

# A análise de fatores na predição de ganhos por seleção em milho (*Zea mays* L.)

Maria José Granate<sup>1\*</sup>, Cosme Damião Cruz<sup>2</sup>, Paulo Roberto Cecon<sup>3</sup> e Cleso Antônio Patto Pacheco<sup>4</sup>

<sup>1</sup>EPAMIG/CTZM, Vila Gianetti, 46, Campus Universitário, 36571-000, Viçosa, Minas Gerais, Brasil. <sup>2</sup>Departamento de Biologia Geral, Universidade Federal de Viçosa, 36571-000 Viçosa, Minas Gerais, Brasil. <sup>3</sup>Departamento de Informática, Universidade Federal de Viçosa, 36571-000 Viçosa, Minas Gerais, Brasil. <sup>4</sup>Embrapa-Centro Nacional de Pesquisa de Milho e Sorgo (CNPMS), C.P. 151, 35701-970 Sete Lagoas, Minas Gerais. \*Author for correspondence.

**RESUMO.** Apresenta-se um histórico e várias utilizações agronômicas da Análise de Fatores, em especial a sua utilização na predição de ganhos por seleção. Foram avaliadas 166 famílias de meios irmãos de milho pipoca no ano agrícola de 1997-1998, em Sete Lagoas, Estado de Minas Gerais, no Centro Nacional de Pesquisa de Milho e Sorgo. A predição dos ganhos por seleção, baseada nos escores obtidos através da análise de fatores, possibilitou ganhos nos caracteres principais. Os escores foram utilizados para a predição de ganhos por seleção direta e indireta e uso de vários índices de seleção. A análise de fatores foi considerada uma técnica multivariada útil para a predição de ganhos por seleção, em substituição do método tradicional, ou complementando as técnicas de seleção simultânea de caracteres baseadas em índices de seleção.

**Palavras-chave:** *Zea mays* L., melhoramento, diferencial de seleção, índices de seleção, análise de fatores, métodos multivariados.

**ABSTRACT. Analysis of prediction factors in yields by maize selection (*Zea mays* L.).**

The history and use of the Factor Analysis in agronomic studies, especially for the prediction of selection yields, is presented. The Factor Analysis was applied on data provided by the evaluation of 166 half-sib families of popcorn at the National Research Center in Maize and Sorghum, Sete Lagoas MG Brazil, in 1997-1998. Selection yield prediction based on scores from the Factor Analysis favored yields in the most important traits. Scores were used for yield prediction by direct and indirect selection and by several selection indices. The Factor Analysis is a useful multivariate method for prediction of selection yields and may either substitute the traditional method or supplement selection by selection indices.

**Key words:** *Zea mays* L., breeding, differential selection, selection index, factor analysis, multivariate methods.

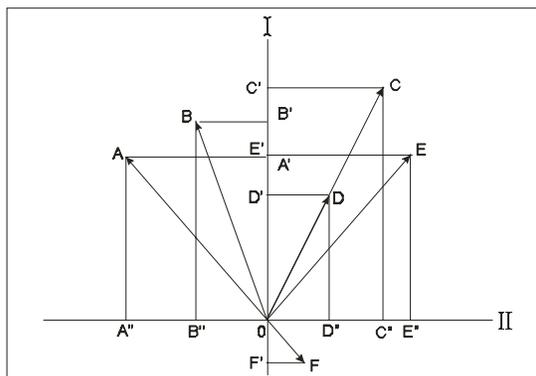
A Análise de Fatores (AF) é um método estatístico multivariado que vem sendo aplicado aos estudos agronômicos há relativamente pouco tempo e com várias finalidades. As imensas dificuldades dos cálculos retardaram a sua difusão junto à comunidade científica até à popularização dos computadores. Considerou-se conveniente apresentar um histórico da Análise de Fatores e algumas aplicações agronômicas recentes, seguidos pela exemplificação da sua utilização na predição de ganhos por seleção, sob duas metodologias.

Em 1904, Charles Spearman, Estatístico e Professor Titular de Psicologia da University College de Londres, desenvolveu a técnica da AF, como um processo que lhe permitisse identificar uma causa subjacente para as numerosas correlações positivas existentes entre os mais diversos testes

mentais aplicados à mesma pessoa. No seu método, as muitas variáveis iniciais são substituídas por um número menor de variáveis latentes, chamadas fatores. A cada fator corresponde um grupo das variáveis iniciais que estejam fortemente correlacionadas entre si, mas fracamente correlacionadas com as outras variáveis iniciais (Gould, 1996).

O sistema complexo de uma matriz de correlações pode ser comparado a um espaço multidimensional, e os fatores podem ser considerados eixos de influência comuns a vários conjuntos de variáveis independentes, inseridos dentro do espaço multidimensional da matriz. Com a aplicação da AF, esse espaço multidimensional pode ser reduzido a um espaço com muito menos dimensões, com perda mínima de informação e mais

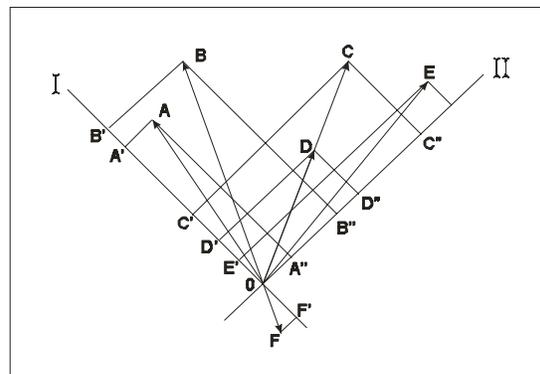
fácil de interpretar. Para isso, recorre-se à fatoração da matriz ao longo dos eixos referidos. Geometricamente, em modo bidimensional, a matriz de correlações pode ser representada por um feixe de vetores irradiando de um ponto central e a sua fatoração, pelas projeções dos vetores sobre eixos cartesianos, conforme se representa na Figura 1. A razão entre a projeção e o vetor representa a porcentagem da informação explicada por cada eixo. O eixo vertical neste exemplo é o que apresenta as maiores projeções, ou seja, é o que explica a maior parte da informação. Spearman chamou esse eixo de primeiro componente principal e o outro de segundo componente principal (Gould, 1996). Segundo esse modelo, duas observações altamente correlacionadas são representadas por vetores muito próximos. O valor do cosseno do ângulo entre esses vetores representa o coeficiente de correlação entre as variáveis originais. Quando os vetores se sobrepõem, como os vetores OC e OD na Figura 1, a correlação é máxima, e tem-se:  $\cos 0^\circ = 1 = \text{coeficiente de correlação}$ ,  $r = 1$ . Quando os vetores são perpendiculares, como os vetores OA e OE na Figura 1, representam observações independentes, e tem-se:  $\cos 90^\circ = 0 = \text{coeficiente de correlação}$ ,  $r = 0$ . Se os vetores são opostos, como os vetores OA e OF na Figura 3, tem-se:  $\cos 180^\circ = -1 = \text{coeficiente de correlação}$ ,  $r = -1$ . Isto é, quando os vetores têm a mesma direção representam correlações positivas, quando os vetores têm direções contrárias representam correlações negativas, quando os vetores se sobrepõem representam a correlação máxima e quando os vetores são perpendiculares representam a correlação nula.



**Figura 1.** Representação geométrica e bidimensional de uma matriz de correlações como um feixe de vetores. As projeções dos vetores representam os fatores da análise de fatores. Todos os vetores apresentam projeção maior sobre o eixo I

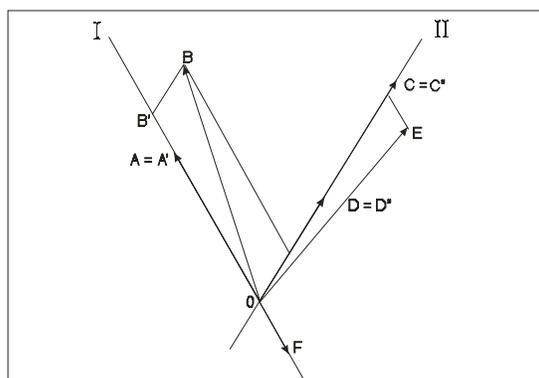
Um vetor muito perto de um dos eixos terá maior projeção sobre esse eixo que sobre outro mais afastado. Então, diz-se que aquele eixo explica a maior parte da informação contida no vetor. No exemplo da Figura 1, o eixo vertical ou primeiro componente principal explica a maior porcentagem da variação. O eixo horizontal ou segundo componente principal explica o restante da variação.

Tendo identificado um valor fortemente positivo no primeiro componente principal, ou seja, uma correlação subjacente existente nos testes de inteligência, impossível de identificar sem a AF, Spearman considerou ter encontrado a característica interior a todas as atividades mentais e que media a “inteligência geral” (Gould, 1996). No entanto, os eixos de Spearman são apenas um dos muitos modos de inserir eixos num espaço, e os vetores podem ser projetados sobre qualquer conjunto de eixos inseridos naquele espaço. Matematicamente, as diferentes projeções são equivalentes. Na Figura 2, os mesmos vetores existentes na Figura 1 são projetados sobre eixos que sofreram uma rotação. Agora as maiores projeções não coincidem sobre o mesmo eixo. Os vetores OA e OB têm maior projeção sobre o eixo I, o qual explica a maior porcentagem da respectiva variação. Os vetores OC e OD têm maior projeção sobre o eixo II, sendo este o eixo que explica a maior parte da respectiva variação. As projeções (ou fatores) formam agora dois grupos ortogonais, não correlacionados ou independentes. A rotação não traz aumento nem diminuição da informação, apenas a sua redistribuição (Gould, 1996).



**Figura 2.** Representação geométrica e bidimensional de uma matriz de correlações como um feixe de vetores. Comparando com a Figura 1, os eixos sofreram uma rotação. Nesta nova posição dos eixos, os vetores apresentam as maiores projeções repartidas em dois grupos, sobre os dois eixos. As projeções dos vetores representam os fatores da análise de fatores

Em 1935, L. L. Thurstone, Professor de Psicologia da Universidade de Chicago, além de ter inventado a rotação e a representação geométrica, também descobriu que os eixos não precisam ser perpendiculares, mas podem ser oblíquos, como se exemplifica na Figura 3. Deste modo, podem ser colocados ainda mais perto de cada grupo de vetores do que quando a rotação é feita, mantendo-os perpendiculares. Pelo mesmo motivo também recebem projeções ainda maiores. Existem vários critérios para posicionar os eixos. Thurstone escolheu o critério que permite que os eixos representem a máxima variância possível, o método Varimax. Esses “vetores da mente”, como denominados por Thurstone, representariam as “habilidades mentais primárias”: para matemática, poesia, esportes, etc. Thomson (1939), citado por Gould (1996), adversário de Spearman, comentou que o pesquisador pode rodar os eixos para qualquer posição e que, em consequência, os mesmos testes mentais, submetidos às mesmas análises, permitem interpretações opostas. Os fatores são, então, apenas abstrações matemáticas convenientes (Burt, 1937, citado por Gould, 1996). Devido à rotação e à fatoração, um número infinito de resultados pode ser obtido (Abreu, 1997). As variáveis reunidas em cada fator são as que melhor explicam o fenômeno biológico relativo a esse fator.



**Figura 3.** Representação geométrica e bidimensional de uma matriz de correlações como um feixe de vetores. Os eixos não são perpendiculares e, em comparação com a Figura 1 e a Figura 2, sofreram uma nova rotação. Nesta posição dos eixos, os vetores também apresentam as projeções repartidas em dois grupos sobre os dois eixos. Os eixos coincidem com alguns vetores, obtendo-se a projeção máxima representando a correlação máxima. As projeções dos vetores representam os fatores da análise de fatores.

Atualmente, designa-se por Análise de Componentes Principais a análise sem a rotação dos eixos, e por Análise de Fatores a análise feita após a rotação dos eixos, apresentando Harris (1975) e

Johnson e Wichern (1992) o modo de proceder à análise e de realizar os cálculos.

As semelhanças entre os locais de experimentação e as regiões produtoras de trigo dos Estados Unidos foram determinadas submetendo os dados de trinta anos de ensaios à análise de fatores (Petersen, 1992). Em milho doce, a associação entre as características das sementes, da porcentagem de emergência, do teor em açúcares, da maciez e da suscetibilidade aos fungos, representadas por vinte e nove variáveis, foi estudada usando a análise de fatores e reduzidas a seis fatores (Juvik *et al.*, 1993). As relações entre a arquitetura da planta e o tamanho das sementes de feijoeiro, representadas por quatro variáveis, foram identificadas usando dois fatores e em outra análise de fatores, de quinze variáveis, relativas à arquitetura da planta, foram extraídos três fatores (Acquaah *et al.*, 1992). Em estudo do rendimento de sementes de feijoeiro, Ranalli *et al.* (1991) utilizaram sete variáveis relacionadas às características morfológicas e à produção as quais, após a análise de fatores, permitiram a extração de três fatores.

Para a utilização da AF no melhoramento ou na seleção de plantas, deve-se procurar o fator cujos elementos ou cargas fatoriais apresentem valores próximos da unidade em relação às características de interesse para o melhorista e aos valores próximos de zero para as outras. A partir desse fator se estimam os escores que são utilizados na seleção como se fossem os valores de uma nova variável relacionada com esse fator. Os dados de Abreu (1997) e Castaldi (1997) permitiram a obtenção do fator de produção, considerando essa a variável mais importante.

O consumo do milho pipoca vem aumentando nos últimos anos: na CEASA, Estado de Minas Gerais, em 1988, comercializaram-se 740 t e em 1997, 1256 t, das quais apenas 456 t foram produzidas em Minas Gerais, o que sugere a necessidade do melhoramento deste tipo de milho, visando obter variedades de alta produtividade e qualidade.

Uma das principais dificuldades deste melhoramento é a existência de correlação negativa entre as duas características mais importantes: a capacidade de expansão (relacionada com a qualidade da pipoca) e a produtividade, o que não permite obter predição de ganhos simultâneos nessas duas características com as metodologias tradicionais (Marques, 2000). A procura de uma metodologia alternativa, que permita a estimação de ganhos preditos simultâneos nessas duas características, constitui o objetivo deste trabalho, apresentando-se a utilização da Análise de Fatores com essa finalidade.

## Material e métodos

Utilizaram-se famílias de meios irmãos do composto de milho, *Zea mays* L. (Poaceae) CMS 43 - Composto Pipoca Branco Tipo Americano, obtido pela recombinação de materiais considerados superiores e tolerantes às principais doenças, existentes no Banco de Germoplasma do Centro Nacional de Pesquisa de Milho e Sorgo. Desde 1979, as populações do composto foram submetidas a mais de cinco ciclos de seleção massal estratificada e em seguida a vários ciclos de seleção entre e dentro de famílias de meios irmãos (Pacheco et al., 1992a, b; Gama, 1997). Os ensaios foram conduzidos no ano agrícola de 1997-1998, nos campos experimentais do CNPMS da Embrapa, em Sete Lagoas, Estado de Minas Gerais, em área homogênea.

Avaliaram-se 179 tratamentos, dos quais 166 famílias de meios irmãos do composto CMS 43 e como testemunhas 13 tratamentos com a variedade Zélia e a população de milho CMS 43 do ciclo de seleção anteriormente obtido. A quantidade de semente disponível apenas permitiu a formação de dois blocos, considerados homogêneos. Cada parcela constou de uma fileira de 5 m, com 50 sementes por parcela, em 25 covas espaçadas 0,20 m. O espaçamento entre fileiras foi de 0,90 m. O plantio, os tratos culturais e as adubações foram os habituais na cultura do milho. A colheita foi manual, após a maturação fisiológica. Os seguintes caracteres referem-se a cada parcela e foram avaliados na época da colheita: ALTP – altura de planta em cm, média das medições feitas do nível do solo a inserção da folha bandeira, em 6 plantas competitivas tomadas ao acaso, em cm, ALTES – altura da espiga em cm, média das distâncias do nível do solo até a inserção da espiga, nas mesmas 6 plantas, em cm, PRE – posição relativa da espiga, razão entre ALTES e ALTP, AC – número de plantas acamadas, contagem das plantas com ângulo de inclinação superior a 45° em relação à vertical, QUE – número de plantas cujo colmo se apresenta quebrado, EST – número de plantas existentes, NESP – número de espigas colhidas, PROLI – prolificidade, razão entre NESP e EST, PRODU – produtividade, peso dos grãos debulhados por unidade de área, expresso em kg/ha, CE – capacidade de expansão, razão entre o volume de pipoca expandida e o volume de grãos antes do pipocamento. Uma amostra de 30 mL de grãos foi tomada de cada parcela, medida em proveta graduada, estourada em pipoqueira elétrica, regulada automaticamente para 237° C, e o volume da pipoca expandida foi medido em proveta graduada de 1000 mL.

As análises de variâncias foram efetuadas no delineamento em blocos casualizados, com duas

repetições, considerando as famílias de meios irmãos de milho pipoca como efeito aleatório e as testemunhas como efeito fixo. Estimaram-se os parâmetros genéticos, verificando-se a existência de variabilidade genética que permite o melhoramento. Os coeficientes de variação estão dentro da faixa encontrada na maioria dos experimentos com variedades e características semelhantes a este. (Marques, 2000).

O modelo de análise de fatores utilizado é (Cruz, 1997):  $X_j = l_{j1}F_1 + l_{j2}F_2 + \dots + l_{jm}F_m + \varepsilon_j$  no qual  $X_j$  é a  $j$ -ésima variável estimada em cada parcela, com  $j=1,2,\dots,v$ ,  $l_{jk}$  é a carga fatorial para a  $j$ -ésima variável associada ao  $k$ -ésimo fator, sendo  $k=1,2,\dots,m$ ,  $F_k$  é o  $k$ -ésimo fator comum,  $\varepsilon_j$  é o fator específico.

Escolheu-se um número de fatores comuns, igual ao número de autovalores superiores à unidade existentes na matriz de correlações fenotípicas das variáveis e optou-se pelo modelo de fatores ortogonais (Castoldi, 1997; Abreu, 1997). A carga fatorial inicial é:  $l_{ij} = \lambda_i V_{ij}^2$ , sendo  $\lambda_i$  o  $i$ -ésimo autovalor maior que 1 obtido da matriz de correlações fenotípicas e  $V_{ij}$  o  $j$ -ésimo valor do  $i$ -ésimo vetor, sendo  $j$  o número de variáveis e  $k$  o número de fatores. A comunalidade representa-se por:  $C_j = l_{j1}^2 + l_{j2}^2 + \dots + l_{jm}^2$ . Após o cálculo das cargas fatoriais iniciais, os fatores foram submetidos a um máximo de 50 rotações pelo método Varimax, para obtenção das cargas fatoriais finais (Cruz, 1997).

A partir dos fatores finais, calcularam-se os escores fatoriais segundo o modelo (Castoldi, 1997):  $F^* = [(G'G)^{-1}G'X]$ , no qual  $F^*$  é o vetor de dimensão  $m \times 1$  de escores fatoriais,  $G$  é a matriz de dimensões  $v \times m$  das cargas fatoriais rotacionadas finais, em que cada uma das  $m$  colunas corresponde a um fator e  $X$  é o vetor de dimensão  $v \times 1$  das variáveis ou caracteres da  $k$ -ésima família. Para a formação do vetor  $X$ , as variáveis devem ser padronizadas (Cruz, 1997).

Este cálculo foi repetido para cada família e em cada repetição. Obtiveram-se assim os valores de  $m$  novas características, tantas quanto o número de fatores, designadas, conforme Castoldi (1997) e Martins (1999), por supervariáveis. Esses valores foram submetidos à análise de variância. As supervariáveis, para maior facilidade de discussão, receberam nomes semelhantes aos fatores dos quais derivam.

Com os resultados das análises de variância das supervariáveis, estimaram-se parâmetros genéticos e os ganhos preditos baseados no diferencial de seleção (Cruz, 1997) e por índices de seleção. Os índices de seleção utilizados foram: Índice de Seleção de Smith (1936) e Hazel (1943), Índice de Seleção de Pesek e Baker (1969), Índice de Seleção de Elston (1963) e

Índice de Seleção de Williams (1962). O aplicativo computacional GENES (Cruz, 1997) foi utilizado para efetuar todas as análises.

### Resultados

Pretendia-se obter um fator relacionado simultaneamente com as características produtividade e capacidade de expansão. Considerando que, quanto menor o número de fatores mais interessantes são os resultados da AF, fez-se, inicialmente, a análise de fatores com a definição antecipada de quatro fatores, o que não permitiu a extração de um fator cujo respectivo autovalor associado fosse superior à unidade e que também estivesse relacionado com as características produtividade e capacidade de expansão, nem pelo menos só com a capacidade de expansão, a característica principal. Então, seguindo o roteiro de Jonhson e Wichern (1992), fizeram-se as análises com cinco, seis e sete fatores e em todas se conseguiu um fator relacionado com a característica capacidade de expansão, mas em nenhuma delas um fator relacionado simultaneamente com as duas características acima referidas. As análises com seis e sete fatores não melhoraram a discriminação dos fatores e por esse motivo apenas se apresentam os resultados da análise com cinco fatores.

As cargas fatoriais iniciais, as comunalidades iniciais e as cargas fatoriais finais, após a rotação pelo método Varimax, são apresentadas na Tabela 1. A partir das cargas fatoriais finais se identificaram os fatores: o primeiro fator está relacionado com a produção porque é o que apresenta maiores cargas fatoriais finais para as características relacionadas com a produção, que neste trabalho são NESP, PROLI e PRODU, foi chamado Fator Produção; o segundo fator é o que está mais influenciado pela altura da planta, pois que apresenta cargas fatoriais finais mais elevadas para as características ALTP e ALTES, foi denominado Fator Altura; o terceiro fator apresenta maior carga fatorial final para o caráter QUE, pelo que se designou por Fator Quebradas. Ao quarto fator não foi possível atribuir a influência de uma ou mais características, pois as cargas fatoriais deste fator são relativamente baixas para todas as características e por isso não foi utilizado. O quinto fator é o que apresenta a maior carga fatorial final para a característica capacidade de expansão, sendo identificado por Fator Expansão. As cargas fatoriais finais foram superiores às correlações com os cinco primeiros componentes principais e confirmaram o relacionamento entre as respectivas variáveis e os fatores.

As cargas canônicas foram utilizadas como coeficientes de ponderação das variáveis padronizadas,

para a obtenção dos escores correspondentes às novas super variáveis, obtidas a partir da análise de fatores e que são identificadas com as siglas dos respectivos fatores. Segundo a teoria da análise de fatores, as variáveis que se concentram em determinado fator devem ser fracamente relacionadas com as variáveis dos outros fatores de modo que os fatores sejam não-correlacionados. Neste estudo essas correlações foram consideradas fracas.

As análises de variância das supervariáveis permitiram verificar a existência de variabilidade genética, o que permite a seleção, embora as herdabilidades não sejam elevadas. O número de 25 famílias foi definido como o número mínimo de famílias a ser selecionado, em todas as predições de ganhos.

A seleção por resposta correlacionada, baseada no diferencial de seleção, foi praticada para aumento da super variável FTCE, aumento da super variável FTPROD, diminuição da super variável FTALT e diminuição da super variável FTQUE, sendo os ganhos preditos para as características originais apresentados na Tabela 2. Verifica-se que a seleção na super variável FTPROD não permitiu a predição de ganhos nas duas características PRODU e CE, mas a seleção na super variável FTCE permitiu a predição de ganhos nessas duas características. No entanto, essas predições de ganhos foram acompanhadas de aumentos nas características que se deseja diminuir, ALTP, ALTES, PRE e QUE. A seleção para diminuição da super variável FTALT possibilitou a predição de ganhos simultâneos em PRODU e CE, mas a seleção para diminuição da super variável FTQUE não proporcionou predições de ganhos interessantes, conforme se apresenta na Tabela 2.

Os ganhos preditos por índices de seleção, sendo a seleção praticada nos escores FTPROD e FTCE, são apresentados na Tabela 3. Verifica-se que apenas o Índice de Smith (1936) e Hazel (1943), com pesos econômicos obtidos por tentativas, permitiu a predição de ganhos nas características PRODU e CE. No entanto, não houve predição de diminuição de ALTES nem de PRE e aparece predição de diminuição de EST, que são três ocorrências indesejáveis.

Os ganhos preditos por índices de seleção, sendo a seleção praticada nos escores FTALT, para a sua diminuição e FTPROD e FTCE para seu aumento, são apresentados na Tabela 4. Mais uma vez, apenas o Índice de Smith (1936) e Hazel (1943), com pesos econômicos obtidos por tentativas, permitiu a predição de ganhos com o sentido desejado, nas características ALTP, PRODU e CE mas não nos caracteres ALTES e PRE, simultaneamente.

**Tabela 1.** Cargas fatoriais iniciais, comunalidades iniciais e cargas fatoriais após rotação, estimadas nos caracteres avaliados nas famílias de meios irmãos do milho pipoca CMS 43, em Sete Lagoas, Estado de Minas Gerais, no ano agrícola de 1997-1998

Caracteres	Cargas fatoriais iniciais					Comunalidades iniciais	Cargas fatoriais após rotação				
	FTPRO	FTALT	FTQUE	Fator 4	FTCE		FTPRO	FTALT	FTQUE	Fator 4	FTCE
ALT	0,4834	0,3117	0,3243	-0,7112	-0,2240	0,991931	0,1631	0,9746	0,0181	-0,1223	-0,0142
ALTES	0,5419	0,7266	0,4000	-0,0768	0,0325	0,988700	0,1541	0,7377	-0,0444	-0,0656	-0,0656
PRE	0,2162	0,6219	0,2013	0,6558	0,2952	0,991341	0,0400	-0,5556	-0,0594	-0,0580	-0,0580
AC	-0,4420	0,4202	-0,5230	-0,1558	0,0661	0,674093	-0,2480	-0,1117	-0,7708	0,0464	0,0464
QUE	0,1292	-0,6006	0,4829	0,1625	-0,0929	0,645704	0,0282	-0,1002	0,7860	0,0004	0,0004
EST	0,2142	0,2585	-0,4950	0,1741	-0,5225	0,661024	0,2098	-0,0214	-0,3793	-0,6875	-0,6875
NESP	0,9095	-0,1647	-0,2887	0,0838	0,1056	0,955874	0,9587	0,0909	0,0909	-0,1304	-0,1304
PROLI	0,8764	-0,2482	-0,1412	0,0353	0,28785	0,933836	0,9301	0,2125	0,2125	0,0921	0,0921
PRODU	0,8225	-0,1248	-0,3101	-0,0930	-0,0024	0,796912	0,8583	0,0037	0,0037	-0,1562	-0,1562
CE	-0,2373	0,0038	-0,1548	-0,3612	0,7010	0,702218	0,0132	-0,3182	-0,3182	0,7674	0,7674

**Tabela 2.** Estimativas dos ganhos preditos por seleção direta e indireta (resposta correlacionada) nos caracteres avaliados em famílias de meios irmãos do milho pipoca CMS43, em Sete Lagoas, Estado de Minas Gerais, no ano agrícola de 1997-1998, por seleção nos escores dos fatores, fornecidos pela análise de fatores, expressos como combinação linear dos caracteres

Seleção em		Ganho em									
		ALTP	ALTES	PRE	AC	QUE	EST	NESP	PROLI	PRODU	CE
FTPROD a	GS	1,97	-0,22	-0,0004	-0,10	0,26	0,03	1,21	0,05	187,86	-0,21
FTPROD a	GS%	0,88	-0,16	-0,0600	-1,87	6,83	0,14	5,41	4,00	3,79	-0,98
FTCE a	GS	0,99	0,94	0,0001	-0,44	0,17	0,02	1,26	0,05	190,22	0,10
FTCE a	GS%	0,44	0,68	0,0100	-7,84	4,42	0,12	5,61	4,40	3,84	0,45
FTALT d	GS	-	-	-	-	-	-	-	-	102,82	0,09
FTALT d	GS%	-	-	-	-	-	-	-	-	2,07	0,42
FTQUE d	GS	-	-	-	-	-	-	-	-	135,89	-0,08
FTQUE d	GS%	-	-	-	-	-	-	-	-	2,74	-0,36

a – seleção para aumento da característica, d – seleção para decréscimo da característica, GS - ganho por seleção baseado no diferencial de seleção, GS% - ganho por seleção percentual baseado no diferencial de seleção

**Tabela 3.** Estimativas dos ganhos percentuais preditos, por seleção simultânea nos escores FTPROD e FTCE, fornecidos pela análise de fatores dos caracteres avaliados nas famílias de meios irmãos do milho pipoca CMS 43, em Sete Lagoas, Estado de Minas Gerais, no ano agrícola de 1997-1998, para formação de uma população com 25 famílias

Índices de Seleção	Seleção para aumento de FTPROD e FTCE									
	SHCV <sub>g</sub>	SH $\hat{\sigma}_g$	SHCV <sub>g</sub> /CV	SHh <sup>2</sup>	SHr <sub>g</sub>	SHt	PB $\hat{\sigma}_g$	PB10% <sub>m</sub>	LPP	BCV <sub>g</sub>
Ganho em										
ALTP	0,94	0,88	0,78	0,94	-0,20	-0,20	1,02	0,80	1,01	0,99
ALTESP	-0,28	-0,16	-0,33	-0,28	-1,23	0,29	1,00	0,98	0,80	0,12
PRE	-0,07	-0,06	-0,06	-0,07	-0,06	0,03	Indetermi.	0,01	-0,01	-0,05
AC	-1,66	-1,87	-1,54	-1,66	3,53	-5,06	-7,63	-8,00	-6,69	-3,67
QUE	6,23	6,83	6,43	6,23	3,01	-1,41	7,23	6,83	7,63	8,03
EST	0,14	0,14	0,14	0,14	0,03	-0,01	0,02	0,08	-0,01	0,07
NESP	4,68	5,41	5,07	4,68	-0,92	0,90	6,69	5,56	6,94	5,86
PROLI	3,30	4,00	3,63	3,30	-1,13	0,85	6,02	4,60	6,37	4,80
PRODU	3,98	3,79	3,79	3,98	-1,21	0,84	4,26	3,55	4,88	3,31
CE	-0,37	-0,98	-0,60	-0,37	2,75	1,22	-0,70	-0,34	-1,64	-1,25

SH – índice de seleção de Smith (1936) e Hazel (1943), SHCV<sub>g</sub> - SH usando o coeficiente de variação genético como peso econômico, SH  $\hat{\sigma}_g$  - SH usando o desvio padrão genético como peso econômico, SHCV<sub>g</sub>/CV - SH usando a razão CV<sub>g</sub>/CV como peso econômico, SHh<sup>2</sup> - SH usando a herdabilidade como peso econômico, SHr<sub>g</sub> - SH usando o coeficiente de correlação genético como peso econômico, SHt - SH usando pesos econômicos 100 e 100000 para FTPROD e FTCE, respectivamente, obtidos por tentativas, PB  $\hat{\sigma}_g$  - índice de seleção de Pesek & Baker (1969) usando como ganhos desejados o desvio padrão genético, PB10%<sub>m</sub> - índice de seleção de Pesek & Baker (1969) usando como ganhos desejados 10% da média dos escores, LPP - índice de seleção Livre de Pesos e Parâmetros (Elston, 1963) usando a média dos escores como ponto de corte para FTPROD e FTCE, BCV<sub>g</sub> - índice de seleção Base (Williams, 1962) usando o CV<sub>g</sub> como peso econômico, Indetermi - indeterminado

Na Tabela 5, apresentam-se os resultados da seleção praticada nos escores FTALT, FTQUE, para a sua diminuição e nos escores FTPROD e FTCE, para seu aumento, o que possibilitou a obtenção de ganhos preditos nas características PRODU e CE e em algumas outras, com quase todos os índices utilizados. Os índices de Pesek e Baker (1969), com peso econômico igual ao

desvio padrão genético, e o de Elston (1963) foram os que não predisseram ganhos simultâneos nas duas características principais. No entanto, o Índice de Pesek e Baker (1969), com pesos econômicos iguais a 10% da média, permitiu a melhor distribuição de ganhos.

**Tabela 4.** Estimativas dos ganhos percentuais preditos, por seleção simultânea nos escores estimados por meio da análise de fatores dos caracteres avaliados nas famílias de meios irmãos do milho pipoca CMS 43, em Sete Lagoas, Estado de Minas Gerais, no ano agrícola de 1997-1998, para formação de uma população com 25 famílias

Seleção para diminuição de FTALT e para aumento de FTPRODU e de FTCE										
Índices de Seleção	SHCV <sub>g</sub>	SH $\hat{\sigma}_g$	SHCV <sub>g</sub> /CV	SHh <sup>2</sup>	SHr <sub>g</sub>	SHt	PB $\hat{\sigma}_g$	PB10% <sub>m</sub>	LPP	BCV <sub>g</sub>
Ganho em										
ALTP	0,92	0,92	0,92	0,95	-0,30	-0,29	-1,48	-1,44	1,19	1,14
ALTESP	-0,25	-0,25	-0,25	-0,24	0,20	0,12	-1,52	-1,43	1,07	-0,13
PRE	-0,07	-0,07	-0,07	-0,07	0,03	0,02	Ind.	Ind.	Ind.	-0,08
AC	-0,23	-0,23	-0,23	-0,35	-3,42	-3,79	6,28	6,11	-5,83	-0,88
QUE	3,61	3,61	3,61	4,12	-0,70	0,90	-0,20	0,60	6,23	3,92
EST	0,04	0,04	0,04	0,09	0,05	0,06	0,12	0,16	Ind.	0,08
NESP	0,36	0,36	0,36	1,00	1,93	6,00	1,59	1,78	5,37	1,14
PROLI	-0,04	-0,04	-0,04	0,21	1,51	5,17	0,75	0,67	4,96	0,37
PRODU	1,31	1,31	1,31	1,36	1,98	4,31	-0,12	-0,26	2,65	2,07
CE	0,78	0,78	0,78	1,06	0,06	0,24	1,41	1,41	-1,80	0,42

SH – índice de seleção de Smith (1936) e Hazel (1943), SHCV<sub>g</sub> - SH usando o coeficiente de variação genético como peso econômico, SH  $\hat{\sigma}_g$  - SH usando o desvio padrão genético como peso econômico, SHCV<sub>g</sub>/CV - SH usando a razão CV<sub>g</sub>/CV como peso econômico, SHh<sup>2</sup> - SH usando a herdabilidade como peso econômico, SHr<sub>g</sub> - SH usando o coeficiente de correlação genético como peso econômico, SHt - SH usando pesos econômicos 1, -10, e 100 para FTALT, FTPRODU e FTCE, respectivamente, obtidos por tentativas, PB  $\hat{\sigma}_g$  - índice de seleção de Pesek & Baker (1969) usando como ganhos desejados o desvio padrão genético, PB10%<sub>m</sub> - índice de seleção de Pesek & Baker (1969) usando como ganhos desejados 10% da média dos escores, LPP - índice de seleção Livre de Pesos e Parâmetros (Elston, 1963) usando a média como ponto de corte, BCV<sub>g</sub> - índice de seleção Base (Williams, 1962) usando o CV<sub>g</sub> como peso econômico, Ind. - indeterminado

**Tabela 5.** Estimativas dos ganhos percentuais preditos, por seleção simultânea nos escores da análise de fatores dos caracteres avaliados nas famílias de meios irmãos do milho pipoca CMS 43, em Sete Lagoas, Estado de Minas Gerais, no ano agrícola de 1997-1998, para formação de uma população com 25 famílias

Seleção para diminuição de FTALT e FTQUE e para aumento de FTPROD e FTCE										
Índices de seleção	SHCV <sub>g</sub>	SH $\hat{\sigma}_g$	SHCV <sub>g</sub> /CV <sub>c</sub>	SHh <sup>2</sup>	SHr <sub>g</sub>	SHt	PB $\hat{\sigma}_g$	PB10% <sub>m</sub>	LPP	BCV <sub>g</sub>
Ganho em										
ALTP	0,87	0,87	0,87	0,87	-0,25	-0,25	-1,41	-0,03	1,22	1,14
ALTESP	-0,35	-0,35	-0,35	-0,35	0,30	0,31	-1,36	0,56	1,30	-0,13
PRE	-0,07	-0,07	-0,07	-0,07	0,03	0,03	Ind.	0,04	0,01	-0,08
AC	-0,88	-0,88	-0,88	-0,88	-3,91	-3,95	6,60	8,24	-5,43	-0,88
QUE	3,81	3,81	3,81	3,81	-1,01	-0,91	-0,30	-4,22	4,82	3,92
EST	0,06	0,06	0,06	0,06	0,04	0,04	0,15	0,18	0,05	0,08
NESP	0,56	0,56	0,56	0,56	2,96	3,84	1,24	2,03	6,54	1,14
PROLI	-0,05	-0,05	-0,05	-0,05	2,55	3,36	0,21	0,76	5,77	0,37
PRODU	1,60	1,60	1,60	1,60	2,65	3,36	-0,78	1,31	3,03	2,07
CE	0,78	0,78	0,78	0,78	0,32	0,13	1,24	1,43	-1,54	0,42

SH – índice de seleção de Smith (1936) e Hazel (1943), SHCV<sub>g</sub> - SH usando o coeficiente de variação genético como peso econômico, SH  $\hat{\sigma}_g$  - SH usando o desvio padrão genético como peso econômico, SHCV<sub>g</sub>/CV - SH usando a razão CV<sub>g</sub>/CV como peso econômico, SHh<sup>2</sup> - SH usando a herdabilidade como peso econômico, SHr<sub>g</sub> - SH usando o coeficiente de correlação genético como peso econômico, SHt - SH usando 10, -15, 4 e -4 como pesos econômicos obtidos por tentativas, respectivamente para FTALT, FTQUE, FTPROD e FTCE, PB  $\hat{\sigma}_g$  - índice de seleção de Pesek & Baker (1969) usando como ganhos desejados o desvio padrão genético, PB10%<sub>m</sub> - índice de seleção de Pesek & Baker (1969) usando como ganhos desejados 10% da média dos escores, LPP - índice de seleção Livre de Pesos e Parâmetros (Elston, 1963) usando a média mais  $\hat{\sigma}_g$  para FTALT e FTQUE e a média para FTPROD e FTCE como pontos de corte, BCV<sub>g</sub> - índice de seleção Base (Williams, 1962) usando o CV<sub>g</sub> como peso econômico, Ind. - indeterminado.

**Discussão**

A Análise de Fatores permite reduzir um número elevado de variáveis originais observadas a um pequeno número de variáveis abstratas, os fatores (Castoldi, 1997), sendo tanto mais útil quanto menor o número de fatores e maior o número de variáveis iniciais. Cada fator reúne variáveis originais fortemente correlacionadas entre si, mas fracamente correlacionadas com as dos outros fatores (Johnson e Wichern, 1992; Harris, 1975). O conhecimento das

relações biológicas e funcionais existentes entre as várias características mensuráveis possibilita uma escolha mais criteriosa das variáveis sobre as quais se deverá exercer a seleção, pois as variáveis reunidas em cada fator são as que melhor explicam cada um dos referidos fenômenos biológicos (Acquaah *et al.*, 1992). Neste trabalho, a AF permitiu a criação de quatro fatores que reúnem em cada um várias características sobre as quais se pretende exercer a seleção, facilitando a interpretação dos resultados, visto que, em vez de interpretar simultaneamente

dez variáveis, o melhorista poderá se concentrar em quatro.

No melhoramento de plantas, a seleção simultânea em várias características, se exercida sobre poucos fatores que representam vários caracteres originais fortemente correlacionados, pode tornar-se mais eficiente. Castoldi (1997), que a usou para a seleção de famílias de meios-irmãos de milho comum, pondera que a AF tem possibilidades de vir a ser útil na seleção simultânea de características. Abreu (1997), trabalhando com frangos, afirma que a AF tem grande potencial em estudos genéticos porque cria fatores com significado tanto biológico como econômico, o que facilita a interpretação dos resultados.

No entanto, Martins (1999), trabalhando com dados de *Eucalyptus grandis*, de seis ambientes, com cinco variáveis, não obteve a predição de ganhos com a distribuição que pretendia, ao praticar a seleção nos escores dos fatores extraídos pela análise de fatores, em nenhum dos ambientes. Castoldi (1997), nos seus quatro ensaios de milho, com nove variáveis em todos, ao aplicar a seleção nos escores dos fatores, usando a resposta correlacionada e índices de seleção, em nenhum ambiente conseguiu a combinação desejada para todos os caracteres, existindo sempre um ou mais caracteres com sinal contrário ao que se pretendia.

Neste trabalho a utilização dos escores obtidos com a AF para a predição de ganhos por seleção simultânea de características (índices de seleção) forneceu predições superiores e com melhor distribuição que a seleção por resposta correlacionada, baseada no diferencial de seleção. Assim os índices de seleção podem ser considerados superiores à resposta correlacionada, para a obtenção de ganhos simultâneos em várias características. Cruz et al. (1993) conseguiram predizer ganhos simultâneos nas características teor de óleo e rendimento de espigas, em progênies de irmãos completos de milho comum, usando índices de seleção, o que não foi possível quando usaram a seleção por resposta correlacionada, embora não tenham usado a Análise de Fatores.

A supervariável FTALT poderá ser utilizada para se obter, por resposta correlacionada baseada no diferencial de seleção, predição de ganhos por seleção em PRODU e CE. A identificação desta supervariável através da Análise de Fatores é uma das utilidades desta técnica.

Considera-se que a análise de fatores é uma técnica alternativa viável para a obtenção de escores a serem utilizados sob vários critérios de seleção.

## Referências

- ABREU, V.M.N. *Estudo da capacidade de combinação de linhagens de matrizes de frango de corte, utilizando técnicas uni e multivariadas*. 1997. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa, 1997.
- ACQUAAH, G. et al. A factor analysis of plant variables associated with architecture and seed size in dry bean. *Euphytica*, Dordrecht, v. 60, n. 3, p. 71-177, 1992.
- CASTOLDI, F.L. *Comparação de métodos multivariados aplicados na seleção em milho*. 1997. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa, 1997.
- CRUZ, C.D. *Programa genes, aplicativo computacional em genética e estatística*. Viçosa: Universidade Federal de Viçosa, 1997.
- CRUZ, C.D. et al. Comparison of gains from selection among corn progenies, based on different criteria. *Rev. Bras. Gen.*, Ribeirão Preto, v. 16, n. 1, p. 79-89, 1993.
- ELSTON, R.C. A weight-free index for the purpose of ranking or selection with respect to several traits at a time. *Biometrics*, Wahington D.C., v. 19, n. 1, p. 85-97, 1963.
- GAMA, E.E.G. Melhoramento de milhos especiais. In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 2, 1997, Lavras, MG. *Simpósio...* Lavras: UFLA, 1997. p. 251-254.
- GOULD, S.J. *The mismeasure of man*. New York: W. W Norton, 1996.
- HARRIS, R. J. *A primer of multivariate statistics*. New York: Academic Press, 1975.
- HAZEL, H.N. The genetic basis for constructing selection indexes. *Genetics*, Bethesda, v. 28, n. 6, p. 476-490, 1943.
- JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. Englewood Cliffs: Prentice Hall, 1992.
- JUVIK, J.A. et al. Kernel changes in a shrunken 2 maize population associated with selection for increased field emergence. *J. Amer. Soc. Hort. Sci.*, v. 118, n. 1, p. 135-140, 1993.
- MARQUES, M.J.B.S.G.S.M. *Número mínimo de famílias de meios irmãos de milho pipoca, critérios de seleção e predição de ganhos por seleção*. 2000. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa, 2000.
- MARTINS, I.S. *Comparação entre métodos uni e multivariados aplicados na seleção em Eucalyptus grandis*. 1999. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa, 1999.
- PACHECO, C.A.P. et al. Formação de compostos de milho pipoca. In: Embrapa/CNPMS (Ed.). *Relatório Técnico Anual*, Sete Lagoas, 1992a. p. 165-166.
- PACHECO, C.A.P. et al. Seleção entre e dentro de progênies de meios-irmãos em duas populações de milho pipoca. In: Embrapa/CNPMS (ed.) *Relatório Técnico Anual*. Sete Lagoas, MG, 1992b. p. 166-167.

PESEK, J., BAKER, R.J. Desired improvement in relation to selected indices. *Can. J. Plant Sci.*, Ottawa, v. 49, n. 6, p. 803-804, 1969.

PETERSEN, C. J. Similarities among test sites based on cultivar performance in the hard red winter-wheat region. *Crop Sci.*, Madison, v. 32, n. 4, p. 907-912, 1992.

RANALLI, P. *et al.* Response to selection for seed yield in bean (*Phaseolus vulgaris*). *Euphytica*, Dordrecht, v. 57, n. 2, p. 117-123, 1991.

SMITH, H.F. A discriminant function for plant selection. *Ann. Eug.*, v. 7, p. 240-250, 1936.

WILLIAMS, J.S. The evolution of a selection index. *Biometrics*, Washington D.C., v. 18, n. 4, p. 375-393, 1962

*Received on December 12, 2000.*

*Accepted on April 05, 2001.*