

MODELOS MISTOS NA SELEÇÃO DE FAMÍLIAS DE CANA DE AÇÚCAR APARENTADAS SOB O ENFOQUE CLÁSSICO E BAYESIANO

Mariane Alves Gomes da SILVA¹
Luiz Alexandre PETERNELLI¹
Moysés NASCIMENTO¹
Felipe Lopes da SILVA²

- **RESUMO:** A base do agronegócio de cana-de-açúcar é o desenvolvimento de novas variedades a partir do melhoramento genético. A estratégia ótima de seleção de plantas é por meio da predição de valores genotípicos usando o BLUPI (Best Linear Unbiased Predictor Individual). Recentemente um procedimento operacionalmente mais prático, denominado BLUPIS (BLUP individual simulado), foi proposto. No BLUPIS os dados são coletados ao nível de parcela. O BLUPIS permite selecionar as melhores famílias e, posteriormente, simular o número de plantas a serem selecionadas dentro dessas famílias. O objetivo deste trabalho foi comparar o BLUPIS sob o enfoque clássico (REML/BLUP) e sob o enfoque bayesiano. Os dados de produtividade foram expressos como uma estimativa da tonelada de cana por hectare (TCH). Foram avaliadas as variâncias genética e residual, a herdabilidade e os valores genotípicos das famílias, em função de seus intervalos de confiança ou de credibilidade. O procedimento bayesiano foi mais eficiente quando se consideraram, no modelo, a informação de parentesco e a distribuição *a priori* como informativa, nos demais casos os dois procedimentos apresentaram resultados equivalentes.
- **PALAVRAS-CHAVE:** Análise bayesiana; melhoramento genético; BLUPIS

1 Introdução

A composição genética de diversas espécies de plantas é o resultado de um longo processo natural de melhoramento a que elas são submetidas durante sua existência. Entretanto, algumas espécies tem seu progresso genético direcionado de forma mais rápida e intencional com a ajuda de melhoristas de plantas e com o desenvolvimento ou uso de métodos estatísticos apropriados para otimizar o processo de seleção. A cana-de-açúcar (*Saccharum spp.*) é uma dessas espécies que, devido a sua grande importância na produção de açúcar e etanol, é bastante estudada em programas de melhoramento.

¹ Universidade Federal de Viçosa - UFV, Departamento de Estatística, Centro de Ciências Exatas e Tecnológicas, CEP:36570-000, Viçosa, MG, Brasil. E-mail: mariane.estadistica@gmail.com / petermelli@ufv.br / moysesnascim@ufv.br

² Universidade Federal de Viçosa - UFV, Departamento de Fitotecnia, Centro de Ciências Agrárias, CEP:36570-000, Viçosa, MG, Brasil.

A utilização das equações de modelos mistos para análises de dados vem sendo de grande aplicabilidade na seleção de genótipos superiores (Lachos, 2005). Esses modelos compreendem tanto efeitos fixos como aleatórios em sua formulação e assumem distribuição multivariada tanto para os efeitos aleatórios como para os erros.

A estratégia ótima de seleção de genótipos superiores é obtida a partir do cálculo dos valores genotípicos preditos pelo BLUP (melhor preditor linear não viciado) individual (BLUPI) (Resende, 2007). Este método contempla simultaneamente as informações de família e de indivíduo para a seleção. No entanto, este método não tem sido utilizado em cana-de-açúcar devido às dificuldades práticas em se obter dados de plantas individuais.

Na prática, na maioria dos programas de melhoramento da cana, é aplicada a seleção individual de clones, ou seleção massal (Matsuoka et al., 2005). No entanto pode ser mostrado que a seleção massal é ineficiente (Kimbeng & Cox, 2003). Esses autores discutem que o ganho genético a partir da seleção de famílias seguida da seleção individual de clones é maior do que o ganho com a simples seleção de clones ou de famílias. Além disso, a economia de recursos seria muito maior.

Segundo Resende & Barbosa (2006) a variação genética dentro de famílias pode ser um parâmetro para descrever a superioridade de um determinado cruzamento, mas sua estimativa apresenta erros maiores do que para a variância entre as famílias e envolve custos adicionais elevados que irá resultar em geral, baixa eficiência de seleção.

Um método que consiste em determinar, de forma dinâmica, o número de indivíduos a serem selecionados em cada família, sem que seja necessária as suas avaliações individuais foi descrito por Resende & Barbosa (2006). Esse método é denominado BLUP Individual Simulado (BLUPIS).

O método BLUPIS é usado via enfoque clássico (REML/BLUP), onde a análise é realizada em duas etapas: estimação dos componentes de variância via REML (máxima verossimilhança restrita) seguida da predição dos valores genéticos via BLUP. No entanto, uma possibilidade de análise desse tipo de dados seria usando a abordagem bayesiana. Apesar de o enfoque bayesiano ser geralmente mais trabalhoso e sua obtenção pela forma analítica ser bastante difícil, o que dificulta seu uso (Gianola, 1996), esse tipo de abordagem é uma técnica elegante e flexível que permite a estimação simultânea dos componentes de variância, dos efeitos fixos (sob o ponto de vista frequentista), e dos valores genéticos de maneira precisa, podendo inserir informações prévias (à priori), mesmo para amostras de tamanho finito (Resende, 2001). Essa última consideração justificaria a tentativa de se implementar o procedimento bayesiano na análise de dados de cana-de-açúcar.

Deve-se ficar claro que na presença de prioris não informativas os resultados obtidos por meio da abordagem bayesiana são semelhantes aos obtidos pela análise usual de modelos mistos. Assim, pode-se dizer que sob tais condições a análise clássica de modelos mistos é um caso particular da abordagem bayesiana, por apresentar soluções semelhantes (Sorensen & Gianola, 2002, p.318). Porém são abordagens diferentes quanto as formas de implementação. Na abordagem bayesiana todos os efeitos do modelo, sejam fixos ou aleatórios conforme definidos sob o ponto de vista frequentista, são tratados com aleatórios. Em outras palavras, as definições de efeitos fixos e aleatórios na abordagem bayesiana não surgem naturalmente como na abordagem frequentista (Sorensen & Gianola, 2002, p.313 a 319).

O presente trabalho tem por objetivo fazer uma comparação inédita entre a obtenção dos efeitos genotípicos de famílias de cana-de-açúcar sob o enfoque clássico

(REML/BLUP) e sob o enfoque bayesiano para a aplicação do método BLUPIS em um conjunto de dados reais.

2 Metodologia

Foram avaliadas 22 famílias de irmãos germanos de cana-de-açúcar. Os cruzamentos foram realizados em 2006 na Estação de Floração e Cruzamentos, a qual se localiza na Serra do Ouro, no município de Murici – AL, situada à latitude 9°13'S, longitude 35°50' W e a 450-500 m de altitude.

As plântulas obtidas de cada família foram transplantadas em maio de 2007, no Centro de Experimentação em Cana-de-açúcar (CECA), localizado em Oratórios – MG (latitude 20°25' S, longitude 42°48' W e 494 m de altitude), pertence à Universidade Federal de Viçosa.

Cada parcela foi constituída por dois sulcos espaçados de 1,40 m, sendo cada sulco composto por dez plantas equidistantes a 0,5 m. O delineamento experimental utilizado foi o de blocos casualizados com cinco repetições, totalizando um número de 100 genótipos avaliados por família.

Os dados de cana planta foram coletados em junho de 2008. A variável estudada neste trabalho foi referente à tonelada de cana por hectare (TCH). Assim, a estimativa da produtividade de cada família foi obtida via o estimador:

$$TCH = \frac{MTP \times 10}{tp}$$

em que, MTP é a massa total da parcela em kg, e tp é o tamanho da parcela em m².

O desempenho do enfoque clássico (REML/BLUP) e do enfoque bayesiano na obtenção dos efeitos genotípicos de famílias para a aplicação do método de seleção BLUPIS foi realizado sobre a variável TCH.

As equações de modelos mistos foram utilizadas para o cálculo dos efeitos genotípicos das famílias. O modelo estatístico adotado é expresso pela equação $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}$, em que \mathbf{y} é o vetor (n×1) de observações, \mathbf{X} é a matriz (n×p) de incidência dos efeitos fixos, \mathbf{Z} é a matriz (n×q) de incidência dos efeitos aleatórios (genéticos), $\boldsymbol{\beta}$ é o vetor (p×1) de efeitos fixos a serem estimados, \mathbf{g} é o vetor (q×1) de efeitos aleatórios (efeito genotípico das famílias) a serem preditos, $\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios associados a cada observação.

Admitiu-se, neste trabalho, que os erros são não correlacionados, isto é, $\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{y}) = \text{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}) = \text{Var}(\boldsymbol{\varepsilon}) = \mathbf{R} = \mathbf{I}\sigma_{\varepsilon}^2$, em que \mathbf{I} é a matriz identidade. Para o modelo adotado foi considerado $\mathbf{g} \sim N(0, \mathbf{A}\sigma_g^2)$ e $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_{\varepsilon}^2)$.

A obtenção da estimativa de $\boldsymbol{\beta}$, demonstrada por Henderson (1959), é dada por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

E a predição de \mathbf{g} demonstrada por Henderson (1973) é dada por:

$$\hat{\mathbf{g}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

em que $\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}$ é a inversa da matriz de covariância de \mathbf{y} , e $\mathbf{G} = \mathbf{A}\sigma_g^2$, em que σ_g^2 é a variância genética aditiva. Além disso, nas expressões

acima σ_ε^2 é a variância residual e \mathbf{A} é a matriz que indica o grau de associação genética entre os indivíduos, podendo ser obtida conforme apresentado em Peternelli et al. (2009). Utilizou-se o programa Selegen-REML/BLUP (Sistema Estatístico e Seleção Genética Computadorizada via Modelos Lineares Mistos) (Resende, 2007) para obter a análise clássica. Para a obtenção da análise bayesiana foi construído um algoritmo no software livre R (R Development Core Team, 2011).

Na análise clássica, o método de estimação dos componentes de variância adotado foi o de Máxima Verossimilhança Restrita, ou Residual (REML). Este método utiliza uma função de verossimilhança que não envolve $\boldsymbol{\beta}$, ou seja, é definida somente em termos dos parâmetros de variância e covariância (Leotti, 2007).

Para a realização da análise bayesiana considerou-se, inicialmente, a distribuição condicional dos dados (\mathbf{y}) como uma normal multivariada:

$$\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \sigma_\varepsilon^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g}, \mathbf{I}\sigma_\varepsilon^2)$$

em que \mathbf{I} é a matriz de identidade.

Logo, a função de verossimilhança é dada por:

$$L(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \sigma_\varepsilon^2) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})\right\}.$$

O segundo passo foi definir a distribuição *a priori* dos parâmetros do modelo.

Para o efeito fixo, foi definida uma *a priori* do tipo constante, já que a noção do conhecimento *a priori* é vago para $\boldsymbol{\beta}$.

A distribuição *a priori* dos componentes de variância σ_g^2 (variância genética aditiva) e σ_ε^2 (variância residual) foi especificada como uma distribuição qui-quadrado escalonada invertida, da forma:

$$p(\sigma_i^2 | v_i, S_i^2) \propto (\sigma_i^2)^{-(\frac{v_i}{2}+1)} \exp\left[\frac{-v_i S_i^2}{2\sigma_i^2}\right], \text{ com } i = \varepsilon, \mathbf{g}.$$

Quando foi assumida uma distribuição *a priori* informativa substituiu-se em v os graus de liberdade da distribuição qui-quadrado e em S_i^2 , o valor inicial da variância. Tal estratégia é adotada visando fornecer mais informação ao modelo, de forma aumentar a acurácia das estimativas. O ganho por meio do uso de distribuições *prioris* informativas pode ser visto nos trabalhos de Morais et al. (2010) e Nascimento et al. (2012).

Para assumir uma distribuição *a priori* não-informativa, a qui-quadrado foi reduzida em uma uniforme imprópria, substituindo-se $v = -2$ e $S_i^2 = 0$.

Após definidas as distribuições acima, considerando independência entre as distribuições *a prioris*, pôde-se escrever a distribuição conjunta *a posteriori* dos parâmetros do modelo, $p(\boldsymbol{\beta}, \mathbf{g}, \sigma_g^2, \sigma_\varepsilon^2 | \mathbf{y})$.

$$p(\boldsymbol{\beta}, \mathbf{g}, \sigma_g^2, \sigma_\varepsilon^2 | \mathbf{y}) \propto p(\boldsymbol{\beta}, \mathbf{g}, \sigma_g^2, \sigma_\varepsilon^2) p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{g}, \sigma_g^2, \sigma_\varepsilon^2) \\ p(\boldsymbol{\beta}, \mathbf{g}, \sigma_g^2, \sigma_\varepsilon^2 | \mathbf{y}) \propto \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\left(\frac{n+v_\varepsilon}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{g}) + v_\varepsilon S_\varepsilon^2]\right\} \\ \left(\frac{1}{\sigma_g^2}\right)^{\left(\frac{q+v_g}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma_g^2}[\mathbf{g}'\mathbf{A}^{-1}\mathbf{g} + v_g S_g^2]\right\}.$$

Dentro do contexto bayesiano, qualquer inferência em relação às quantidades desconhecidas do modelo (incluindo parâmetros e variáveis não observáveis) é obtida a partir da distribuição conjunta *a posteriori* (Resende, et al, 2001; Sorensen e Gianola, 2002; Gamerman e Lopes, 2006).

Logo, a partir da distribuição conjunta obtiveram-se as distribuições *a posteriori* condicionais completas para cada parâmetro:

$$\beta_i | \beta_{-i}, \mathbf{g}, \sigma_g^2, \sigma_\varepsilon^2, \mathbf{y} \sim N(\hat{\beta}_i, (\mathbf{X}'_i \mathbf{X}_i)^{-1} \sigma_\varepsilon^2);$$

$$g_i | \beta, \mathbf{g}_{-i}, \sigma_g^2, \sigma_\varepsilon^2, \mathbf{y} \sim N(\hat{g}_i, (\mathbf{Z}'_i \mathbf{Z}_i + \mathbf{A}_{i,i}^{-1} \mathbf{g})^{-1} \sigma_\varepsilon^2);$$

$$\sigma_g^2 | \beta, \mathbf{g}, \sigma_\varepsilon^2, \mathbf{y} \sim \chi^{-2}(q + v_g), \text{ em que o parâmetro de escala é } \mathbf{g}' \mathbf{A}^{-1} \mathbf{g} + v_g S_g^2;$$

$$\sigma_\varepsilon^2 | \beta, \mathbf{g}, \sigma_g^2, \mathbf{y} \sim \chi^{-2}(n + v_\varepsilon), \text{ em que parâmetro de escala é } (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{g})'(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{g});$$

Por se tratar de distribuições tais que já estão implementadas em softwares, as aproximações para as distribuições marginais foram feitas através da implementação do Amostrador de Gibbs.

Além disso, a herdabilidade, $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\varepsilon^2}{r}}$, foi obtida em cada iteração.

Para a determinação do número n_k de indivíduos a serem selecionados em cada família k foi utilizado a expressão: $n_k = \left(\frac{\hat{g}_k}{\hat{g}_j}\right) n_j$, em que: \hat{g}_j refere-se ao efeito genotípico da melhor família; \hat{g}_k refere-se ao efeito genotípico da k -ésima família; e, n_j equivale ao número de indivíduos selecionados na melhor família (Resende & Barbosa, 2006). Estes autores também comentam que este método elimina automaticamente as famílias com efeito genotípico negativo, ou seja, aquelas abaixo da média geral do experimento.

Os intervalos referentes às predições dos componentes da variância, herdabilidade e valores genotípicos das famílias pela análise bayesiana foram calculados através do intervalo de credibilidade central, que calcula os valores relativos aos quantis de 2,5 % e 97,5% da distribuição *a posteriori* $\left(\frac{\alpha}{2} \text{ e } 1 - \frac{\alpha}{2}\right)$, para os limites inferior e superior, respectivamente (Holsinger, 2005).

Para a análise bayesiana foi desenvolvido uma função no software R (R Development Core Team, 2011), e também usados os seguintes pacotes: “geoR” (Ribeiro Junior & Diggle, 2001) e “boa” (Smith, 2007). As análises dos dados foram baseadas nas comparações entre os resultados obtidos pelos dois métodos. Foram verificados a variância genética, a variância residual, a herdabilidade e os valores genotípicos das famílias junto a seus intervalos de confiança ou de credibilidade, dependendo do enfoque clássico ou bayesiano, respectivamente. Também foram conferidas a importância da ausência ou presença de informação de parentesco (IP) e priori informativa (PI) no modelo, e se as seleções das famílias por ambos os métodos foram coincidentes.

Na análise bayesiana, para cada análise do modelo adotado, foram consideradas 70000 iterações no algoritmo Gibbs *sampler*, com um período de aquecimento (“burn-in”) mínimo de 9 iterações. Para obter-se uma amostra não correlacionada, considerou-se um

espaçamento entre os pontos amostrados de no mínimo três iterações (“*thinning*”). As amostras representam amostras das distribuições marginais *a posteriori* de cada parâmetro, sob as quais foram realizadas a inferência de cada parâmetro nas diferentes análises. A convergência das cadeias foi analisada no pacote “boa” (SMITH, 2007).

3 Resultados e discussão

Em relação à convergência, para todas as cadeias simuladas, o fator de dependência de Raftery & Lewis (1992) forneceu valores menores que cinco, e o p-valor do critério de Geweke (1992) foi sempre maior que o nível de significância pré-fixado (5%), ou seja, ambos os critérios indicaram a convergência das cadeias geradas pelo Gibbs sampler. Todas as análises foram realizadas num Microcomputador Portátil Dell Inspiron 1525, com processador Intel Core 2 duo (2 ghz, 3 gb Ram) e SO de 32 bits. A análise clássica foi executada em cerca de 30 segundos, enquanto que cada análise bayesiana necessitava, em média, de 5 vezes esse tempo.

Os resultados das estimativas e predições de parâmetros obtidos por meio do modelo misto com enfoque clássico foram semelhantes aos obtidos a partir do enfoque bayesiano sem informação de parentesco e com a distribuição *a priori* não-informativa (Tabela 1). Esse resultado já era esperado, visto que, nessa situação a maior contribuição nas estimativas é proveniente da função de verossimilhança, ou seja, do conjunto de dados em análise. Além disso, esse resultado é corroborado pelos trabalhos de Nogueira et al. (2003) e Nascimento et al. (2011).

Os resultados das estimativas de parâmetros de um modelo misto com enfoque Clássico e as predições de parâmetros dos modelos com enfoques bayesianos sem informação de parentesco e com *priori* informativa, foram considerados similares.

Observou-se também que a inclusão da distribuição *a priori* informativa (coluna 3 em comparação a 4, e a coluna 5 em comparação a 6) faz com que a maioria dos valores dos parâmetros diminuam e a informação de parentesco ajuda a aumentar valores dos efeitos genotípicos e a diminuir as magnitudes das variâncias e da herdabilidade (quando se compara a coluna 3 com a 5 e a coluna 4 com a 6). Verifica-se também que a informação de parentesco acarretou em aumento na maioria dos parâmetros (comparação da coluna 4 com a 5), o que torna evidente que a informação de parentesco tem um peso de informação maior em relação *a priori* informativa.

No entanto, não se pode desconsiderar que a distribuição *a priori* informativa ajuda a aumentar os efeitos genotípicos e a diminuir as variâncias e a herdabilidade (coluna 3 em comparação a 6), pois conforme Falconer & Mackay (1996), a eficiência da seleção de famílias é preferida quando o caráter selecionado apresenta baixa herdabilidade. As quatro diferentes maneiras de implementação do modelo misto adotadas (coluna 3, 4, 5 e 6) para a variável TCH pelo método bayesiano resultaram nas mesmas três melhores famílias que no método clássico, que foram a 61, 65 e 66.

Os resultados a respeito dos efeitos genéticos da família, tanto pelo método bayesiano quanto pelo método clássico (REML/BLUP) foram consideravelmente parecidos analisando os resultados encontrados neste estudo.

Tabela 1 - Estimativas e predições (média *a posteriori*) dos parâmetros de um modelo misto com enfoque clássico (REML/BLUP) e bayesiano para tonelada de colmos por hectare de cana-de-açúcar

Parâmetros	Clássico	Bayesiano			
		Sem IP Sem PI	Sem IP Com PI	Com IP Sem PI	Com IP Com PI
σ_g^2	229,0589	301,5960	241,2471	295,4213	240,1148
σ_e^2	538,5974	566,2199	544,4732	567,2167	544,3562
h^2	0,6801	0,6958	0,6799	0,6923	0,6793
\hat{g}_{61}	26,3301	26,9133	26,3281	27,9316	27,4169
\hat{g}_{65}	17,3817	17,7771	17,4778	18,2076	17,7736
\hat{g}_{66}	16,2544	16,6632	16,3395	17,0779	16,5408
\hat{g}_{67}	9,0838	9,2933	9,1121	9,6387	9,4663
\hat{g}_{50}	8,4911	8,8697	8,6068	10,3587	10,2149
\hat{g}_{56}	7,6944	8,0567	7,6866	8,1471	7,9132
\hat{g}_{53}	7,5002	7,6020	7,5437	6,7703	6,2672
\hat{g}_{54}	6,7229	6,9297	6,6939	9,7827	9,8734
\hat{g}_{45}	3,3416	3,3559	3,3831	3,1184	3,0269
\hat{g}_{49}	2,1075	2,0753	2,1449	0,9079	0,8619
\hat{g}_{55}	1,3205	1,4786	1,3962	1,7801	1,6571
\hat{g}_{68}	-0,6131	-0,6282	-0,5798	-0,1576	-0,2792
\hat{g}_{60}	-0,817	-0,6957	-0,8802	-0,4726	-0,6127
\hat{g}_{64}	-2,9644	-2,9524	-2,9790	-1,2729	-1,3511
\hat{g}_{48}	-5,3056	-5,2977	-5,3655	-4,9298	-4,9838
\hat{g}_{63}	-5,4129	-5,5547	-5,2997	-4,8907	-4,9585
\hat{g}_{52}	-8,182	-8,4902	-8,0695	-5,2224	-4,8791
\hat{g}_{51}	-8,7649	-8,9974	-8,8056	-8,5574	-8,4239
\hat{g}_{46}	-10,3391	-10,5774	-10,2793	-10,7033	-10,5772
\hat{g}_{57}	-19,7541	-20,2289	-19,7063	-19,6410	-19,3588
\hat{g}_{59}	-22,0277	-22,6229	-21,9477	-21,9287	-21,6986
\hat{g}_{62}	-22,0471	-22,4056	-22,0414	-21,9449	-21,5557

IP = informação de parentesco; PI = distribuição *a priori* informativa; σ_g^2 : variância genotípica; σ_e^2 : variância residual; h^2 : herdabilidade da média da família; \hat{g}_i : efeito genotípico da *i*-ésima família.

Na Tabela 2 podemos ver que, no enfoque bayesiano, cada família tem seu próprio Intervalo de Credibilidade e a maior influência na obtenção de menores intervalos de credibilidade foi a utilização de uma distribuição *a priori* informativa, pois mesmo sem ou com a informação de parentesco os intervalos de confiança e credibilidade foram bem semelhantes. Entretanto, a informação de parentesco fez com que ocorresse uma mudança na ordem de classificação das melhores famílias, mostrando que sua informação, por menor que seja, mudou de certo modo os resultados dos valores genotípicos. Esse fato também foi observado por Wright et al. (2000) e Nogueira et al. (2003), quando compararam-se a abordagem clássica e bayesiana aplicados em dados de melhoramento animal.

Tabela 2 - Amplitude dos intervalos de confiança” no método Clássico (para BLUP via REML) e intervalos de credibilidade para predições bayesiana de parâmetros de um modelo misto para tonelada de colmos por hectare de cana-de-açúcar

Parâmetros	Clássico	Bayesiana			
		Sem IP Sem PI	Sem IP Com PI	Com IP Sem PI	Com IP Com PI
σ_g^2	630,4	573,6	247,6	571,6	248,8
σ_e^2	328,6	353,4	219,2	358,5	220,3
h^2	0,4	0,40	0,20	0,40	0,20
\hat{g}_{61}	33,6	40,0	36,4	40,2	37,1
\hat{g}_{65}	33,6	39,1	36,1	38,4	35,6
\hat{g}_{66}	33,6	39,0	35,7	38,4	35,7
\hat{g}_{67}	33,6	37,6	35,8	37,9	35,4
\hat{g}_{50}	33,6	37,1	35,3	38,0	35,4
\hat{g}_{56}	33,6	37,2	35,5	37,5	35,6
\hat{g}_{53}	33,6	37,6	35,6	38,3	36,6
\hat{g}_{54}	33,6	36,4	35,9	37,5	35,5
\hat{g}_{45}	33,6	36,6	35,4	37,4	36,3
\hat{g}_{49}	33,6	37,1	35,2	37,1	35,5
\hat{g}_{55}	33,6	36,9	36,0	37,2	35,4
\hat{g}_{68}	33,6	37,2	35,4	37,4	35,4
\hat{g}_{60}	33,6	37,2	35,5	36,8	35,5
\hat{g}_{64}	33,6	37,1	35,4	37,2	35,6
\hat{g}_{48}	33,6	36,7	35,8	37,1	36,4
\hat{g}_{63}	33,6	37,0	35,4	37,0	35,8
\hat{g}_{52}	33,6	36,5	35,4	37,6	37,0
\hat{g}_{51}	33,6	37,3	35,2	37,5	35,6
\hat{g}_{46}	33,6	37,7	35,5	37,6	36,5
\hat{g}_{57}	33,6	39,3	35,9	39,3	36,3
\hat{g}_{59}	33,6	39,2	35,7	39,5	36,9
\hat{g}_{62}	33,6	38,7	35,9	39,0	36,0

IP = informação de parentesco; PI = distribuição *a priori* informativa; σ_g^2 : variância genotípica; σ_e^2 : variância residual; h^2 : herdabilidade da média da família; \hat{g}_i : efeito genotípico da *i*-ésima família.

Segundo Resende & Barbosa (2006) o método BLUPIS nos fornecem quantos indivíduos são selecionados em cada família utilizando informações apenas da família. Isto pode ser observado por meio da Tabela 3 que teve como resultado o mesmo número de indivíduos a serem selecionados por família quando comparado o REML/BLUP com o bayesiano (Sem IP, Com PI).

Na Tabela 3 se verifica que quanto maior o efeito, maior o número de indivíduos a serem selecionado, como também, a família com maior efeito genotípico é a 61 e com menor efeito é a 62. Já para as famílias 68, 60, 64, 48, 63, 52, 51, 46, 57, 59 e 62 o método BLUPIS indicou como zero a seleção de indivíduos dentro das mesmas devido aos seus valores de efeitos genotípicos terem sido negativos.

Tabela 3 - Valores dos efeitos genotípicos (\hat{g}_k) e do número de indivíduos a serem selecionados (n_k) em cada família utilizando a metodologia BLUP sob enfoque clássico e bayesiano para a característica tonelada de colmos por

hectare considerando o número de indivíduos a serem selecionados na melhor família (n_k) igual a 50

Famílias	$n_j = 50$					
	Clássico		Bayesiano			
	\hat{g}_k	n_k	\hat{g}_k (Sem IP Com PI)	n_k	\hat{g}_k (Com IP Com PI)	n_k
\hat{g}_{61}	26,3301	50	26,9133	50	27,4170	50
\hat{g}_{65}	17,3817	33	17,7771	33	17,7736	32
\hat{g}_{66}	16,2544	31	16,6632	31	16,5408	30
\hat{g}_{67}	9,0838	17	9,2933	17	9,4663	17
\hat{g}_{50}	8,4911	16	8,8697	16	10,2149	19
\hat{g}_{56}	7,6944	15	8,0567	15	7,9132	14
\hat{g}_{53}	7,5002	14	7,6020	14	6,2672	11
\hat{g}_{54}	6,7229	13	6,9297	13	9,8734	18
\hat{g}_{45}	3,3416	6	3,3559	6	3,0269	6
\hat{g}_{49}	2,1075	4	2,0753	4	0,8619	2
\hat{g}_{55}	1,3205	3	1,4786	3	1,6571	3
\hat{g}_{68}	-0,6131	0	-0,6282	0	-0,2792	0
\hat{g}_{60}	-0,817	0	-0,6957	0	-0,6127	0
\hat{g}_{64}	-2,9644	0	-2,9524	0	-1,3511	0
\hat{g}_{48}	-5,3056	0	-5,2977	0	-4,9838	0
\hat{g}_{63}	-5,4129	0	-5,5547	0	-4,9585	0
\hat{g}_{52}	-8,182	0	-8,4902	0	-4,8791	0
\hat{g}_{51}	-8,7649	0	-8,9974	0	-8,4239	0
\hat{g}_{46}	-10,3391	0	-10,5774	0	-10,5772	0
\hat{g}_{57}	-19,7541	0	-20,2289	0	-19,3588	0
\hat{g}_{59}	-22,0277	0	-22,6229	0	-21,6986	0
\hat{g}_{62}	-22,0471	0	-22,4056	0	-21,5557	0

IP = informação de parentesco; PI = distribuição *a priori* informativa, σ_g^2 : variância genotípica, σ_e^2 : variância residual

Conclusões

Na análise bayesiana, a inclusão da distribuição *a priori* informativa leva a intervalos de credibilidade mais estreitos.

De acordo com os intervalos de credibilidade, na maioria dos casos (80%), a informação de parentesco leva ao aumento dos valores genotípicos preditos e à redução nas estimativas dos componentes de variâncias e da herdabilidade.

Os intervalos de credibilidade obtidos por meio da análise bayesiana e os intervalos de confiança obtidos pelo procedimento REML/BLUP clássico foram bastante parecidos, com exceção apenas no caso das variâncias genotípica e residual.

Agradecimentos

À CAPES, pela bolsa de estudo do primeiro autor. À FAPEMIG e ao CNPq, pelo apoio financeiro. Aos revisores anônimos pelas sugestões e correções apontadas.

SILVA, M. A. G.; PETERNELLI, L. A.; NASCIMENTO, M.; SILVA, F. L. Mixed models in the selection of related sugarcane families under classical and bayesian approaches. *Rev. Bras. Biom.*, São Paulo, v.31, n.1, p. 1-12, 2013.

- **ABSTRACT:** *Sugarcane agribusiness is founded upon the development of new varieties from breeding programs. The optimal strategy for selection of plants is through prediction of their genotypic values using BLUPI (Best Linear Unbiased Predictor individual). Recently a procedure operationally more practical, called BLUPIS (simulated individual BLUP), has been proposed. In BLUPIS, data is collected at plot level. The BLUPIS allows one to select the best families and then simulate the number of plants to be selected within these families. The objective of this study was to compare BLUPIS under the classical approach (REML / BLUP) and under the Bayesian approach. Yield data was expressed in tons of stalks per hectare. The Bayesian procedure was more efficient when considered, in the model, a kinship information and also an informative prior distribution. For the other cases, the two procedures showed similar results.*
- **KEYWORDS:** *Bayesian analysis, plant breeding, BLUPIS*

Referências

FALCONER, D.S.; MACKAY, T.F.C. *Introduction to quantitative genetics*. 4. ed. London: Longman, 1996. 464p.

GAMERMAN, D. ; LOPES, H.F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. 2nd. edition. Londres: Chapman & Hall/CRC, 2006. v.1.

GEWEKE, J. chapter. *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. In: BERNARDO J. M. et al. (Ed.). *Bayesian Statistics*. Oxford University Press, New York. v.4, 1992.

GIANOLA, D. *Bayesian analysis with applications to genetics e biology*. Ribeirão Preto: USP, 1996. Class notes of a course taught at the departament of genetics, Medical School, University of São Paulo at Ribeirão Preto, 1996.

HENDERSON, C.R., KEMPTHORNE, O., SEARLE, S.R. The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, Chichester, v.15, p.192, 1959.

HENDERSON, C. R. *Sire evolution and genetic trends*. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM, 10, 1973, Champaign. Proceedings... Champaign: American Society of Animal Science, 1973, p.10-41.

HOLSINGER, K. E., *Bayesian population genetic data analysis*. Boston: department of Ecology & Evolutionary Biology, University of Connecticut, 2005. 124p.

KIMBENG, C. A.; COX, M. C. Early generation selection of sugarcane families and clones in Australia: a review. *J. Am. Soc. Sugarcane Technol.*, Baton Rouge, v.23, p.20-39, 2003.

LACHOS, V. H. *Modelos lineares mistos assimétricos*. Tese (Doutorado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2005.

LEOTTI, V.B. *Comparação via simulação dos estimadores clássicos e bayesianos no modelo de coeficientes aleatórios para dados longitudinais*. 2007. 99 f. Dissertação (Mestrado em Epidemiologia) - Universidade Federal do Rio Grande do Sul, RS, Porto Alegre, 2007.

MATSUOKA, S.; GARCIA, A. A. F.; ARIZONO, H. Melhoria da cana-de-açúcar. In: BORÉM, A. *Melhoria de espécies cultivadas*. Viçosa: UFV, 2005. p.225-274

MORAIS, T. S. S.; SILVA, F. F.; SILVA, C. H. O.; MARTINS FILHO, S.; NASCIMENTO, M.; SÁFADI, T. Análise bayesiana de sensibilidade do modelo AR(1) para dados em painel: uma aplicação em dados temporais de microarrays. *Rev. Bras. Biom.*, v.28, p.171-192, 2010.

NASCIMENTO, M.; SILVA, F. F.; SÁFADI, T.; NASCIMENTO, CAMPANA, A. C.; FERREIRA, R. P.; CRUZ, C. D. Abordagem bayesiana para avaliação da adaptabilidade e estabilidade de genótipos de alfafa. *Pesquis. Agropec. Bras.*, Brasília, v46, p.26-32, 2011, 1977.

NOGUEIRA, D. A.; SÁFADI, T.; BEARZOTI E; FILHO, J. S. D. S. B Análise clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. *Cienc. Agrotec.*, Lavras, n. esp., p. 1614-1624, 2003.

PETERNELLI, L. A.; FERREIRA, F. M.; ROCHA, R. B.; BARROS, W. S.; BARBOSA, M. H. P. Análise dos coeficientes de endogamia e de parentesco para qualquer nível de ploidia usando o pacote estatístico R. *Bragantia*, Campinas, v.68, n.4, p.849-855, 2009.

R DEVELOPMENT CORE TEAM, R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2012, Disponível em: <<http://www.R-project.org>>.

RAFTERY, A. L.; LEWIS, S. How many iterations in the gibbs sampler. In: BERNARDO, J. M.. *Bayesian statistics*, New York: Oxford University Press, 1992. v.4.

RESENDE, M. D. V., DUDA, L. L.; GUIMARÃES, P. R. Q.; FERNANDES, J. S. C. *Mixed linear models analysis using bayesian inference*. *Rev. Mat. Estat.*, São Paulo, v.19, p.41-70, 2001.

RESENDE, M. D. V. *Seleção-Reml/Blup*. Colombo: Editora Embrapa, 2007, 360p.

RESENDE, M. D. V. de; BARBOSA M. E. P. Selection via simulated individual BLUP based on family genotypic effects in sugarcane. *Pesquis. Agropec. Bras.*, Brasília, v.41, n.3, p.421-429, 2006.

RIBEIRO Jr, P. J.; DIGGLE, P. J. GeoR: a package for geostatistical analysis. *R-NEWS*, Viena, v.1, n.2, p.15-18, 2001

SILVA, M. A. G. *Modelos mistos na seleção entre e dentro de famílias de cana de açúcar sob o enfoque bayesiano*. 2012. 59f. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, Viçosa, 2012.

SMITH, B. J. Boa: an R package for MCMC output convergence Assessment and posterior inference. *J. Stat. Softw.*, Los Angeles, v.21, n.11, p.1-37, 2007.

SORENSEN, D; GIANOLA, D. *Likelihood, bayesian and MCMC methods in quantitative genetics*. Amsterdam: Springer, 2002. 740p.

WRIGHT, D. R.; STERN, H. S.; BERGER, J. Comparing traditional and Bayesian analyses of selection experiments in animal breeding. *J. Agric. Biol. Environ. Stat.*, Alexandria, v.5, n.2, p.240-256, 2000.

Recebido em 25.03.2013

Aprovado após revisão em 18.06.2013