



Análise de componentes principais em características de produção de aves de postura

André Luis da Costa Paiva¹, Rafael Bastos Teixeira¹, Marcos Yamaki¹, Gilberto Romeiro de Oliveira Menezes¹, Carla Daniela Sugimoto Leite¹, Robledo de Almeida Torres¹

¹ Universidade Federal de Viçosa-MG.

RESUMO - Para avaliar a possibilidade de descarte de variáveis de produção em 942 aves de postura por meio de componentes principais, visando eliminar características redundantes e de difícil mensuração, foram utilizados os dados obtidos de linhagens de aves de postura do Programa de Melhoramento Genético da Universidade Federal de Viçosa. As características analisadas foram: taxa de postura da 26^a à 58^a semana de idade (TP); peso médio individual na 34^a (PMI₁), na 42^a (PMI₂), na 50^a (PMI₃), na 58^a (PMI₄) e na 66^a semana (PMI₅); e peso médio do ovo na 34^a (PMO₁), na 42^a (PMO₂), na 50^a (PMO₃), na 58^a (PMO₄) e na 66^a semana de idade (PMO₅). Dos 11 componentes principais, oito apresentaram variância inferior a 0,7 (autovalor inferior a 0,7), o que sugere oito variáveis para descarte. As variáveis descartadas foram aquelas com maiores coeficientes, em valor absoluto, a partir do último componente principal, uma vez que variáveis altamente correlacionadas aos componentes principais de menor variância representam variação praticamente insignificante. As variáveis descartadas apresentaram correlação linear simples significativa com as demais, ou seja, foram redundantes. Com base nesses resultados, recomendam-se as seguintes variáveis para utilização em experimentos futuros: TP, PMI₁ e PMO₄.

Palavras-chave: análise multivariada, correlação, descarte de variáveis, Legorne

Principal component analysis in laying hen production traits

ABSTRACT - To assess the possibility of discarding production variables in 942 laying hens by principle component analysis to eliminate unnecessary and difficult to measure characteristics, data were obtained from laying hen lines of the genetic breeding program of the Universidade Federal de Viçosa. The traits analyzed were egg production rate (TP) from the 26th to the 58th week, individual mean weight at the 34th (PMI₁), 42th (PMI₂), 50th (PMI₃), 58th (PMI₄) and 66th week (PMI₅); egg mean weight at the the 34th (PMO₁), 42nd (PMO₂), 50th (PMO₃), 58th (PMO₄) and 66th weeks of age (PMO₅). Eight of the 11 principal components showed variance lower than 0.7 (eigenvalue lower than 0.7), suggesting 8 variables to discard. The discarded variables were those that showed the highest coefficients, in absolute value, from the last principal component because variables highly correlated with the principal components of smaller variance represent practically insignificant variation. The discarded variables presented significant simple linear correlation with the others, therefore, they were redundant. Based on these results, the following variables are recommended for use in future experiments: TP, PMI₁, PMO₄.

Key Words: correlations, discarding of variables, Leghorn, multivariate analysis

Introdução

A técnica de componentes principais, segundo Cruz & Regazzi (1994), foi originalmente descrita por Pearson (1901) e posteriormente aplicada por Hotelling (1933) em diversas áreas da ciência. De acordo com Manly (1986), o uso desta técnica só foi acentuado quando houve disponibilidade de recursos na área computacional. O método consiste na transformação do conjunto original de variáveis em outro, os componentes principais (CP), de dimensões equivalentes, porém com a vantagem de que

cada componente retém porcentagem da variância original e que as variâncias decrescem do primeiro ao último componente principal (Morrison, 1976).

Segundo Baker et al. (1988), a análise dos componentes principais pode revelar relações não-identificadas previamente, contribuindo para melhor interpretação dos dados.

Decisões acerca da seleção de animais para determinada característica podem ser tomadas de forma intuitiva, negligenciando seus componentes e as ligações entre eles (Roso & Fries 1995). Nesse contexto, a análise de componentes principais deve contribuir na interpretação

das relações entre as variáveis e consequentemente nas decisões (Baker et al., 1988; Roso & Fries, 1995).

A correlação entre as variáveis e as informações obtidas por análises univariadas pode ser incompleta, principalmente quando há correlação entre as variáveis. Nesses casos, é de grande interesse o uso de análise multivariada, pois essa análise combina as informações múltiplas provenientes da unidade experimental. Em muitas situações, os pesquisadores tendem a avaliar maior número de características, gerando, com isso, acréscimo considerável de trabalho. Quando o número de características é elevado, muitas delas podem contribuir pouco para a discriminação dos indivíduos avaliados. Essa situação aumenta o trabalho de caracterização, mas não melhora a precisão, além de tornar mais complexa a análise e interpretação dos dados. Assim, podem-se eliminar aquelas características redundantes e de difícil mensuração, o que reduziria o tempo e os custos de experimentos.

Este trabalho foi realizado com o objetivo de reduzir a dimensionalidade, por meio da análise de componentes principais, do conjunto original de variáveis com a menor perda de informação possível, eliminando as informações redundantes e reduzindo o custo e tempo das pesquisas.

Material e Métodos

Os dados utilizados foram provenientes de linhagens de postura de Legorne do Programa de Melhoramento Genético da Universidade Federal de Viçosa (UFV) totalizando 942 animais avaliados. As aves foram alojadas em gaiolas individuais às 24 semanas de idade. Inicialmente, foi realizado o teste para diagnóstico do efeito da multicolinearidade ou dependência linear entre as variáveis, que pode levar à formação de matrizes singulares ou mal condicionadas. Após essa análise, não foi diagnosticado efeito problemático de multicolinearidade.

As características analisadas foram: taxa de postura da 26^a à 58^a semana de idade (TP), peso médio individual na 34^a (PMI₁), na 42^a (PMI₂), na 50^a (PMI₃), na 58^a (PMI₄) e na 66^a semana (PMI₅); peso médio do ovo na 34^a (PMO₁), na 42^a (PMO₂), na 50^a (PMO₃), na 58^a (PMO₄) e na 66^a semana de idade (PMO₅).

O método de análise de componentes principais, a partir da matriz de correlação, consiste em transformar um conjunto de variáveis Z_1, Z_2, \dots, Z_p em um novo conjunto de variáveis $Y_1 (CP_1), Y_2 (CP_2), \dots, Y_p (CP_p)$ (Regazzi, 2002). Dessa forma, um novo conjunto de p variáveis não-correlacionadas entre si e arranjadas numa ordem decrescente de variâncias é definido. A ideia principal com esse procedimento é de que poucos, entre os primeiros

componentes principais, contenham a maior variabilidade dos dados originais; contudo, pode-se racionalmente descartar os demais componentes, reduzindo o número de variáveis.

Tendo em vista o grande número de variáveis medidas em unidades diferentes, foi necessária a padronização das variáveis originais X_j ($j = 1, 2, \dots, p$) em Z_j . Após essa transformação, todas as variáveis originais têm a mesma variância unitária e, dessa maneira, a base para extração dos componentes principais se torna a matriz de correlação R . Neste caso, os coeficientes dos componentes principais são indicadores da magnitude da correlação entre as variáveis padronizadas e os componentes principais, pois a correlação pode ser obtida pela seguinte expressão:

$$r_{CP_i Z_j} = \sqrt{\lambda_i} \times a_{ij}$$

em que λ_i = autovalor associado ao CP_i , que é uma constante; a_{ij} = coeficiente do CP_i relacionado à variável Z_j .

O critério para descarte de variáveis utilizado foi recomendado por Jolliffe (1973) para situações em que se trabalha com dados simulados e reais, com a análise de componentes principais a partir da matriz de correlação (Regazzi, 2002). Esse critério estabelece que o número de variáveis descartadas deve ser igual ao número de componentes cuja variância (autovalor) é inferior a 0,7.

No descarte de variáveis, a variável que apresenta o maior coeficiente (valor absoluto) no componente principal de menor autovalor (menor variância) deve ser menos importante para explicar a variância total e, portanto, passível de descarte (Regazzi, 2002). Assim, o processo de descarte consiste em considerar o autovetor (coeficientes do CP) correspondente ao menor autovalor e rejeitar a variável associada ao maior coeficiente (valor absoluto). Então, o próximo menor autovetor é avaliado. Esse processo continua até que o autovetor associado ao autovalor inferior a 0,7 seja considerado. A razão para isso é que variáveis altamente correlacionadas aos componentes principais de menor variância representam variação praticamente insignificante (Mardia et al., 1997).

Todas as análises foram feitas utilizando-se o programa SAS versão 8.0, licenciado pela Universidade Federal de Viçosa.

Resultados e Discussão

Com base nos resultados obtidos para os componentes principais, seus respectivos autovalores e porcentagens da variância explicada por cada um (Tabela 1), dos 11 componentes principais, oito (72,2%) apresentaram variância inferior a 0,7 (autovalor inferior a 0,7).

Tabela 1 - Componentes principais (CP), autovalores (λ_i) e porcentagem da variância explicada pelos componentes (% VCP) das características de produção

Componente principal	λ_i^*	% VCP	% VCP (acumulada)
CP ₁	4,98	0,45	0,45
CP ₂	2,54	0,23	0,68
CP ₃	0,97	0,09	0,77
CP ₄	0,56	0,05	0,82
CP ₅	0,40	0,04	0,86
CP ₆	0,33	0,03	0,89
CP ₇	0,30	0,03	0,92
CP ₈	0,28	0,02	0,95
CP ₉	0,23	0,02	0,97
CP ₁₀	0,21	0,02	0,99
CP ₁₁	0,17	0,01	1,00

* Valores em negrito indicam os autovalores superiores a 0,7.

As oito variáveis que apresentaram maiores coeficientes, em valor absoluto, a partir do último componente principal, são passíveis de descarte (Tabela 2). As variáveis sugeridas para descarte são, respectivamente, em ordem de menor importância para explicar a variação total: PMI₂, PMO₃, PMI₅, PMI₄, PMI₃, PMO₂, PMO₅, PMO₁.

Em estudo de divergência genética entre acessos de capim-elefante, Daher et al. (1997) utilizaram a técnica de componentes principais e observaram que, de um total de 22 caracteres avaliados em três anos, apenas oito (36,4%) foram selecionados como os mais importantes para

determinação da divergência genética. Strapasson et al. (2000), selecionando descritores na caracterização de germoplasma de *Paspalum sp.* por meio de componentes principais com o intuito de descartar os descritores considerados redundantes ou não-discriminantes, selecionaram oito descritores, considerados mais importantes na descrição da variabilidade presente na coleção de acessos do germoplasma estudado.

Neste trabalho, as características sugeridas para descarte apresentaram correlação linear simples significativa com as demais, ou seja, são redundantes; por outro lado, as variáveis selecionadas apresentaram menor correlação entre si (Tabela 3).

Com base nesses resultados, recomendam-se, para serem mantidas em experimentos futuros, as variáveis taxa de postura da 26^a a 58^a semana, peso médio individual na 34^a semana, e peso médio do ovo na 58^a semana. A variável taxa de postura * semana sempre foi uma característica importante nos programas de melhoramento de aves de postura por ter relação com a produção de ovos. A importância da característica PMI₁ pode ser justificada pela coincidência em corresponder ao mesmo período do pico de produção de ovos. Nesta fase a exigência das aves é muito alta, em decorrência da produção de ovos, e ao mesmo tempo é um período em que as aves não alcançaram sua maturidade física. Já a PMO₄ relaciona-se à capacidade

Tabela 2 - Coeficientes dos 11 componentes principais (autovetores)¹

Característica ²	Coeficiente										
	CP ₁	CP ₂	CP ₃	CP ₄	CP ₅	CP ₆	CP ₇	CP ₈	CP ₉	CP ₁₀	CP ₁₁
TP	0,025 (0,06)	-0,170 (-0,27)	0,971 (0,96)	-0,101 (-0,08)	0,041 (0,03)	0,031 (0,02)	0,089 (0,05)	0,078 (0,04)	0,008 (0,01)	0,023 (0,01)	0,022 (0,01)
PMO ₁	0,230 (0,51)	0,334 (0,53)	0,120 (0,12)	0,834 (0,62)	0,213 (0,14)	0,245 (0,14)	-0,031 (-0,02)	0,088 (0,05)	-0,025 (-0,01)	-0,030 (-0,01)	-0,019 (-0,01)
PMO ₂	0,294 (0,66)	0,335 (0,53)	0,153 (0,15)	0,053 (0,04)	-0,491 (-0,31)	-0,468 (-0,27)	-0,373 (-0,20)	-0,355 (-0,19)	0,194 (0,09)	-0,075 (-0,03)	0,081 (0,03)
PMO ₃	0,303 (0,68)	0,334 (0,53)	0,012 (0,01)	-0,215 (-0,16)	-0,205 (-0,13)	0,223 (0,13)	0,270 (0,15)	-0,115 (-0,06)	-0,426 (-0,20)	0,593 (0,27)	-0,135 (-0,06)
PMO ₄	0,293 (0,65)	0,362 (0,58)	-0,021 (-0,02)	-0,301 (-0,22)	-0,216 (-0,50)	0,249 (0,14)	0,159 (0,09)	0,527 (0,28)	0,158 (0,08)	-0,505 (-0,23)	0,002 (0,01)
PMO ₅	0,285 (0,64)	0,319 (0,51)	0,004 (0,01)	-0,316 (-0,24)	0,785 (-0,03)	-0,256 (-0,15)	-0,080 (-0,04)	-0,101 (-0,05)	0,112 (0,05)	0,025 (0,01)	0,066 (0,03)
PMI ₁	0,348 (0,78)	-0,263 (-0,42)	-0,092 (-0,09)	0,163 (0,12)	-0,043 (-0,03)	-0,422 (-0,24)	0,466 (0,25)	0,188 (0,01)	-0,161 (-0,08)	0,025 (0,01)	0,561 (0,23)
PMI ₂	0,350 (0,78)	-0,293 (-0,47)	-0,047 (-0,05)	0,087 (0,06)	0,007 (0,01)	-0,357 (-0,21)	0,012 (0,01)	0,277 (0,15)	0,014 (0,01)	0,057 (0,03)	-0,757 (-0,31)
PMI ₃	0,336 (0,75)	-0,290 (-0,46)	-0,028 (-0,03)	-0,095 (-0,07)	0,030 (0,02)	0,204 (0,12)	-0,673 (-0,37)	0,171 (0,09)	-0,475 (-0,23)	-0,048 (-0,02)	0,202 (0,08)
PMI ₄	0,351 (0,78)	-0,256 (-0,41)	-0,025 (0,03)	-0,025 (-0,02)	0,040 (0,03)	0,247 (0,14)	0,268 (0,15)	-0,646 (-0,34)	-0,092 (-0,04)	-0,465 (-0,21)	-0,137 (-0,06)
PMI ₅	0,348 (0,78)	-0,277 (-0,44)	-0,081 (-0,08)	-0,027 (-0,02)	-0,044 (-0,03)	0,346 (0,20)	-0,072 (-0,04)	-0,003 (0,01)	0,695 (0,33)	0,403 (0,18)	0,148 (0,06)

¹ Em negrito, os oito menos importantes para explicar a variação total e abaixo, entre parênteses, as correlações entre as variáveis e os componentes principais.

² TP - taxa de postura da 26^a a 58^a semana, PMI₁ - peso médio individual na 34^a semana, PMI₂ - peso médio individual na 42^a semana, PMI₃ - peso médio individual na 50^a semana, PMI₄ - peso médio individual na 58^a semana, PMI₅ - peso médio individual na 66^a semana; PMO₁ - peso médio do ovo 34^a semana, na 42^a semana, PMO₂ - peso médio do ovo, PMO₃ - peso médio do ovo na 50^a semana, PMO₄ - peso médio do ovo na 58^a semana e PMO₅ - peso médio do ovo na 66^a semana.

Tabela 3 - Coeficientes de correlação simples entre as características avaliadas

Característica ¹	TP	PMO ₁	PMO ₂	PMO ₃	PMO ₄	PMO ₅	PMI ₁	PMI ₂	PMI ₃	PMI ₄	PMI ₅
TP	1,0000										
PMO ₁	-,0441	1,0000									
PMO ₂	0,0023	0,5786	1,0000								
PMO ₃	-,0911	0,5494	0,7071	1,0000							
PMO ₄	-,1103	0,5135	0,6755	0,7660	1,0000						
PMO ₅	-,0751	0,4920	0,5887	0,6667	0,6566	1,0000					
PMI ₁	0,0743	0,2036	0,2811	0,2717	0,2511	0,2581	1,0000				
PMI ₂	0,1208	0,1684	0,2730	0,2344	0,2332	0,2589	0,8094	1,0000			
PMI ₃	0,1336	0,1223	0,2410	0,2388	0,2347	0,2507	0,6952	0,7577	1,0000		
PMI ₄	0,1255	0,1803	0,2796	0,3089	0,2620	0,2898	0,7334	0,7340	0,7225	1,0000	
PMI ₅	0,0924	0,1588	0,2539	0,2728	0,2699	0,2546	0,7278	0,7634	0,7508	0,7590	1,0000

¹ De acordo com a Tabela 2.

das aves de manterem a produção de ovos com pesos de bom valor comercial até o final do primeiro ciclo de postura.

Barbosa et al. (2005), avaliando 11 características de desempenho em suínos, concluíram que seis variáveis (54,5%) analisadas foram redundantes e podem ser descartadas em experimentos futuros. Abreu et al. (1999), trabalhando com produção de ovos de matrizes de frango de corte, utilizando componentes principais, observaram que os dois primeiros componentes principais explicaram mais de 98% da variação total disponível entre as médias de cruzamentos disponíveis.

Mascioli et al. (2000) analisaram componentes principais para características de crescimento na raça Canchim e observaram que os dois primeiros foram responsáveis por mais de 75% da variação total, tanto para os ganhos como para os pesos, e concluíram que a variância fenotípica pode ser utilizada para o estudo dos componentes principais.

Conclusões

Características relacionadas à taxa de postura, ao peso médio individual e ao peso médio do ovo medidas em diferentes idades podem ser restritas a três componentes principais. Há possibilidade de redução do número de variáveis avaliadas em experimentos com aves de postura. Recomendam-se as seguintes variáveis para serem mantidas em estudos futuros: taxa de postura, peso médio individual na 34^a semana de idade e peso médio do ovo na 58^a semana de idade.

Referências

ABREU, N.M.V.; SILVA, M.A.; CRUZ, C.D. et al. Capacidade de combinação de características de produção de ovos de

linhagens de matrizes de corte usando componentes principais. **Revista Brasileira de Zootecnia**, v.28, n.5, p.955-959, 1999.

BAKER, J.F.; STEWART, T.S.; LONG, C.R. et al. Multiple regression and principal components analysis of puberty and growth in cattle. **Journal of Animal Science**, v.66, n.9, p.2147-2158, 1988.

BARBOSA, L.; LOPES, P.S.; REGAZZI, A.J. et al. Seleção de variáveis de desempenho de suínos por meio da análise de componentes principais. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.57, n.6, p.805-810, 2005.

CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. 1.ed. Viçosa, MG: UFV, Imprensa Universitária, 1994. 390p.

DAHER, R.F.; MORAES, C.F.; CRUZ, C.D. et al. Seleção de caracteres morfológicos discriminantes em capim-elfante (*Pennisetum purpureum* Schum.). **Revista Brasileira de Zootecnia**, v.26, n.2, p.247-254, 1997.

HOTELLING, H. Review of the triumph of mediocrity in business, by Horace Secrist. **Journal of American Statistical Association**, v.28, p.463-5, 1933.

JOLLIFFE, I.T. Discarding variables in a principal component analysis. II. Real data. **Applied Statistics**, v.22, p.21-31, 1973.

MANLY, B.F.J. **Multivariate statistical methods: a primer**. London: Chapman & Hall, 1986. 159p.

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. 6.ed. London: Academic Press, 1997. 518p.

MASCIOLI, A.S.; EL FARO, L.; ALENCAR, M.M. et al. Estimativas de parâmetros genéticos e fenotípicos e análise de componentes principais para características de crescimento na raça Canchim. **Revista Brasileira de Zootecnia**, v.29, n.6, p.1654-1660, 2000.

MORRISON, D.F. **Multivariate statistical methods**. 2.ed. Singapore: McGraw Hill, 1976. 415p.

PEARSON, K. On lines and planes of closest fit to system of point in space. **Philosophical Magazine**, v.2, n.6, p.550-572, 1901. Disponível em: <<http://pbil.univ-lyon1.fr/R/pearson190.pdf>>. Acesso em: 28/7/2006.

REGAZZI, A.J. **Análise multivariada**. Viçosa, MG: Universidade Federal de Viçosa, 2002. (INF-766 - notas de aula).

ROSO, V.M.; FRIES, L.A. Componentes principais em bovinos da raça Polled Hereford à desmama e sobreano. **Revista da Sociedade Brasileira de Zootecnia**, v.24, n.5, p.728-735, 1995.

STRAPASSON, E.; VENCOVSKY, R.; BATISTA, L.A.R. Seleção de descritores na caracterização de germoplasma de *Paspalum sp.* por meio de componentes principais. **Revista Brasileira de Zootecnia**, v.29, n.2, p.373-381, 2000.