

DELINEAMENTO E ANÁLISE ESTATÍSTICA DE EXPERIMENTOS EM SILVICULTURA: EFEITOS GENOTÍPICOS, ERROS ALEATÓRIOS E ESTIMADORES MELHORADOS PARA A MÉDIA DE TRATAMENTOS

Marcos Deon Vilela de Resende*

José Alfredo Sturion**

Os experimentos envolvendo tratamentos silviculturais são geralmente delineados com os objetivos de comparação destes tratamentos e de inferência sobre as médias dos mesmos. Em tais experimentos, os tratamentos são realizados sobre materiais genéticos que podem ser mais ou menos homogêneos dependendo das condições e objetivos da pesquisa. Os efeitos dos tratamentos podem ser influenciados pelos erros aleatórios a que está sujeita a pesquisa e pelos efeitos genotípicos particulares a cada indivíduo que recebe o tratamento. Assim, este trabalho tem por objetivo apresentar estimadores para médias de tratamentos, que forneçam informações livres das influências genotípicas e dos erros aleatórios.

Considerando o delineamento experimental de blocos ao acaso, com t tratamentos, b blocos e n plantas por parcela (plantas obtidas aleatoriamente de uma população), tem-se o modelo matemático:

$Y_{ijk} = \mu + t_i + b_j + e_{ij} + d_{ijk}$, onde μ , t_i , b_j , e_{ij} e d_{ijk} referem-se aos efeitos da média geral, de tratamentos, de blocos, de parcelas e de indivíduos dentro de parcela, respectivamente.

Assumindo t_i , e_{ij} e d_{ijk} como efeitos aleatórios e b_j como efeito fixo (estas suposições não alteram os resultados da análise de variância e das comparações das médias pelo teste F, conforme Steel & Torrie, 1980), para efeito desta apresentação, tem-se a seguinte distribuição dos efeitos genotípicos (G) através dos efeitos do modelo estatístico:

* Eng.-Agrônomo, Doutor, CREA no 50.602/D, Pesquisador da *Embrapa Florestas*.

** Eng. Florestal, Doutor, CREA no 47.263/D, Pesquisador da *Embrapa Florestas*.

$G = b_1(Y_{ijk} - \bar{Y}_{ij.}) + b_2(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) + b_3(\bar{Y}_{i..} - \bar{Y}_{...})$, em que:

$$b_1 = \frac{\sigma_G^2}{\sigma_d^2} = \frac{h_G^2}{1-t_1} = \frac{h_G^2}{1-c^2};$$

$$b_2 = \frac{\sigma_G^2/n}{\frac{\sigma_d^2}{n} + \sigma_e^2} = \frac{h_G^2}{1+(n-1)t_1} = \frac{h_G^2}{1+(n-1)c^2};$$

$$b_3 = \frac{\sigma_G^2/nb}{\frac{\sigma_d^2}{nb} + \frac{\sigma_e^2}{b} + \sigma_t^2} = \frac{h_G^2}{1+(n-1)t_1 + (nb-1)t_2} = \frac{h_G^2}{1+(n-1)c^2 + (nb-1)t_t}, \text{ em que:}$$

$t_1 = c^2 = \frac{\sigma_e^2}{\sigma_F^2}$ = correlação intraclasse entre indivíduos na parcela ou correlação devida ao ambiente comum da parcela.

$t_2 = t_t = \frac{\sigma_t^2}{\sigma_F^2}$ = correlação intraclasse entre os indivíduos que recebem um mesmo tratamento.

h_G^2 = herdabilidade no sentido amplo ao nível de indivíduo.

$\sigma_F^2 = \sigma_t^2 + \sigma_e^2 + \sigma_d^2$ = variância fenotípica.

Verifica-se pelas expressões de b_1 , b_2 e b_3 que a maioria da variação genotípica presente no material experimental fica retida nos efeitos de indivíduo dentro de parcela ($Y_{ijk} - \bar{Y}_{ij.}$) e de parcela ($\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$). Apenas a fração σ_G^2/nb permanece confundida nos efeitos de tratamentos.

Com base neste resultado, pode-se dizer que, com nb muito baixo, os valores genotípicos das plantas que recebem os tratamentos podem mascarar os efeitos de tratamento, exceto quando o material experimental for homogêneo ($\sigma_G^2 = h_G^2 = 0$) ou a herdabilidade no sentido amplo for muito baixa.

Por outro lado, o uso de material experimental homogêneo só é desejável quando o objetivo for a realização de recomendações específicas para o material em experimentação. Quando o objetivo for a extrapolação dos resultados para outros materiais genéticos, é recomendável a utilização de material genético heterogêneo e neste caso, com um número razoável de plantas para cada tratamento.

O efeito de tratamento ($\bar{Y}_{i..} - \bar{Y}_{...}$) pode ser assim decomposto:

$$\begin{aligned} (\bar{Y}_{i..} - \bar{Y}_{...}) &= \left(\frac{nb t_t}{1 + (n-1) c^2 + (nb-1) t_t} + \frac{h_G^2}{1 + (n-1) c^2 + (nb-1) t_t} + \right. \\ &+ \left. \frac{nc^2}{1 + (n-1) c^2 + (nb-1) t_t} + \frac{(1 - h_G^2 - t_t - c^2)}{1 + (n-1) c^2 + (nb-1) t_t} \right) (\bar{Y}_{i..} - \bar{Y}_{...}) \\ &= \left(\frac{nb t_t + h_G^2 + nc^2 + (1 - h_G^2 - t_t - c^2)}{1 + (n-1) c^2 + (nb-1) t_t} \right) (\bar{Y}_{i..} - \bar{Y}_{...}) \end{aligned}$$

Assim, o efeito de tratamento, na realidade, é devido a quatro fatores:

- (i) Efeito do tratamento propriamente dito, que é função de $nb t_t$;
- (ii) Efeito genotípico, que é função de h_G^2 ;
- (iii) Efeito ambiental entre parcelas, que é função de nc^2 ;
- (iv) Efeito ambiental dentro de parcelas, que é função de $(1 - h_G^2 - t_t - c^2)$.

Para a detecção de diferenças significativas entre tratamentos, é desejável que o efeito descrito em (i) seja muito superior aos demais e que explique a quase totalidade de $(\bar{Y}_{i..} - \bar{Y}_{...})$. Uma ação que pode permitir atingir tal objetivo é aumentar o número de repetições, pois neste caso, aumenta-se $nb t_t$ sem aumentar nc^2 e reduzindo também (σ_G^2 / nb) .

Com base no exposto, tem-se que a fração de $(\bar{Y}_{i..} - \bar{Y}_{...})$ que realmente determina a média do tratamento é:

$$R^2 = \frac{nb t_t}{1 + (n-1) c^2 + (nb-1) t_t} = \frac{\sigma_t^2}{\frac{\sigma_d^2}{nb} + \frac{\sigma_e^2}{b} + \sigma_t^2} = 1 - 1/F, \text{ onde } F \text{ é o}$$

valor do teste F de Snedecor da análise de variância.

Assim, a determinação (R^2) associada ao efeito e tratamento pode ser generalizada para efeitos fixos ou aleatórios de tratamentos, pois depende apenas de F que é idêntico para ambos os modelos. A média (valor esperado em uma nova repetição do ensaio ou em uma aplicação do tratamento ao nível operacional) verdadeira (M) do tratamento é então, melhor predita fazendo-se

$$M = (1 - 1/F)(\bar{Y}_{i..} - \bar{Y}_{...}) + \bar{Y}_{...} = (1 - 1/F) \bar{Y}_{i..} + (1/F) \bar{Y}_{...} \text{ (Resende, 1999).}$$

Em outras palavras, a melhor estimativa da média é dada pela fração $(1-1/F)$ da média amostral do tratamento mais a fração $(1/F)$ da média geral. Esta expressão para a média, além de ser mais coerente com a prática, considera adequadamente a precisão experimental na estimativa da média de tratamento, pois quanto mais preciso ($> F$) o experimento, menor é a correção $(-1/F)$ aplicada sobre as médias amostrais de tratamento. Neste caso, níveis de significância diferentes (por exemplo 5% e 1%) conduzirão a médias ajustadas diferentes, embora na metodologia convencional conduzissem a inferências idênticas sobre as médias.

Considere o exemplo apresentado por Steel & Torrie (1980), referente à avaliação do teor de óleo de sementes de linho, cujas plantas foram inoculadas em seis diferentes estádios de crescimento. Estes seis tratamentos foram avaliados no delineamento de blocos casualizados com quatro repetições, produzindo os seguintes resultados:

Fontes de Variação	GL	SQ	QM	E (QM)	F
Blocos	3	3,14	1,05	-	
Tratamentos	5	31,65	6,33	$\sigma_d^2 / n + \sigma_e^2 + b\sigma_t^2$	4,83**
Resíduo	15	19,72	1,31	$\sigma_d^2 / n + \sigma_e^2$	

** significativo para $p < 0,01$ de probabilidade de erro Tipo I.

Tratamentos	Médias
1	5,1
2	4,3
3	4,0
4	6,7
5	6,0
6	7,0
Média Geral	5,517

Supondo diferentes valores de F, ter-se-ia as seguintes inferências sobre

Tratamentos	M original	M com $\rightarrow\infty$ (P valor $\rightarrow 0$)	M com F = 4,56 (P valor = 0,01)	M com F = 2,90 (P valor = 0,05)	M com F = 2,27 (P valor = 0,10)	M com F $\rightarrow 1,00$
1	5,1	5,1	5,2	5,2	5,3	5,517
2	4,3	4,3	4,6	4,7	4,8	5,517
3	4,0	4,0	4,3	4,5	4,7	5,517
4	6,7	6,7	6,4	6,3	6,2	5,517
5	6,0	6,0	5,9	5,8	5,7	5,517
6	7,0	7,0	6,8	6,5	6,3	5,517

Constata-se que a média amostral eqüivale ao valor mais provável da média, apenas quando $F \rightarrow \infty$ ou seja quando P valor $\rightarrow 0$. Com a redução de $F \rightarrow \infty$ para $F = 1$, o valor mais provável da média converge para a média geral e a variância entre as médias converge para zero. No caso, a inferência pontual mais correta para a média do melhor tratamento (tratamento 6) é 6,5 (visto que o P – valor obtido foi de 0,05) e não 7,0. Verifica-se também que a diferença absoluta entre o melhor e o pior tratamento cai de 3,0 para 2,0, reduzindo-se em 33%. É importante relatar que o cômputo do P-valor só é válido quando se tem os efeitos de tratamentos como fixos, sendo que neste artigo a apresentação do P-valor foi apenas ilustrativa.

Do ponto de vista da estatística clássica ou freqüentista, a média amostral é o melhor estimador não viciado (BLUE) da média populacional. O estimador apresentado, que apresenta um regressor (no caso $1-1/F$) da média amostral do tratamento em relação à média geral é viciado (Gianola & Fernando, 1986). Entretanto, existem estimadores e preditores viciados que propiciam menor erro quadrático médio do que os estimadores BLUE (Henderson, 1984), sendo assim vantajosos em relação aos estimadores não viciados (Efron, 1975).

Os três principais procedimentos de estimação caracterizam os estimadores: melhor estimador linear não viciado (BLUE); estimador de mínimo erro quadrático médio (EMMSE) e estimador Bayesiano (EB). Esses três estimadores são similares e idênticos sob certas circunstâncias (Henderson, 1984).

A utilização e reconhecimento das vantagens dos estimadores viciados iniciou-se com o trabalho de Stein (1955) que constituiu um verdadeiro paradoxo na Estatística. Stein demonstrou que a média aritmética é estimador não admissível, isto é, que existem estimadores que propiciam menor erro quadrático médio ou menor risco que a média aritmética. Neste contexto, James & Stein (1961) apresentaram um estimador melhorado para a média populacional, o qual é dado por $M^* = k (\bar{Y}_{i..} - \bar{Y}_{...}) + \bar{Y}_{...}$, onde k é um fator regressor da média amostral de determinado tratamento sobre a média geral. Constata-se assim, a similaridade entre M e M^* .

O fator k apresentado por James & Stein é dado por

$$K = 1 - \frac{(t-3) \sigma^2}{\sum (\bar{Y}_{i..} - \bar{Y}_{...})^2},$$
 onde t refere-se ao número de tratamentos e σ^2

refere-se à variância residual. Notando a similaridade entre $(t-3)$ e o número de graus de liberdade $(t-1)$ para tratamentos, verifica-se também a grande similaridade entre k e $(1-1/F)$. Stein (1955) demonstrou que a média aritmética é um estimador admissível apenas quando existem apenas uma ou duas médias a serem estimadas, de forma que a correção somente torna-se necessária quando três ou mais tratamentos forem considerados, conforme intuitivamente esperado. O estimador melhorado apresentado por James & Stein (1961) não necessita de qualquer suposição referente aos efeitos fixos ou aleatórios ou distribuições das médias a serem estimadas (Efron & Morris, 1977). Os estimadores de mínimo erro quadrático médio tem recebido mais atenção recentemente (Bibby & Toutenburg, 1977).

O procedimento de estimação Bayesiana é mais antigo (1763) do que o método de Stein e também minimiza o erro quadrático esperado, de forma que o estimador de James-Stein é muito similar ao estimador de Bayes, tornando-se idênticos para grande número de tratamentos (Efron & Morris, 1977), sendo denominados de estimadores de Bayes-Stein. Em inferência Bayesiana não existe qualquer distinção entre efeitos fixos ou aleatórios, sendo que os parâmetros a serem estimados são considerados variáveis aleatórias (Gianola & Fernando, 1986) que devem ser estimadas considerando as incertezas associadas a elas. No caso de inferências sobre médias populacionais de tratamentos, sob o enfoque Bayesiano, Box & Tiao (1973) apresentam como regressor exatamente a quantidade $(1-1/F)$.

Com base no exposto, recomenda-se, para inferências (com menor risco) práticas na área de pesquisa agropecuária e florestal, a apresentação das médias de tratamento como $M = (1-1/F) \bar{Y}_{i..} + (1/F) \bar{Y}_{...}$, onde F , $\bar{Y}_{i..}$ e $\bar{Y}_{...}$ referem-se ao valor de F para tratamentos na análise de variância, média de tratamentos e média geral, respectivamente.

Referências Bibliográficas

BIBBY, J.; TOUTENBURG, H. **Prediction and improved estimation in linear models**. New York: J. Wiley and Sons, 1977.

- BOX, G.E.P.; TIAO, G.C. **Bayesian inference in statistical analysis**. Reading: Addison-Wesley Publ., 1973, 588p.
- EFRON, B. Biased versus unbiased estimation. **Advances in Mathematics**, v.16, p.259-277, 1975.
- EFRON, B.; MORRIS, C. Stein's paradox in statistics. **Scientific American**, v.236, n.5, p.119-127, 1977.
- GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217-244, 1986.
- HENDERSON, C.R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984, 462p.
- JAMES, W.; STEIN, C. Estimation with quadratic loss. **Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability**, v.1, p.361-379, 1961.
- RESENDE, M.D.V. de. **Predição de valores genéticos, componentes de variância, delineamentos de cruzamento e estrutura de populações no melhoramento florestal**. Curitiba: Universidade Federal do Paraná, 1999. 434p. Tese Doutorado.
- STEEL, R.G.D.; TORRIE, J.H. **Principles and procedures of statistics**. New York: Mc Graw-Hill, 1980. 633p.
- STEIN, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. **Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability**, v.1, p.197-206, 1955.