre 30/11/00 e 25/12/00; e lote 5, animais nascidos entre 19/01/01 e 12/02/01. Os três primeiros lotes foram obtidos de fêmeas do primeiro parto e os dois últimos, de fêmeas de segunda parição.

Informações detalhadas sobre o manejo dos animais foram descritas por Pires (2003).

Após ser feita a consistência dos dados, 435 observações foram utilizadas na análise. Preliminarmente, foi realizado um teste para diagnóstico do efeito da multicolinearidade ou dependência linear entre as variáveis, o qual pode levar à formação de matrizes singulares ou mal condicionadas. Após essa análise, foi identificada e descartada a variável peso aos 105 dias de idade, visto que provocou forte multicolinearidade nas características de desempenho.

Foram avaliadas as seguintes características de desempenho: tamanho da leitegada ao nascer (TLN), tamanho da leitegada na desmama – 21

Seleção de variáveis de desempenho...

dias (TLD), número de tetos (NT), peso ao nascer (PN), pesos aos 21 (P21), 42 (P42), 63 (P63) e 77 (P77) dias de idade, ganho de peso médio diário (GPD), consumo de ração (CR) e conversão alimentar, dos 77 aos 105 dias de idade (CA).

Os dados submetidos à análise de componentes principais foram previamente ajustados para efeitos fixos e covariáveis. O modelo estatístico adotado foi:

$$y_{ijk} = \mu + S_i + L_j + b(C_{ijk} - \bar{C}) + e_{ijk}, \text{ em que:}$$

y_{ijk} = característica observada no animal k , do lote j e do sexo i ;

μ = constante inerente a toda observação;

S_i = efeito fixo do sexo i , $i = 1, 2$ (1 = macho castrado e 2 = fêmea);

L_j = efeito fixo do lote j , $j = 1, 2, 3, 4, 5$;

C_{ijk} = valor observado da covariável, no animal k , do lote j , do sexo i ;

\bar{C} = média da covariável;

b = coeficiente de regressão linear da característica, em função da covariável;

e_{ijk} = erro aleatório associado a cada observação.

Com base no trabalho de Pires (2003), para ajuste dos dados, foram utilizadas as seguintes características e respectivas covariáveis: PN – TLN; P21, P42, P63 e P77 – TLD; CR, GPD e CA – P77.

O método de análise de componentes principais, discutido por Barbosa (2003), a partir da matriz de correlação, consiste em transformar um conjunto de variáveis Z_1, Z_2, \dots, Z_p em novo conjunto de variáveis $Y_1 (CP_1), Y_2 (CP_2), \dots, Y_p (CP_p)$, não-correlacionadas entre si e arranjadas numa ordem decrescente de variância. A idéia principal desse procedimento é de que poucos, dentre os primeiros componentes principais, contenham a maior variabilidade dos dados originais, assim, os demais componentes podem ser descartados, o que reduz o número de variáveis.

Em razão do grande número de variáveis medidas em unidades diferentes, foi necessária a padronização dessas variáveis X_j ($j = 1, 2, \dots, p$);

nesse caso, a estrutura de dependência de X_j foi dada pela matriz de correlação R .

Para descarte, a variável que domina (aquela com maior correlação) no componente principal de menor autovalor (menor variância) deve ser menos importante para explicar a variância total e, portanto, passível de descarte (Regazzi, 2002).

Para o descarte de variáveis, adotou-se a recomendação de Jolliffe (1972, 1973) pela qual o número de variáveis descartadas deveria ser igual ao número de componentes cuja variância (autovalor) fosse inferior a 0,7.

Utilizou-se o programa SAS System for Windows^{NT}, versão 8.0, licenciado pela Universidade Federal de Viçosa (SAS, 1999).

RESULTADOS E DISCUSSÃO

O número de observações, as médias corrigidas e os desvios-padrão das características de desempenho são apresentados na Tab. 1.

Tabela 1. Número de observações, médias corrigidas e desvios-padrão das características de desempenho

Característica	Número de observações	Média corrigida	Desvio-padrão *
TLN (n°)	798	10,42052	2,13024
TLD (n°)	713	9,27431	1,89787
NT (n°)	804	13,19901	1,28136
PN (kg)	798	1,18584	0,26711
P21 (kg)	584	4,94395	1,10331
P42 (kg)	595	8,38423	1,81944
P63 (kg)	578	16,54801	3,14553
P77 (kg)	535	21,55254	3,96670
CR (kg)	595	40,29999	6,52833
GPD (kg)	585	0,53587	0,12155
CA (kg/kg)	580	2,80183	0,62664

TLN - tamanho da leitegada ao nascer; TLD - tamanho da leitegada no desmame; NT - número de tetos; PN - peso ao nascer, P21 - peso aos 21, P42 - peso aos 42, P63 - peso aos 63 e P77 - peso aos 77 dias de idade; GPD - ganho de peso médio diário; CR - consumo de ração; e CA - conversão alimentar, dos 77 aos 105 dias de idade.

Desvio-padrão dos dados corrigidos para efeitos fixos e covariáveis.

Os resultados obtidos para os componentes principais, os autovalores e as percentagens da variância explicada pelos componentes são apresentados na Tab. 2.

Os três primeiros componentes principais explicaram 67,9% da variação total. Seis variáveis (54,5% das 11 características avaliadas) podem ser descartadas, de acordo com o critério de Jolliffe (1972, 1973), pois o número de componentes que apresentaram autovalores menores que 0,7 foi seis.

Tabela 2. Componentes principais (CP), autovalores (λ_i) e percentagem da variância explicada pelos componentes (% var. CP) das características de desempenho

Componente principal	λ_i	% var. CP	% var. CP (acumulada)
CP ₁	3,66645313	33,33	33,33
CP ₂	2,03806818	18,53	51,86
CP ₃	1,76553181	16,05	67,91
CP ₄	0,98916384	8,99	76,90
CP ₅	0,90767612	8,25	85,15
CP ₆	0,65728174	5,98	91,13
CP ₇	0,38070829	3,46	94,59
CP ₈	0,23790821	2,16	96,75
CP ₉	0,19937425	1,81	98,57
CP ₁₀	0,10214737	0,93	99,49
CP ₁₁	0,05568705	0,51	100

Strapasson et al. (2000), ao utilizarem a análise de componentes principais em forrageiras do gênero *Paspalum*, observaram redução de 53, 68 e 43% nos descritores reprodutivos, vegetativos e agrônômicos, respectivamente, do conjunto inicialmente considerado.

As seis variáveis que apresentaram maiores coeficientes, em valor absoluto, a partir do último componente principal, são passíveis de descarte, conforme apresentado na Tab. 3. A razão para isso é que variáveis altamente correlacionadas com os componentes principais de menores variâncias representam variação praticamente insignificante. As variáveis passíveis de descarte são, respectivamente, em ordem de menor importância para explicar a variação total: GPD, P63, P42, TLD, P21 e PN. Com base nesses resultados, recomendam-se as seguintes variáveis para serem mantidas em experimentos futuros: TLN, NT, P77, CR e CA.

Do ponto de vista econômico, a CA deve receber atenção especial, pois a alimentação representa

considerável parte do custo total de produção de suínos. Uma restrição à avaliação dessa característica, para fins de melhoramento genético, é o custo elevado de sua mensuração, uma vez que os animais têm de ser criados em baias individuais e o fornecimento de ração deve ser anotado. Mesmo assim, as referências encontradas justificam o custo de mensuração, visto que o ganho genético obtido nessa característica é compensador (Torres Filho, 2001).

Pires (2003) encontrou QTL (loco de característica quantitativa) significativo para consumo de ração. Segundo o autor, um mesmo gene ou grupo gênico localizado em torno de 100cM do cromossomo 6 pode atuar nas características CR e GPD, o que sugere correlação significativa entre essas duas variáveis. Neste trabalho, a correlação foi de 0,62 (Tab. 4), o que explica, em parte, o fato de a variável GPD ter sido sugerida para descarte.

Apesar de terem sido avaliadas cinco pesagens, apenas uma foi selecionada, sendo as demais redundantes. Além disso, a redução no número de pesagens causaria menos estresse ao animal na avaliação dessas características.

O tamanho da leitegada é o principal componente de produtividade da matriz, e seu melhoramento genético é de importância crescente em mercados nos quais outras características, como espessura de toucinho, estão alcançando níveis satisfatórios (Sorensen, 1991), ou em mercados em que o aumento na produtividade da porca reduz, substancialmente, os custos de produção. No aspecto econômico, o tamanho da leitegada é uma característica importante, já considerada em vários programas de melhoramento (Rydhmer, 2000). Como os tamanhos da leitegada ao nascer e na desmama são fortemente correlacionados (Tab. 4), a análise de componentes principais sugeriu, para descarte, o TLD, por ela ser redundante. Omtvedt et al. (1966) verificaram correlação positiva entre TLN e TLD ($r_p=0,70$). Johansson e Kennedy (1983) relataram correlações genéticas positivas entre o tamanho da leitegada ao nascer e no desmame.

Seleção de variáveis de desempenho...

Tabela 3. Coeficientes de ponderação (CP) das características e suas correlações (em porcentagem) com os seis últimos componentes principais menos importantes para explicar a variação total de 11 características de desempenho de suínos

Característica	Coeficiente de ponderação (Correlação - %)					
	CP ₁₁	CP ₁₀	CP ₉	CP ₈	CP ₇	CP ₆
TLN	-0,063016 (-1,487)	-0,003536 (-0,113)	0,170324 (7,605)	0,656094 (32,002)	0,133733 (8,252)	0,120510 (9,770)
TLD	0,080231 (1,893)	0,032339 (1,034)	-0,170240 (-7,601)	-0,674098 (-32,880)	-0,115558 (-7,130)	0,003342 (0,271)
NT	0,008138 (0,192)	0,006693 (0,214)	0,012571 (0,561)	-0,059993 (-2,926)	-0,017670 (-1,090)	0,167317 (13,565)
PN	-0,028369 (-0,669)	-0,014371 (-0,459)	0,068005 (3,037)	-0,001533 (-0,075)	0,311037 (19,191)	-0,869807 (-70,518)
P21	0,067130 (1,584)	0,002410 (0,077)	-0,375472 (-16,765)	0,264599 (12,906)	-0,754422 (-46,549)	-0,127693 (-10,352)
P42	-0,113538 (-2,679)	0,240499 (7,686)	0,775471 (34,626)	-0,165141 (-8,055)	-0,199369 (-12,301)	0,160086 (12,979)
P63	0,030643 (0,723)	-0,782567 (25,011)	-0,063809 (-2,849)	-0,059501 (-2,902)	0,251109 (15,494)	0,273581 (22,180)
P77	0,047853 (1,129)	0,572141 (18,286)	-0,405022 (-18,085)	0,031015 (1,513)	0,444946 (27,454)	0,284868 (23,095)
CR	-0,429773 (-10,142)	-0,027883 (-0,891)	-0,064364 (-2,874)	0,032115 (1,566)	-0,027363 (-1,688)	0,036083 (2,925)
GPD	0,710125 (16,758)	0,016526 (0,528)	0,119256 (5,325)	0,041187 (2,009)	-0,003996 (-0,247)	0,025803 (2,092)
CA	0,528311 (12,467)	-0,006158 (0,197)	0,091722 (4,096)	0,083764 (4,086)	0,031736 (1,958)	-0,031788 (2,577)

TLN - tamanho da leitegada ao nascer, TLD - tamanho da leitegada no desmame, NT - número de tetos, PN - peso ao nascer, P21 - peso aos 21, P42 - peso aos 42, P63 - peso aos 63 e P77 - peso aos 77 dias de idade, GPD - ganho médio de peso diário, CR - consumo de ração, CA - conversão alimentar dos 77 aos 105 dias de idade.

Variáveis em negrito foram descartadas.

Tabela 4. Coeficientes de correlação simples entre as características de desempenho de suínos

	TLN	TLD	NT	PN	P21	P42	P63	P77	CR	GPD	CA
TLN	1,0000										
TLD	0,7520	1,0000									
NT	0,0099	-0,0367	1,0000								
PN	0,0407	0,1010	0,1366	1,0000							
P21	0,0006	0,0693	0,1222	0,5381	1,0000						
P42	-0,0281	0,0343	0,0893	0,4764	0,7496	1,0000					
P63	-0,0026	0,0265	0,1083	0,4688	0,6945	0,8308	1,0000				
P77	-0,0257	-0,0154	0,1057	0,4544	0,6397	0,7649	0,8848	1,0000			
CR	0,0565	0,1169	0,0271	0,0504	-0,0094	-0,0101	-0,0080	0,0128	1,0000		
GPD	0,0273	-0,0007	0,0191	0,0636	-0,0345	-0,0193	-0,0144	-0,0135	0,6239	1,0000	
CA	0,0061	0,0468	-0,0162	-0,0407	-0,0018	0,0236	-0,0126	-0,0032	-0,0808	-0,7530	1,0000

TLN - tamanho da leitegada ao nascer, TLD - tamanho da leitegada no desmame, NT - número de tetos, PN - peso ao nascer, P21 - peso aos 21, P42 - peso aos 42, P63 - peso aos 63 e P77 - peso aos 77 dias de idade, GPD - ganho médio de peso diário, CR - consumo de ração, CA - conversão alimentar dos 77 aos 105 dias de idade.

Adicionalmente, observou-se que as variáveis sugeridas para descarte apresentaram correlação linear simples significativa com as demais, ou seja, foram redundantes, e as correlações entre as variáveis selecionadas foram menores (Tab. 4). Desse modo, cada variável selecionada deve ser responsável por um tipo de informação biológica exclusiva, e a ação conjunta delas será complementar para descrição geral da população ou dos indivíduos estudados.

CONCLUSÕES

Em razão do grande número de variáveis redundantes, 54,5% delas podem ser descartadas, resultando em economia de tempo e custo do experimento, sem que haja perda considerável da informação.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARBOSA, L. *Utilização de técnicas de análise multivariada na avaliação de características quantitativas de uma população F2 de suínos*. 2003. 80f. Dissertação (Mestrado em Zootecnia) – Universidade Federal de Viçosa, Viçosa, MG.
- JOHANSSON, K.; KENNEDY, B.W. Estimation of genetics parameters for reproduction traits in pigs. *Acta Agric. Scand.*, v.35, p.421-431, 1983.
- 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.
- LIBERATO, J.R. *Aplicação de técnica de análise multivariada em fitopatologia*. 1995. 144f. Dissertação (Mestrado em Fitopatologia) – Universidade Federal de Viçosa, Viçosa, MG.
- OMTVEDT, I.T.; WHATLEY Jr., J.A.; WILLHAM, R.L. Some production factors associated with weaning records in swine. *J. Anim. Sci.*, v.25, p.373-376, 1966.
- PIRES, A.V. *Mapeamento de locos de características quantitativas, no cromossomo seis suíno*. 2003. 86f. Dissertação (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.
- REGAZZI, A.J. *Análise multivariada*. Viçosa: Universidade Federal de Viçosa, 2002. (INF-766) (notas de aula).
- RYDHMER, L. Genetics of sow reproduction, including puberty, oestrus, pregnancy, farrowing and lactation. *Livest. Prod. Sci.*, v.66, p.1-12, 2000.
- SAKAGUTI, E.S. *Utilização de técnicas de análise multivariada na avaliação de cruzamento dialélicos em coelhos*. 1994. 170f. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.
- SAS. Software: versão 8.0. Cary, NC: SAS Institute Inc., 1999.
- SORENSEN, D.A. Predicted breeding values for litter size with an animal model used in the Danish pig breeding program. Report from the Natl. Inst. *Anim. Sci.*, v.1, p.27, 1991.
- STRAPASSON, E.; VENCOVSKY, R.; BATISTA, L.A.R. Seleção de descritores na caracterização de germoplasma de paspalum sp. por meio de componentes principais. *Rev. Bras. Zootec.*, v.29, p.373-381, 2000.
- TORRES FILHO, R.A. *Avaliação genética de características de desempenho e reprodutivas em suínos*. 2001. 79f. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.