

PRECISÃO E CONTROLE DE QUALIDADE EM EXPERIMENTOS DE AVALIAÇÃO DE CULTIVARES¹

Marcos Deon Vilela de Resende², João Batista Duarte³

ABSTRACT

PRECISION AND QUALITY CONTROL IN VARIETY TRIALS

This study had as objective to propose a new approach for quality evaluation of variety trials for determination of cropping and use values (VCU), which considers three attributes simultaneously: magnitude of the residual variation, replication number, and genetic control of the trait under selection. It was also emphasized the need for using shrinkage estimators/predictors of genotypic values instead of unshrunk phenotypic means of varieties, i.e., the procedures should consider the genetic coefficient of determination of the traits, as well as the eventual heterogeneity of residual variance within varieties. Targeting an accuracy of 90%, it was concluded that Snedecor F test values associated to treatment effects in the analysis of variance should be above 5.0. The magnitude of genotypic variability of the traits is also involved in the F statistics. This means that the approach of fixing minimum values for replication number and maximum values for residual variation coefficient (C_{Ve}) is not sufficient. For traits related to yield (with low genetic coefficient of determination) the normally used replication number, between two and four, does not permit to reach the targeted accuracy, even if residual variation coefficients below 10% are aimed, and the experimentation is conducted on several sites and years. For that target accuracy it is recommended the use of at least six replications. It was also shown that shrinkage estimators provide more precise and reliable inferences concerning genotypic means of the varieties, and their use is encouraged.

KEY WORDS: accuracy, shrinkage estimator, variance heterogeneity, biased estimator, variation coefficient.

RESUMO

O presente estudo teve como objetivo propor uma nova abordagem para a avaliação da qualidade dos ensaios de avaliação do valor de cultivo e uso (VCU) de cultivares, a qual considera, simultaneamente, três atributos: magnitude da variação residual, número de repetições e controle genético dos caracteres. Enfatizou também a necessidade do uso de métodos de estimação/predição de valores genotípicos que promovam shrinkage sobre a média fenotípica do cultivar; ou seja, que considerem um coeficiente de determinação genética, bem como a possível heterogeneidade de variâncias residuais entre as cultivares. Concluiu-se que, para atingir uma meta de acurácia de 90%, os valores do teste F de Snedecor associados aos efeitos de cultivares, na análise da variância, devem ser superiores a 5,0. O uso da estatística F é uma alternativa para também levar em consideração o nível de variabilidade genética dos caracteres. Assim, não é suficiente fixar-se um número mínimo de repetições e um valor máximo para o coeficiente de variação experimental (C_{Ve}). Para caracteres de produção, em geral com baixo coeficiente de determinação genética, os números de repetições usualmente empregados (entre dois e quatro) não permitem atingir essa meta de acurácia seletiva, mesmo quando se perseguem valores de C_{Ve} inferiores a 10% e a experimentação é realizada em vários locais e anos. Para isso, recomendam-se ao menos seis repetições na avaliação desses caracteres. Demonstra-se, ainda, que os estimadores/preditores que promovem shrinkage sobre a média fenotípica garantem inferências mais realistas e precisas acerca das médias genotípicas das cultivares, devendo ser encorajados.

PALAVRAS-CHAVE: acurácia, estimador shrinkage, heterogeneidade de variâncias, estimador viciado, coeficiente de variação.

INTRODUÇÃO

Experimentos de campo são essenciais nos programas de melhoramento genético e também no processo de recomendação de cultivares melhoradas. Nesses experimentos é desejável um alto grau de

precisão experimental e, conseqüentemente, uma alta acurácia na inferência sobre as médias genotípicas, isto é, sobre os valores genotípicos dos tratamentos genéticos em avaliação. No Brasil, essa média genotípica é denominada valor de cultivo e uso (VCU), pelo Ministério da Agricultura, para fins legais

1. Trabalho recebido em dez./2005 e aceito para publicação em set./2007 (registro nº 677).

2. Embrapa Florestas, Caixa Postal 319, CEP 83411-000 Colombo-PR. E-mail: deon@cnpf.embrapa.br

3. Escola de Agronomia e Engenharia de Alimentos, Setor de Melhoramento de Plantas, Universidade Federal de Goiás. Caixa Postal 131, CEP 74001-970 Goiânia-GO. Email: jbduarte@agro.ufg.br

de recomendação, registro e proteção de cultivares (Brasil 2001).

Alguns parâmetros estatísticos têm sido recomendados para a avaliação da precisão e qualidade dos experimentos agrícolas. Tradicionalmente, tem sido recomendado (Pimentel Gomes 1987) e utilizado (Estefanel *et al.* 1987, Garcia 1989, Scapim *et al.* 1995, Amaral *et al.* 1997, Lúcio 1997, Judice *et al.* 1999, Clemente & Muniz 2000, 2002, Ramalho *et al.* 2000, Storck *et al.* 2000, Judice *et al.* 2002, Costa *et al.* 2002) o coeficiente de variação experimental (CVe), que deve ser mantido em níveis adequados para cada espécie cultivada e caráter sob avaliação. O valor de CVe depende apenas da variação residual como proporção da média do experimento. Outro parâmetro relacionado à precisão experimental é o índice de variação proposto por Pimentel Gomes (1991) e, também, denominado coeficiente de precisão experimental (CPe), por Storck *et al.* (2000). O índice CPe é mais adequado que o CVe, pois considera também o número de repetições usado na experimentação, além da variação de natureza residual.

Os ensaios de avaliação de cultivares devem ser abordados do ponto de vista genético e estatístico, e não apenas sob a perspectiva estatística. Nesse contexto, um dos parâmetros mais relevantes para a avaliação da qualidade de um experimento, embora ainda pouco utilizado, é a acurácia seletiva. Esta tem a propriedade de informar sobre o correto ordenamento das cultivares para fins de seleção e, também, sobre a eficácia da inferência acerca do valor genotípico da cultivar, isto é, do seu VCU (Resende 2002). Tal parâmetro não depende apenas da magnitude da variação residual e do número de repetições, mas também da proporção entre as variações de natureza genética e residual associadas ao caráter em avaliação.

O presente estudo teve como objetivo propor uma nova abordagem para a avaliação da qualidade dos experimentos de avaliação de cultivares, a qual considera, simultaneamente, os três atributos mencionados: magnitude da variação residual, número de repetições e controle genético dos caracteres. Sem perda de generalidade, este artigo enfoca principalmente os chamados ensaios de VCU planejados e conduzidos sob o delineamento de blocos completos casualizados. Buscou-se também demonstrar que métodos de estimação/predição de valores genotípicos que promovem *shrinkage* das médias

fenotípicas e permitem acomodar heterocedasticidade entre as cultivares garantem inferências mais precisas e, por conseguinte, mais realistas acerca da avaliação genotípica.

METODOLOGIA

A avaliação de tratamentos genéticos em experimentos de campo tem dois objetivos: *i)* inferir sobre os valores genotípicos desses materiais; e *ii)* ordená-los com base em seus valores genotípicos para fins de seleção. O interesse, portanto, não está em estimar as suas médias fenotípicas, mas sim valores genotípicos, isto é, médias futuras desses genótipos, quando estes forem cultivados novamente em plantios comerciais. Neste caso, mesmo que o sejam no mesmo local ou região da experimentação, os efeitos de blocos e de parcelas muito provavelmente não se repetirão. Como tais efeitos estão embutidos, em alguma proporção, nas médias fenotípicas, isso demonstra que tais médias não são adequadas para a inferência sobre os valores genotípicos. Assim, em trabalhos científicos e em catálogos de avaliação genética, a apresentação de médias fenotípicas não é desejável. Pelo contrário, deve-se buscar a apresentação das médias genotípicas (livres de efeitos ambientais), isto é, dos valores de cultivo e uso (VCU) propriamente ditos.

Na estimação ou predição dos valores genotípicos, o mais importante é a escolha do método de estimação/predição. Esse método deve propiciar a inferência mais precisa e realista possível, o que deve ser avaliado segundo parâmetros estatísticos adequados. Segundo Henderson (1984), no contexto da avaliação genotípica, o parâmetro estatístico mais importante é a acurácia seletiva (\hat{r}_{gg}). Este parâmetro refere-se à correlação entre o valor genotípico verdadeiro do tratamento genético e aquele estimado ou predito a partir das informações dos experimentos. Como correlação, varia de 0 a 1, e os valores adequados de acurácia são aqueles próximos à unidade ou 100%. Logo, é natural que valores elevados de acurácia sejam almejados nos experimentos de avaliação de cultivares. A acurácia é tanto mais alta quanto menores forem os desvios absolutos entre os valores genéticos paramétricos e os valores genéticos estimados ou preditos. Tais desvios podem ser avaliados pela estatística erro quadrático médio (EQM). Este equivale à distância Euclideana média entre as estimativas e os correspondentes parâmetros,

e é dado por: $EQM = (\text{Vício})^2 + PEV$; em que PEV é a variância do erro de predição. Assim, EQM congrega, simultaneamente, os conceitos de vício e de precisão, os quais estão implícitos no conceito de acurácia. Um estimador ou preditor acurado do valor genotípico apresenta vício nulo ou pequeno, e alta precisão (baixa variância do erro de predição). Por isso, minimizar EQM significa maximizar a acurácia, e o método ideal de estimação ou predição é aquele que permite alcançar este objetivo. Demonstra-se que esse método pode ser viciado em pequeno grau, pois o que importa é minimizar a soma: $(\text{Vício})^2 + PEV$ (Henderson 1984).

Na classe dos estimadores/preditores não viciados, EQM quantifica a precisão, a qual é dada pela variância do erro de predição (PEV). Esta estatística é relacionada à acurácia por meio da equação (Henderson 1984): $\hat{r}_{gg} = (1 - PEV / \sigma_g^2)^{1/2}$; a qual resulta em: $PEV = (1 - \hat{r}_{gg}^2) \sigma_g^2$, sendo σ_g^2 a variação genotípica entre os tratamentos genéticos (neste momento, ainda livre de qualquer suposição associada aos seus efeitos). Assim, quanto menor o valor de PEV, maior é a acurácia e, também, a precisão. Logo, na classe dos estimadores/preditores não viciados, a estratégia de minimizar PEV conduz também à maximização da acurácia. Mas, de maneira geral, relaxando-se a necessidade de não vício, o que deve ser minimizado é EQM.

A média fenotípica ou média aritmética simples estimada pelo método de quadrados mínimos ordinário, embora sabidamente livre de vício, não é estimador de EQM mínimo para mais que dois tratamentos genéticos. O trabalho de Stein (1955), que constituiu um verdadeiro paradoxo na Estatística, demonstrou que a média aritmética é estimador não admissível, pois existem estimadores que propiciam menor erro quadrático médio ou, ainda, menor risco, quando mais que duas médias necessitam ser estimadas. Neste contexto, James & Stein (1961) apresentaram um estimador melhorado para a média populacional, dado por: $M^* = k (\bar{Y}_{i..} - \bar{Y}_{...}) + \bar{Y}_{...}$, em que k é um fator regressor ou de *shrinkage* da média amostral de determinado tratamento i ($\bar{Y}_{i.}$) sobre a média geral ($\bar{Y}_{...}$).

Os métodos que minimizam EQM, viciados ou não, em geral conduzem a estimadores/preditores do tipo *shrinkage* (Efron & Morris 1977). Genericamente, um estimador desse tipo é expresso pela multiplicação de um escalar, entre 0 e 1, por um vetor

de médias estimadas por quadrados mínimos ou por máxima verossimilhança; isto é, multiplicam-se as médias fenotípicas por um fator de "encolhimento" que depende da confiabilidade que se tem naquelas estimativas. Tal procedimento, às vezes, é usado empiricamente por economistas, administradores e melhoristas, que já constataram a não reprodução das médias fenotípicas dos ensaios, nos plantios comerciais, bem como o fato de a média desses plantios geralmente ser inferior à média experimental. Então, multiplicam as médias fenotípicas por seus coeficientes empíricos de confiabilidade. Embora esses profissionais, a rigor, estejam usando estimado-res *shrinkage*, eles não o fazem de maneira otimizada.

No caso dos testes genéticos, o fator de confiabilidade ótimo é função do coeficiente de determinação genotípica associado ao caráter em avaliação, o que corresponderia ao coeficiente de herdabilidade, num processo de seleção intra-populacional. Assim, a forma de eliminar os efeitos residuais de ambiente embutidos nos dados fenotípicos é multiplicando-se o valor fenotípico corrigido, pelo fator de *shrinkage*. Essa abordagem é generalizada para outros tipos experimentos, sob o enfoque da chamada recuperação de informação inter-efeitos (Federer 1996, Wolfinger *et al.* 1997, Federer 1998, Federer & Wolfinger 1998).

Vários autores (Hill & Rosenberger 1985, Stroup & Miltz 1991, Piepho 1994, Resende *et al.* 1996, Piepho 1998, Smith *et al.* 2001, Duarte 2000, Duarte & Vencovsky 2001, Resende 1999, 2002, 2004) enfatizam a necessidade de se utilizar estimadores/preditores do tipo *shrinkage*, mesmo se esses efeitos em estudo forem tidos como fixos por abordagens tradicionais. Tal justificativa decorre da equivalência algébrica desta abordagem com aquela baseada na suposição de aleatoriedade para tais efeitos, vinculando-os à uma população conceitual de média e variância comuns (Henderson 1984, Wolfinger *et al.* 1997). No caso dos tratamentos genéticos, essa população corresponde àquela representada pela amostra das cultivares em teste, cuja variância intergenotípica denota-se aqui por σ_g^2 .

Estimadores do tipo *shrinkage* são usados no método da melhor predição linear não viciada (BLUP), conforme Henderson (1984). Esse método assume os efeitos de genótipos como aleatórios e o BLUP é, adicionalmente, um preditor não viciado. Segundo Stein (1955), para mais que dois tratamentos, estimadores do tipo *shrinkage* são sempre

necessários, independentemente se os efeitos forem tomados como fixos ou aleatórios. O estimador melhorado de James & Stein (1961) não necessita mesmo de qualquer suposição a esse respeito, ou sobre distribuições das médias a serem estimadas (Efron & Morris 1977). Requer apenas o relaxamento da suposição de não vício. Tal estimador é viciado, mas tem menor erro quadrático médio que o estimador de quadrados mínimos. No contexto da avaliação genotípica, o vício associado ao estimador de James-Stein é pequeno e assume relevância somente quando o número de tratamentos é baixo (inferior a dez). À medida que este número se eleva, o vício tende a ser menor e, por isso, é denominado como um estimador aproximadamente não viciado. Estimadores assim ganham aplicabilidade quando aqueles não viciados geram estimativas fora do espaço paramétrico admissível.

Em síntese, na avaliação genotípica a relevância maior não está entre escolher se os tratamentos são de efeitos fixos ou aleatórios, mas, sim, na escolha de estimadores/preditores mais acurados e de mínimo erro quadrático médio. É nesse sentido que ganharam importância os estimadores de James & Stein (1961), também reconhecidos como estimadores *shrinkage*. Estes não assumem aleatoriedade, mas propiciam, com o aumento do número de tratamentos em avaliação, uma transição natural de efeitos puramente fixos para efeitos completamente aleatórios. E isso só depende do tamanho da amostra (número de tratamentos). Com grande número de tratamentos (algo acima de dez), tais estimadores e o método BLUP se equivalem (Gianola 1990, Weigel *et al.* 1991). Neste caso, porém, o BLUP tem as vantagens da facilidade de implementação e da extensão para o caso não balanceado.

Por outro lado, pela abordagem aqui defendida, os estimadores de VCU, quase sempre, deveriam ser baseados em modelos aleatórios ou mistos com tratamentos genéticos de efeitos aleatórios. Isso porque o uso da informação intergenotípica, decorrente dessa suposição de aleatoriedade, corresponde a obter uma "concessão" para explorar estatisticamente a informação de dependência entre as cultivares, as quais podem estar relacionadas por uma origem comum à já referida população conceitual (Henderson 1984). Neste caso, a interação dos genótipos com os efeitos ambientais (GE) também será um efeito aleatório, fato que permite produzir inferências para toda a população de ambientes.

Admitindo-se, portanto, os efeitos de tratamentos como aleatórios, o estimador da acurácia é dado por (Resende 2002):

$$\hat{r}_{gg} = \left[\frac{bh_i^2}{1 + (b-1)h_i^2} \right]^{1/2} = \left[\frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/b} \right]^{1/2} = \left[\frac{1}{1 + (\sigma_e^2/b)/\sigma_g^2} \right]^{1/2} \quad (I)$$

em que:

b: número de repetições ou blocos (assumindo-se os ensaios típicos em blocos completos casualizados, usuais na avaliação de VCU);

$h_i^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$: coeficiente de determinação genética entre os cultivares (interpopulacional), em nível de parcelas, associado ao caráter em avaliação (a notação h^2 decorre da correspondência com o coeficiente intrapopulacional de herdabilidade no sentido amplo, em nível individual);

σ_g^2 : variância genotípica entre cultivares; e

σ_e^2 : variância residual ou variância dentro de cultivares.

Verifica-se, portanto, que a acurácia depende basicamente da proporção entre a variação residual média e a variação genotípica, sendo que a variação residual média depende do número de repetições.

Os parâmetros de precisão usados tradicionalmente são:

$CVe = \frac{\sigma_e}{m} \times 100$: coeficiente de variação experimental, expresso em porcentagem, sendo "m" a média geral do experimento; e

$CPe = \frac{\sigma_e}{(b)^{1/2} m} \times 100 = \frac{CVe}{(b)^{1/2}}$: coeficiente de precisão experimental, expresso também em porcentagem.

Em termos de parâmetros tradicionais como esses, a acurácia definida na equação I pode, ainda, ser expressa como:

$$\hat{r}_{gg} = \left[\frac{1}{1 + CPe^2 / CVg^2} \right]^{1/2} = \left[\frac{1}{1 + (CVe^2 / CVg^2) / b} \right]^{1/2} = \left[1 - \frac{1}{1 + b CPe^2} \right]^{1/2} \quad (II)$$

em que:

$CVg = \frac{\sigma_g}{m} \times 100$: coeficiente de variação genotípica; e

$CVr = \frac{CVg}{CVe}$: coeficiente de variação relativa.

Portanto, em função destes parâmetros, a acurácia mostra-se dependente da proporção entre os coeficientes de variação experimental e de variação genotípica, bem como do número de repetições. Dessa forma, avaliar a qualidade dos

experimentos de competição de cultivares com base apenas no coeficiente de variação experimental ou no coeficiente de precisão experimental revela-se como critérios inadequados, haja vista não se considerar o nível da variação genotípica expressa no caráter. Ademais, esses critérios não informam sobre a acurácia que está sendo praticada na seleção das cultivares. O uso da relação entre os coeficientes de variação genotípica e residual foi proposto por Vencovsky (1987), para inferir sobre as possibilidades de sucesso no melhoramento genético. Entretanto, também nessa proposta não se vinculou tal relação ao número de repetições.

Conforme Resende (2002), a acurácia pode ainda ser expressa como: $\hat{r}_{gg} = (1 - 1/F)^{1/2}$, em que F (de Snedecor) é o valor da razão de variâncias para os efeitos de tratamentos (cultivares), associada à análise de variância (ANOVA). Assim, um parâmetro adequado para avaliar a qualidade dos experimentos de avaliação de cultivares pode ser sumarizado em termos de uma única estatística, a qual contempla, simultaneamente, o coeficiente de variação experimental, o número de repetições e o coeficiente de variação genotípica. A expressão a seguir mostra como o valor F contempla os três parâmetros mencionados: $F = 1 + bCV_g^2/CV_e^2$.

A estatística PEV também pode ser expressa, alternativamente, em função dos parâmetros anteriormente considerados:

$$PEV = \left[\frac{\sigma_e^2/b}{1 + (\sigma_e^2/b)/\sigma_g^2} \right] = \left[\frac{\sigma_e^2/b}{1 + CP_e^2/CV_g^2} \right] = \left[\frac{\sigma_e^2/b}{1 + (CV_e^2/CV_g^2)/b} \right] \quad (III)$$

No contexto da estimação/predição por modelos mistos, a abordagem tradicional que considera os efeitos de cultivares como fixos corresponde, implicitamente, à suposição de variação genotípica tendendo ao infinito (SAS Institute 1997), ou seja, de que o coeficiente de determinação genotípica tende para 1,0. Neste caso, a expressão de PEV torna-se: $PEV = \sigma_e^2/b$, o que equivale ao quadrado do erro padrão da média (EPM^2), ou, ainda, à variância da média de tratamento num modelo fixo. Outra forma de expressar esta estatística é:

$$PEV = [(1 - 1/F)]\sigma_e^2/b = \hat{r}_{gg}^2\sigma_e^2/b = \hat{r}_{gg}^2EPM^2 \quad (IV)$$

Por essa expressão verifica-se também que a abordagem tradicional de se considerar genótipos como de efeitos fixos assume, ainda, que a acurácia seletiva tende para a unidade. E, isso não é verdadeiro para caracteres quantitativos. Ademais, verifica-se que PEV é menor ou igual ao quadrado do erro padrão

da média, como função da confiabilidade ou quadrado da acurácia (\hat{r}_{gg}^2), que varia entre 0 e 1. Logo, usar estimadores do tipo *shrinkage* ou assumir genótipos como de efeitos aleatórios constitui-se numa abordagem que, quase sempre, é mais realista e resulta em maior precisão (menor PEV).

Neste contexto, assumir os efeitos de genótipos como aleatórios não implica, necessariamente, a obrigação de estimar componentes de variância explicitamente. O coeficiente de determinação genética necessário ao processo de estimação/predição pode ser obtido diretamente do valor F da ANOVA, o qual é normalmente obtido mesmo quando os efeitos de genótipos são tratados como fixos. Nesse caso, o valor genotípico predito é dado por Resende (1999, 2002): $\hat{g} = VCU = m + (1 - 1/F)(\bar{Y}_i - m)$, com $i = 1, 2, \dots, T$ tratamentos genéticos. Esta abordagem usa implicitamente todos os fundamentos genéticos da avaliação genotípica e atende, simultaneamente, às quatro abordagens científicas para o problema da estimação/predição de efeitos de tratamentos (Weigel *et al.* 1991, Resende 2002): abordagem tradicional de efeitos fixos (admitindo-se o uso de estimadores viciados); metodologia de modelos mistos (com tratamentos de efeitos aleatórios); estimadores *shrinkage* de James-Stein; e abordagem Bayesiana. É importante mencionar que, sob este último enfoque, os efeitos de tratamentos são sempre considerados aleatórios, de forma que naturalmente se utilizam estimadores que promovem *shrinkage*, mesmo com pequeno número de tratamentos (menor que dez).

Do ponto de vista da estatística clássica ou freqüentista, a média amostral é o melhor estimador linear não viciado (BLUE) da média populacional. O estimador apresentado, que introduz um regressor da média amostral do tratamento, $(1 - 1/F)$, em relação à média geral, no caso 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).

Stein (1955) também demonstrou que a média aritmética é um estimador admissível, porém, apenas quando existem uma ou duas médias a serem estimadas. Assim, o fator k apresentado por James

& Stein (1961) é dado por: $k = 1 - \frac{(T-3)\sigma^2}{\sum (\bar{Y}_{i..} - \bar{Y}_{...})^2}$, sendo $T \geq 3$ tratamentos e σ^2 a variância residual. Notando-

se a similaridade entre $(T - 3)$ e o número de graus de liberdade para tratamentos, $(T - 1)$, verifica-se também grande similaridade entre k e $(1 - 1/F)$. Logo, a correção é necessária somente quando mais que três tratamentos forem considerados. Os estimadores desse tipo, que envolvem *shrinkage* e EQM mínimo, sem qualquer suposição referente a efeitos fixos ou aleatórios, têm recebido grande atenção dos cientistas (Theil 1971, Bibby & Toutenburg 1977, Vinod & Ullah 1981, Casella 1985, Judge *et al.* 1985).

O procedimento de estimação bayesiana também minimiza o erro quadrático esperado. Por isso, o estimador de James-Stein é muito similar ao estimador de Bayes, tornando-se, inclusive, idênticos para grande número de tratamentos (Efron & Morris 1977). Assim, são também denominados estimadores de Bayes-Stein, Bayes empírico ou regra empírica de Bayes. Em inferência Bayesiana não existe distinção entre efeitos fixos ou aleatórios, e os parâmetros a serem estimados são considerados variáveis aleatórias (Gianola & Fernando 1986). No caso de inferências sobre médias populacionais de tratamentos, sob o enfoque Bayesiano, Box & Tiao (1973) apresentam como regressor, exatamente, a quantidade $(1 - 1/F)$.

RESULTADOS E DISCUSSÃO

Inferência sobre a precisão experimental

A partir das expressões apresentadas, geraram-se tabelas de resultados que permitem simular situações experimentais e, com isso, avaliar a qualidade dos respectivos experimentos, com base na abordagem proposta. Na Tabela 1 apresentam-se valores da estatística F (de Snedecor) para os efeitos de tratamentos na análise da variância, relacionados ao nível de acurácia obtido ou requerido: $F = 1/(1 - \hat{r}_{gg}^2)$. São destacados na tabela os valores de F que devem ser obtidos, visando-se estimações/predições de VCU com elevada acurácia. Observa-se, por exemplo, que para se atingir uma acurácia seletiva ideal de 90% ou mais – correspondente a valores de coeficiente de determinação acima de 80%, conforme preconizam Steel & Torre (1980), para uma inferência estatística segura – os valores de F para cultivares devem ser iguais ou superiores a 5,26. Por conseguinte, este pode ser um valor de referência para os experimentos de avaliação de VCU. Tal valor independe da espécie cultivada e do

caráter sob avaliação, podendo servir como padrão para avaliar experimentos dessa natureza.

Atualmente, têm sido estabelecidos valores máximos do coeficiente de variação para experimentos de VCU com diversas culturas, por exemplo, 20% para feijão e 30% para capim elefante (Brasil 2001). Esses valores, contudo, são relativamente empíricos na medida em que não informam sobre a acurácia seletiva da avaliação e, conseqüentemente, não consideram o nível de variação genotípica e o número de repetições. Contrariamente, o uso da estatística F é uma alternativa que permite considerar simultaneamente todos esses aspectos na avaliação dos ensaios. Admitindo-se, por exemplo, que nos processos de seleção em melhoramento de plantas, valores de acurácia superiores a 70% devam ser buscados, ensaios cujo valor de F para cultivares for inferior a 2,0 (Tabela 1) não devem ser aproveitados. Isso porque sequer permitem alcançar esta acurácia seletiva mínima.

Outra estatística empregada em avaliação genotípica é o coeficiente de variação relativa (CVr), que se relaciona à acurácia através da expressão final na equação II. Fixando-se o número de repetições, a magnitude de CVr pode ser utilizada para inferir sobre a acurácia e a precisão na avaliação genotípica (Tabela 2). Nesta tabela destacam-se os valores de CVr necessários para se obter acurácias seletivas em torno de 90%, para diferentes números de repetições empregados na experimentação. Tais valores variam, por exemplo, de 1,50 para o caso de

Tabela 1. Valores adequados do teste F (de Snedecor) para os efeitos de tratamentos (cultivares), na análise de variância, visando-se atingir determinada acurácia seletiva, e as categorias de precisão requerida na avaliação genotípica.

Acurácia	Classes de precisão	Valor de F
0,99	Muito alta	50,25
0,975	Muito alta	20,25
0,95	Muito alta	10,26
0,90	Muito alta	5,26
0,85	Alta	3,60
0,80	Alta	2,78
0,75	Alta	2,29
0,70	Alta	1,96
0,65	Moderada	1,73
0,60	Moderada	1,56
0,55	Moderada	1,43
0,50	Moderada	1,33
0,40	Baixa	1,19
0,30	Baixa	1,10
0,20	Baixa	1,04
0,10	Baixa	1,01

duas repetições a 0,70 para o caso de dez repetições. Já com trinta ou quarenta repetições, é possível alcançar-se tal meta de acurácia, mesmo quando a relação CVg/CVe é inferior a 0,40. Logo, valores adequados ou não de CVr devem ser inferidos em conjunto com o número de repetições.

Vencovsky (1987) reporta que valores de CVr em torno da unidade são adequados para a experimentação com milho. De fato, confirma-se na Tabela 2 que este valor garante acurácias altas a muito altas. Entretanto, com número de repetições superiores a cinco, valores de CVr inferiores à unidade também podem propiciar elevadas acurácia e precisão. Tais informações podem, ainda, serem utilizadas de forma reversa, isto é, a partir do valor de CVr observado em ensaios de VCU, é possível inferir sobre o número de repetições adequado para se conseguir determinada meta de acurácia seletiva.

Ainda com base nas expressões de acurácia apresentadas na equação II, simularam-se valores alternativos dos coeficientes de variação genotípica e experimental, necessários para se conseguir acurácias em torno de 90% na avaliação genotípica (Tabela 3). Verifica-se que com os números de repetições normalmente empregados, entre dois e quatro, não é possível atingir níveis desejados de acurácia para a maioria dos caracteres de interesse nessas avaliações. Isso porque, sob baixo número de

repetições, uma acurácia de 90% apenas poderia ser conseguida para caracteres de elevado coeficiente de determinação genética ($h^2_i \geq 0,60$), fato que é improvável de ser realidade para caracteres quantitativos. Nessa situação, as magnitudes do coeficiente de variação experimental são inadequadas para informar sobre precisão, mesmo quando valores baixos como 10% são perseguidos. Assim, para os caracteres de produção, com $h^2_i \leq 0,40$, ao menos seis repetições seriam necessárias. E, neste caso, diversos valores de CVe são aceitáveis, desde que mantenham uma relação próxima a 0,82 com CVg (última coluna da Tabela 3).

Em síntese, algumas recomendações experimentais atualmente difundidas (ex. três repetições e CVe mínimo de 20%), para os testes de VCU, não são adequadas para se atingir metas de acurácia seletiva iguais ou superiores a 90%. Isso, tanto em termos do número de repetições, quanto em relação ao coeficiente de variação experimental. A estatística que, de fato, facilita e agrega as informações mais relevantes sobre a qualidade experimental é o valor do teste F para os efeitos de tratamentos, que, neste caso, deve ser superior a 5,0 (Tabela 1). Esta abordagem pode ser empregada para qualquer caráter e espécie cultivada, embora deva-se relembrar que, para caracteres de produção, pelo menos seis repetições são necessárias para se atingir tal valor de F.

Tabela 2. Valores de acurácia seletiva para diversos coeficientes de variação relativa (CVr), sob diferentes números de repetições (b).

CVr ¹	b											
	2	3	4	5	6	7	8	9	10	20	30	40
0,10	0,14	0,17	0,20	0,22	0,24	0,26	0,27	0,29	0,30	0,41	0,48	0,53
0,20	0,27	0,33	0,37	0,41	0,44	0,47	0,49	0,51	0,53	0,67	0,74	0,78
0,25	0,33	0,40	0,45	0,49	0,52	0,55	0,58	0,60	0,62	0,75	0,81	0,85
0,30	0,39	0,46	0,51	0,56	0,59	0,62	0,65	0,67	0,69	0,80	0,85	0,88
0,40	0,49	0,57	0,62	0,67	0,70	0,73	0,75	0,77	0,78	0,87	0,91	0,93
0,50	0,58	0,65	0,71	0,75	0,77	0,80	0,82	0,83	0,85	0,91	0,94	0,95
0,60	0,65	0,72	0,77	0,80	0,83	0,85	0,86	0,87	0,88	0,94	0,96	0,97
0,70	0,70	0,77	0,81	0,84	0,86	0,88	0,89	0,90	0,91	0,95	0,97	0,98
0,75	0,73	0,79	0,83	0,86	0,88	0,89	0,90	0,91	0,92	0,96	0,97	0,98
0,80	0,75	0,81	0,85	0,87	0,89	0,90	0,91	0,92	0,93	0,96	0,97	0,98
0,90	0,79	0,84	0,87	0,90	0,91	0,92	0,93	0,94	0,94	0,97	0,98	0,98
1,00	0,82	0,87	0,89	0,91	0,93	0,94	0,94	0,95	0,95	0,98	0,98	0,99
1,25	0,87	0,91	0,93	0,94	0,95	0,96	0,96	0,97	0,97	0,98	0,99	0,99
1,50	0,90	0,93	0,95	0,96	0,96	0,97	0,97	0,98	0,98	0,99	0,99	0,99
1,75	0,93	0,95	0,96	0,97	0,97	0,98	0,98	0,98	0,98	0,99	0,99	1,00
2,00	0,94	0,96	0,97	0,98	0,98	0,98	0,98	0,99	0,99	0,99	1,00	1,00
2,25	0,95	0,97	0,98	0,98	0,98	0,99	0,99	0,99	0,99	1,00	1,00	1,00
2,50	0,96	0,97	0,98	0,98	0,99	0,99	0,99	0,99	0,99	1,00	1,00	1,00
2,75	0,97	0,98	0,98	0,99	0,99	0,99	0,99	0,99	0,99	1,00	1,00	1,00
3,00	0,97	0,98	0,99	0,99	0,99	0,99	0,99	0,99	0,99	1,00	1,00	1,00
3,25	0,98	0,98	0,99	0,99	0,99	0,99	0,99	0,99	1,00	1,00	1,00	1,00
3,50	0,98	0,99	0,99	0,99	0,99	0,99	0,99	1,00	1,00	1,00	1,00	1,00
3,75	0,98	0,99	0,99	0,99	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00
4,00	0,98	0,99	0,99	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00

¹- Os valores CVr destacados em negrito são aqueles cuja acurácia seletiva atinge cerca de 90%, para os respectivos números de repetições.

Tabela 3. Valores alternativos dos coeficientes de variação genotípica (CVg) e experimental (CVe), expressos em porcentagem, necessários para se conseguir acurácias em torno de 90% na avaliação genotípica, sob diferentes números de repetições (b) (h^2_i é o coeficiente de determinação genotípica em nível de parcelas, em cada caso)¹.

b = 2		b = 3		b = 4		b = 5		b = 6	
CVr=1,50	h^2_i	CVr=1,20	h^2_i	CVr=1,01	h^2_i	CVr=0,90	h^2_i	CVr=0,82	h^2_i
CVg	CVe	CVg	CVe	CVg	CVe	CVg	CVe	CVg	CVe
5,0	3,0	5,0	4,0	5,0	5,0	5,0	6,0	5,0	6,0
10,0	7,0	10,0	8,0	10,0	10,0	10,0	11,0	10,0	12,0
15,0	10,0	15,0	13,0	15,0	15,0	15,0	17,0	15,0	18,0
20,0	13,0	20,0	17,0	20,0	20,0	20,0	22,0	20,0	24,0
25,0	17,0	25,0	21,0	25,0	25,0	25,0	28,0	25,0	30,0
30,0	20,0	30,0	25,0	30,0	30,0	30,0	33,0	30,0	37,0
35,0	23,0	35,0	29,0	35,0	35,0	35,0	39,0	35,0	43,0
40,0	27,0	40,0	33,0	40,0	40,0	40,0	44,0	40,0	49,0
45,0	30,0	45,0	38,0	45,0	45,0	45,0	50,0	45,0	55,0
50,0	33,0	50,0	42,0	50,0	50,0	50,0	56,0	50,0	61,0
55,0	37,0	55,0	46,0	55,0	54,0	55,0	61,0	55,0	67,0
60,0	40,0	60,0	50,0	60,0	59,0	60,0	67,0	60,0	73,0
65,0	43,0	65,0	54,0	65,0	64,0	65,0	72,0	65,0	79,0
70,0	47,0	70,0	58,0	70,0	69,0	70,0	78,0	70,0	85,0
75,0	50,0	75,0	63,0	75,0	74,0	75,0	83,0	75,0	91,0
80,0	53,0	80,0	67,0	80,0	79,0	80,0	89,0	80,0	98,0
90,0	60,0	90,0	75,0	90,0	89,0	90,0	100,0	90,0	110,0
95,0	63,0	95,0	79,0	95,0	94,0	95,0	106,0	95,0	116,0
100,0	67,0	100,0	83,0	100,0	99,0	100,0	111,0	100,0	122,0

¹- Os valores CVg e CVe destacados em negrito são aqueles aproximadamente associados a CVe iguais a 10%, 20% e 30%, valores esses tradicionalmente usados como referência prática.

Para espécies perenes, em que várias safras são colhidas na mesma unidade experimental, os coeficientes de determinação genética a serem considerados na Tabela 3 referem-se àqueles em nível de médias das várias safras. Em plantas anuais, nos casos em que os experimentos são repetidos em vários ambientes, o mesmo valor F de referência (5,26) pode ser usado; porém, este deve ser o valor computado a partir da análise conjunta dos experimentos, isto é:

$$F = \frac{QM_{\text{Cultivares}}}{QM_{\text{Interação GE}}}$$

Nesse sentido, vale ressaltar que, nos testes de VCU instalados com duas ou três repetições e em dois ou três locais, a eficiência da avaliação genotípica depende, ainda, da magnitude da interação GE. Para caracteres com determinação genética igual a 10%, mesmo com três repetições e avaliação em quatro ambientes (locais e, ou, anos), a acurácia máxima obtida é de 76%, na ausência de interação. Esta condição, porém, é pouco provável de ser realidade, indicando que sob valores menores que 30% de determinação genética, o uso de duas ou três repetições é insuficiente para propiciar acurácias seletivas ideais, mesmo quando os experimentos são repetidos em até quatro ambientes. Já para caracteres com determinação genética superior a 40%, o uso de duas ou três repetições por experimento pode ser suficiente para atingir esse nível de acurácia, desde a interação GE seja moderada, isto é, a correlação genotípica através dos ambientes seja superior a 80% (Resende 2007).

Inferência sobre a média genotípica (VCU)

a) Situação de dados balanceados e homogeneidade de variâncias

Na abordagem aqui defendida, baseada em estimadores *shrinkage* e de mínimo EQM, as inferências sobre as médias genotípicas das cultivares (VCU) devem ser obtidas a partir de: $\hat{g} = VCU = m + (1 - 1/F)(\bar{Y}_i - m)$; com desvio padrão igual a: $SEP = (PEV)^{1/2} = (1 - 1/F)^{1/2} \hat{\sigma}_e / (b)^{1/2}$. Logo, a inferência intervalar sobre o VCU é dada por: $m + (1 - 1/F)(\bar{Y}_i - m) \pm t(1 - 1/F)^{1/2} \hat{\sigma}_e / (b)^{1/2}$, em que “t” é o quantil da distribuição de Student, para o nível de significância de interesse (α), com graus de liberdade associados à estimativa da variância do erro (σ_e^2). Verificando-se ou não a sobreposição dos intervalos de confiança de duas cultivares, pode-se

inferir se estas diferem ou não significativamente entre si, para o nível α de probabilidade.

Reitera-se que, nesta abordagem, a ênfase não está no fato de os tratamentos genéticos atenderem ou não condições tradicionalmente estabelecidas para serem assumidos como efeitos aleatórios, mas na possibilidade de se empregar estimadores de menor erro quadrático médio, mesmo admitindo-se pequeno vício (quando o número de tratamentos é muito pequeno). Desse modo, justifica-se a necessidade de correção das médias fenotípicas por meio de um coeficiente de confiabilidade dado, aproximadamente, por $(1 - 1/F)$. Essa correção, de acordo com o número de tratamentos avaliados, pode ser aplicada da seguinte maneira:

- i) para três tratamentos, James & Stein (1961) recomendaram usar como regressor da média do tratamento: $1 - [(T - 2)/(T)]/F^*$, isto é, $1 - 0,33/F^*$, em que F^* é a estatística de Snedecor centrada em zero. O regressor também é centrado em zero e não na média geral, de modo que deve multiplicar diretamente a média do tratamento e não o seu desvio em relação à média geral. Esse resultado coincide com o estimador *shrinkage* apresentado por Vinod (1976), no contexto da análise de regressão linear múltipla.
- ii) para quatro ou mais tratamentos, Efron & Morris (1977) recomendaram também um regressor centrado na média geral: $1 - [(T - 3)/(T - 1)]/F$ (valores fornecidos na Tabela 4). Este deve multiplicar diretamente os desvios das médias dos tratamentos em relação à média geral, assim como se faz no procedimento de melhor predição linear não tendenciosa (BLUP).

A transição de um modelo de efeitos fixos para um modelo de efeitos aleatórios de genótipos, em função do aumento no número de tratamentos, pode ser constatada na Tabela 4 (construída considerando-se o caso de um modelo com efeitos fixos de genótipos, porém relaxando-se a condição de não-vício na estimação dos referidos efeitos). Verifica-se que a partir de dez tratamentos praticamente existe uma equivalência entre modelo fixo com estimador do tipo *shrinkage* e modelo aleatório. Com número de tratamentos (T) maior que cinco, o estimador *shrinkage* já tende mais ao modelo aleatório do que ao modelo fixo (sem regressor ou, equivalentemente, com regressor igual à unidade), pois o escalar a ser dividido por F é igual ou maior que 0,6, valor mais

Tabela 4. Valores dos regressores dos desvios das médias fenotípicas em relação à média geral, adaptados para experimentos visando à obtenção de estimativas mais precisas de VCU sob diferentes números de tratamentos.

Número de tratamentos	Regressor ¹	Número de tratamentos	Regressor
3	1 - 0,33/F*	14	1 - 0,85/F
4	1 - 0,33/F	15	1 - 0,86/F
5	1 - 0,50/F	16	1 - 0,87/F
6	1 - 0,60/F	17	1 - 0,88/F
7	1 - 0,67/F	18	1 - 0,88/F
8	1 - 0,71/F	19	1 - 0,89/F
9	1 - 0,75/F	20	1 - 0,89/F
10	1 - 0,78/F	21	1 - 0,90/F
11	1 - 0,80/F	38	1 - 0,95/F
12	1 - 0,82/F	135	1 - 0,99/F
13	1 - 0,83/F	400	1 - 1/F

¹- F*: estatística F de Snedecor centrada em zero (este regressor deve multiplicar diretamente a média fenotípica); F: estatística F centrada na média geral.

próximo à unidade (modelo aleatório) do que de zero (modelo fixo). Dessa forma, se houver necessidade de escolha entre modelo aleatório e fixo, sugere-se adotar o modelo aleatório quando $T > 5$, e o modelo fixo apenas quando $T < 5$. Isso está de acordo com Efron & Morris (1977), que relatam que o uso de estimador *shrinkage* reduz substancialmente o risco na inferência quando o número de tratamentos é superior a cinco. Com $T = 5$, podem-se adotar quaisquer dos modelos, embora aquele de efeitos aleatórios seja sempre conservador em relação ao de efeitos fixos. Assim, o uso generalizado do modelo de efeitos aleatórios ou, equivalentemente, o uso de $(1 - 1/F)$ como fator de *shrinkage* é uma opção segura para o melhorista.

Nesse contexto, os testes tradicionais de comparação de médias só seriam recomendáveis para números de tratamentos iguais a dois, três ou quatro, pois foram derivados sob a suposição de efeitos fixos de tratamentos (Steel & Torrie 1980). Enfim, segundo o número de tratamentos a serem comparados, pode-se adotar:

$T = 2$: teste t de Student;

$T = 3$: estimador de James-Stein centrado em zero.

$4 \leq T \leq 10$: estimador de James-Stein centrado na média geral; e

$T > 10$: BLUP.

No procedimento BLUP, lidam-se com efeitos genéticos de tratamentos assumidos como aleatórios e com efeitos de macro-ambiente (blocos, locais etc.), alternativamente, como fixos ou aleatórios. Admitindo-se alguns desses efeitos ambientais como fixos, obtêm-se as melhores estimativas lineares e não viciadas (BLUE) desses efeitos, as quais são

empregadas para se obterem as predições BLUP dos efeitos aleatórios (Searle *et al.* 1992, Littell *et al.* 1996). A questão que surge é se o uso de estimadores viciados, em lugar dos estimadores BLUE, obtidos via quadrados mínimos generalizados, propiciaria um melhoramento do método BLUP. Gianola (1990) considerou exatamente isso como uma forma de aprimorar o procedimento REML/BLUP, e Weigel *et al.* (1991) trataram com mais detalhe o problema, concluindo, via simulação, que ocorre sim uma ligeira melhoria no método.

O efeito *shrinkage* sobre os desvios de ambiente, algumas vezes tratados como efeitos fixos, segue a mesma tendência mostrada na Tabela 4 (basta ler as entradas relativas ao número de tratamentos como se fossem número de blocos, por exemplo). Entretanto, do ponto de vista do melhorista e geneticista, o que é conservador inverte-se em relação ao que foi comentado para o caso dos efeitos genotípicos. A fração $(1 - 1/F)$ penaliza os efeitos ambientais, fato que não é mais conservador para ele. Assim, em caso de dúvida na consideração dos efeitos ambientais como fixos ou aleatórios, o conservador é tratá-los como fixos. Da mesma forma que para os tratamentos, se o número "b" de blocos (repetições no delineamento usual de blocos ao acaso) for menor ou igual a cinco é preferível tratar esses efeitos como fixos; com $b > 10$, pode-se optar por tratá-los como aleatórios; e com $6 \leq b \leq 10$ seria melhor usar os estimadores de James-Stein. Porém, como isso é difícil de ser implementado no contexto computacional do método REML/BLUP, a abordagem conservadora recomendaria tratar os efeitos de blocos como fixos. Em resumo, no delineamento de blocos completos, parece adequado tratar esses efeitos como fixos, quando o número deles for menor ou igual a dez, e como aleatório, quando for superior a dez.

b) Situação de dados desbalanceados e homogeneidade de variâncias

Sob desbalanceamento e homocedasticidade, existe um coeficiente de determinação genotípica em nível de médias (\hat{r}_{gg}^2) e uma acurácia para cada tratamento genético. Dessa forma, a estimação dos valores genotípicos pode ser realizada mais facilmente por meio do procedimento BLUP. Em tais casos, os testes de comparação múltipla tradicionais, aplicados sobre médias fenotípicas, são, mais uma vez, não recomendados pelo mesmo motivo relatado anteriormente.

c) Situação de heterogeneidade de variâncias

No caso de heterocedasticidade entre os tratamentos genéticos, além de \hat{r}_{gg}^2 e acurácias diferentes, cada cultivar terá um coeficiente de determinação genética (h_{ei}^2), função da variação residual dentro deste tratamento (σ_{ei}^2). Nesta situação, um procedimento BLUP que acomoda heterogeneidade de variâncias (BLUP-HET) deve ser preferível. Tal procedimento já se encontra implementado no software Selegen-REML/BLUP (Resende 2002). Para o caso balanceado, uma abordagem alternativa é seguir a recomendação já apresentada, porém, computando-se para cada cultivar o multiplicador

$$\hat{r}_{gg}^2 = \frac{(F-1)}{(F-1) + \sigma_{ei}^2 / \sigma_e^2}, \text{ em lugar de } \hat{r}_{gg}^2 = 1 - 1/F.$$

Essa idéia é similar ao que se faz no teste t (de Student) para a comparação de duas médias, sob variâncias desiguais.

Portanto, sob heterogeneidade de variâncias entre os tratamentos genéticos, os testes de comparações múltiplas novamente não são recomendáveis, tanto para o caso balanceado, quanto sob desbalanceamento. Esses testes, em geral, assumem variâncias residuais homogêneas (aproximadamente iguais) para todos os tratamentos. Tal condição, entretanto, pode não se verificar, sobretudo nos ensaios de VCU em que houver diferentes níveis de segregação genética dentro das cultivares e, também, em razão da amostragem ambiental das parcelas atribuídas a cada tratamento, particularmente sob pequeno número de repetições.

Exemplo prático

Considere-se um experimento de avaliação da produção de grãos de seis cultivares de soja, testadas no delineamento inteiramente ao acaso, com quatro repetições. Os dados originais e a análise da variância são apresentados nas Tabelas 5 e 6, respectivamente. Os resultados associados aos coeficientes de variação já definidos (CVe, CPe, CVr e CVg) e a outras estatísticas de interesse neste estudo são:

$$CVe = \frac{\sigma_e}{m} \times 100 = 14,13\%;$$

$$CPe = \frac{\sigma_e}{(b)^{1/2} m} \times 100 = \frac{CVe}{(b)^{1/2}} = 7,07\%;$$

$$CVr = CVg / CVe = [(F-1) / b]^{1/2} = 0,67 ;$$

Tabela 5. Dados de unidades de produção de grãos por parcela num experimento realizado no delineamento inteiramente ao acaso, para a comparação de seis cultivares de soja, em quatro repetições.

Cultivar	Repetição			
	1	2	3	4
1	64	59	50	63
2	53	51	55	69
3	46	48	43	35
4	56	45	45	42
5	39	59	53	53
6	46	50	65	59

Fonte: Machado *et al.* (2005).

Tabela 6. Análise da variância da produção de grãos de seis cultivares de soja, em quatro repetições (Machado *et al.* 2005).

Fontes de Variação	Graus de Liberdade	Quadrados Médios	Teste F	p valor
Cultivares	5	152,0	2,81	0,0476
Resíduo	18	54,0	-	

¹- Média geral = 52,0.

$$CVg = CVr \times CVe \times 100 = 9,52\% ;$$

$$\hat{h}_i^2 = \frac{(CVr)^2}{1 + (CVr)^2} = 0,31 \text{ (coeficiente de determinação genética entre cultivares, em nível de parcelas); e}$$

$$\hat{r}_{gg} = (1 - 1/F)^{1/2} = 0,8030 \text{ (acurácia seletiva).}$$

Com base nesses resultados, as seguintes inferências podem ser realizadas:

- A estatística F de Snedecor, equivalente a 2,81, propicia uma acurácia seletiva de cerca de 80%. Esta acurácia, embora possa ser classificada como alta (Tabela 1), não atinge um nível mínimo adequado (90%), aqui assumido como 90%, para ensaios de VCU.
- O coeficiente de variação relativa obtido ($CVr = 0,67$), para testes de VCU, é considerado baixo, pois, no caso de quatro repetições, ter-se-ia que atingir um valor igual a 1,0 para propiciar a referida meta de acurácia seletiva (Tabela 2).
- O coeficiente de variação experimental ($CVe = 14,13\%$), embora possa parecer satisfatório à luz de recomendações tradicionais, para testes de VCU com quatro repetições, pode ser considerado alto. Conforme observa-se na Tabela 3, o valor de CVe para propiciar a acurácia de 90% deveria ser aproximadamente igual ao valor do CVg, isto é, em torno de 10%.

iv) Os coeficientes de variação relativa ($CV_r = 0,67$) e o de determinação genética ($\hat{h}_i^2 = 0,31$) são indicativos de que, neste caso, para se atingir a meta de acurácia estabelecida (90%), seriam necessárias mais de seis repetições (Tabela 3), ou, mais precisamente, nove ou dez repetições (Tabela 2).

A análise da heterogeneidade de variâncias residuais entre os tratamentos pode ser feita observando-se a Tabela 7. A maior variância (74,0) foi constatada para a cultivar 6, e a menor (32,7), para a cultivar 3. A relação entre estas variâncias é de 2,26. Este valor, inferior a 3,0, geralmente conduz a não se rejeitar a hipótese de homogeneidade das variâncias, o que daria suporte para a aplicação de um teste de comparações múltiplas, considerando-se a significância ($p < 0,05$) do teste F para cultivares (Tabela 6). Entretanto, em termos de avaliação genotípica, as diferenças nas variâncias residuais são consideráveis, pois os coeficientes de determinação genética individuais variaram de 0,25 a 0,43, fato que corresponde a uma diferença relativa de 72%. Em outras palavras, a cultivar 6 teria maior nível de variação genética dentro e, ou, experimentou ambientes mais heterogêneos do que a cultivar 3. Assim, a média fenotípica da cultivar 3 apresenta um coeficiente de confiabilidade bastante superior (75%) ao da média da cultivar 6 (57%). Isso conduz a diferentes acurácias seletivas entre os genótipos e, conseqüentemente, à necessidade de diferentes predições, via BLUP-HET, para os seus valores genotípicos. Logo, uma abordagem tradicional de comparações múltiplas conduziria a inferências

menos fidedignas sobre os parâmetros de VCU. A diferença fenotípica de 12,0 unidades de produção entre as duas cultivares (6 e 3), estimada pela abordagem tradicional, pouco ultrapassa 8,0 unidades pelo método BLUP-HET, o que corresponde a uma redução bastante considerável de 33%.

Por último, deve-se destacar que, embora neste exemplo não tenha havido mudança no ordenamento das cultivares, segundo os resultados dos estimadores empregados (Tabela 7), as diferenças nas estimativas/predições e nas respectivas precisões, em favor daqueles que promovem *shrinkage* e, ou, acomodam heterocedasticidade, demonstram que tais estimadores produzem inferências mais realistas e confiáveis acerca da avaliação genotípica.

CONCLUSÕES

1. Recomendações como três repetições e coeficiente de variação experimental mínimo de 20% ou 30% são inadequadas para informar sobre a qualidade de ensaios cujo propósito é prever o valor de cultivo e uso (VCU) de tratamentos genéticos.
2. O valor do teste F (de Snedecor) para os efeitos de cultivares e o coeficiente de variação relativa ($CV_r = CV_g / CV_e$), em conjunto com o número de repetições, são mais adequados para este fim, devendo ser empregados para inferir sobre a acurácia seletiva e a precisão em ensaios de VCU.
3. Para se atingir uma meta de acurácia de 90%, deve-se obter valores de F acima de 5,0 para os efeitos de tratamentos (cultivares), na análise da

Tabela 7. Resultados da avaliação genotípica para diferentes estimadores/preditores de valor de cultivo e uso (VCU), associados à produção de grãos de seis cultivares de soja, considerando-se também a heterogeneidade de variâncias residuais entre os tratamentos genéticos.

Cultivar	Média fenotípica ($\bar{Y}_{i.}$)	Variância residual ($\hat{\sigma}_{ei}^2$)	Coef. determ. genét. (\hat{h}_{ei}^2) ¹	Média genotípica James-Stein	Média genotípica BLUP	Média genotípica BLUP-HET	Acurácia BLUP-HET
1	59,0 ± 3,7	40,67	0,376	57,51 ± 3,3	56,52 ± 2,9	56,95 ± 2,7	0,841
2	57,0 ± 3,7	66,67	0,269	55,93 ± 3,3	55,23 ± 2,9	54,98 ± 3,1	0,771
3	43,0 ± 3,7	32,67	0,429	44,92 ± 3,3	46,20 ± 2,9	45,25 ± 2,5	0,866
4	47,0 ± 3,7	38,00	0,392	48,07 ± 3,3	48,78 ± 2,9	48,40 ± 2,6	0,849
5	51,0 ± 3,7	72,00	0,254	51,21 ± 3,3	51,36 ± 2,9	51,42 ± 3,2	0,759
6	55,0 ± 3,7	74,00	0,249	54,36 ± 3,3	53,94 ± 2,9	53,71 ± 3,2	0,755
Homoced. ²	52,00	54,00	0,312	-	52,00	-	0,803

¹- Coeficiente de determinação genética em nível de parcelas (análogo ao coeficiente de herdabilidade individual no sentido amplo).

²- Valores estimados sob a suposição de homogeneidade das variâncias residuais de tratamentos.

variância, e valores de CVr entre 0,70 (com dez repetições) e 1,50 (com duas repetições). Valores de F em torno de 2,0 só permitem acurácias seletivas da ordem de 70%.

4. Os números de repetições atualmente empregados nos ensaios de VCU, entre dois e quatro, não possibilitam níveis ideais de acurácia seletiva (iguais ou superiores a 90%), a menos para caracteres de elevada determinação genética ($h^2_i > 0,60\%$). Logo, ao menos seis repetições são necessárias para caracteres de produção ($h^2_i < 40\%$).
5. Sob reduzido número de repetições, a magnitude do coeficiente de variação experimental (CVe) é inadequada para informar sobre a precisão da avaliação genotípica, mesmo se valores tão baixos como 10% forem perseguidos e a experimentação for realizada em vários locais e anos.

REFERÊNCIAS

- Amaral, A.M., J.A. Muniz & M. Souza. 1997. Avaliação do coeficiente de variação como medida da precisão na experimentação com Citrus. *Pesquisa Agropecuária Brasileira*, 32: 1221-1225.
- Bibby, J. & H. Toutenburg. 1977. Prediction and improved estimation in linear models. John Wiley and Sons, Chichester. 188 p.
- Brasil. Ministério da Agricultura e do Abastecimento. 2001. Registro Nacional de Cultivares (RNC) – Informe Técnico. Requisitos mínimos para determinação do valor de cultivo e uso, para a inscrição no RNC. p. 19. Anexo IV.
- Box, G.E.P. & G. C. Tiao. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading. 588 p.
- Casella, G. 1985. An introduction to empirical Bayes data analysis. *American Statistician*, 39: 83-83.
- Clemente, A.L. & J.A. Muniz. 2000. Estimativas de faixas de coeficiente de variação em leguminosas forrageiras para avaliação da precisão experimental. *Ciência e Agrotecnologia*, 24: 738-743.
- Clemente, A.L. & J.A. Muniz. 2002. Avaliação do coeficiente de variação em experimentos com gramíneas forrageiras. *Ciência e Agrotecnologia*, 26: 197-203.
- Costa, N.H.A.D., J.C. Seraphin, & F.J.P. Zimmermann. 2002. Nova proposta de classificação de coeficientes de variação para a cultura do arroz de terras altas. *Pesquisa Agropecuária Brasileira*, 37: 243-249.
- Duarte, J.B. 2000. Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal. Tese de Doutorado. Escola Superior de Agricultura Luiz de Queiroz / USP. Piracicaba. 293 p.
- Duarte, J.B. & R. Vencovsky. 2001. Estimação e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. *Scientia Agrícola*, 58: 109-117.
- Efron, B. 1975. Biased versus unbiased estimation. *Advances in Mathematics*, 16: 259-277.
- Efron, B. & C. Morris. 1977. Stein's paradox in statistics. *Scientific American*, 236: 119-127.
- Estefanel, V., I.A.B. Pignarato & L. Storck. 1987. Avaliação do coeficiente de variação de experimentos com algumas culturas agrícolas. p. 115-131. In *Simpósio de Estatística Aplicada à Experimentação Agrônômica*, 1. FUEL/RBRAS/ IAPAR, Londrina. Anais.
- Federer, W.T. 1996. SAS Proc GLM and Proc Mixed for recovering inter-effect information. Tech. Rep. Biometrics Unit, Cornell Univ., Ithaca, NY. 8 p. (BU-1330-M).
- Federer, W.T. 1998. Recovery of interblock, intergradient, and intervarietal information in incomplete block and lattice rectangle designed experiments. *Biometrics*, 54: 471-481.
- Federer, W.T. & R.D. Wolfinger. 1998. SAS code for recovering intereffect information in experiments with incomplete block and lattice rectangle designs. *Agronomy Journal*, 90: 545-551.
- Garcia, C.H. 1989. Tabelas para classificação do coeficiente de variação. IPEF, Piracicaba. 12 p. (Circular Técnica 171).
- Gianola, D. 1990. Can BLUP and REML be improved upon? p. 445-449. In *World Congress on Genetics Applied to Livestock Production*, 4th. Joyce Darling, Penicuik, UK. (Proceedings, vol. XIII).
- Gianola, D. & R.L. Fernando. 1986. Bayesian methods in animal breeding theory. *Journal of Animal Science*, 63: 217-244.
- Henderson, C.R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph. 462 p.
- Hill, R.R. & J.L. Rosenberger. 1985. Methods for combining data from germplasm evaluation trials. *Crop Science*, 25: 467-470.
- Hoerl, A.E. & R.W. Kennard. 1970. Ridge regression: biased estimation and applications for non-orthogonal problems. *Technometrics*, 12: 55-82.
- James, W. & C. Stein. 1961. Estimation with quadratic loss. p. 361-379. In *Symposium on Mathematical Statistics and Probability*, 4th. University of Berkeley, Berkeley, USA. (Proceedings, vol. 1).
- Judge, G.G., W.E. Griffiths, R. Carter, H. Lutkepohl & T.C. Lee. 1985. The Theory and Practice of Econometrics. John Wiley & Sons, New York. 1056 p.

- Judice, M.G., J.A. Muniz & R. Carvalheiro. 1999. Avaliação do coeficiente de variação na experimentação suínos. *Ciência e Agrotecnologia*, 23: 170-173.
- Judice, M.G., J.A. Muniz, L.H. de Aquino & E. Bearzoti. 2002. Avaliação da precisão experimental em ensaios com bovinos de corte. *Ciência e Agrotecnologia*, 26: 1035-1040.
- Littell, R.C., G.A. Milliken, W.W. Stroup & R.D. Wolfinger. 1996. SAS® system for mixed models. Statistical Analysis System Institute, Cary (USA). 633 p.
- Lúcio, A.D. 1997. Parâmetros da precisão experimental das principais culturas anuais no Rio Grande do Sul. Dissertação de Mestrado. Universidade Federal de Santa Maria. Santa Maria-RS. 62 p.
- Machado, A. de A., J.G.C. Silva, C.G.B. Demétrio & D.F. Ferreira. 2005. Estatística experimental: uma abordagem baseada no planejamento e no uso de recursos computacionais. p.188-197. In Reunião da RBRAS, 50/ Simpósio de Estatística Aplicada à Experimentação Agrônômica, 11. UEL, Londrina. 290 p.
- Piepho, H.P. 1994. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theoretical and Applied Genetics*, 89: 647-654.
- Piepho, H.P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. *Theoretical and Applied Genetics*, 97: 195-201.
- Pimentel Gomes, F. 1987. Curso de estatística experimental. 12. ed. Nobel, São Paulo. p.
- Pimentel Gomes, F. 1991. Índice de variação: um substituto vantajoso do coeficiente de variação. IPEF, Piracicaba. 4 p. (Circular técnica 178).
- Ramalho, M.A.P., D.F. Ferreira & A.C. de Oliveira. 2000. Experimentação em genética e melhoramento de plantas. UFLA, Lavras. 303 p.
- Resende, M.D.V. de. 1999. Predição de valores genéticos, componentes de variância, delineamentos de cruzamento e estrutura de populações no melhoramento florestal. Tese de Doutorado. Universidade Federal do Paraná, Curitiba. 434 p.
- Resende, M.D.V. de. 2002. Genética biométrica e estatística no melhoramento de plantas perenes. Embrapa Informação Tecnológica, Brasília. 975 p.
- Resende, M.D.V. de. 2004. Métodos estatísticos ótimos na análise de experimentos de campo. Embrapa Florestas, Colombo. 65 p. (Documentos 100).
- Resende, M.D.V. de. 2007. Matemática e estatística na análise de experimentos e no melhoramento genético. Embrapa Florestas, Colombo. 435 p.
- Resende, M.D.V. de, D.F. Prates, A. Jesus & C.K. Yamada. 1996. Estimação de componentes de variância e predição de valores genéticos pelo método da máxima verossimilhança restrita (REML) e melhor predição linear não viciada (BLUP) em Pinus. *Boletim de Pesquisa Florestal*, 32/33: 18-45.
- SAS Institute. 1997. SAS/Stat® software: changes and enhancements through release 6.12. Statistical Analysis System Institute, Cary (USA). 1167 p.
- Searle, S.R., G. Casella & C.E. McCulloch. 1992. Variance components. John Wiley & Sons, New York. 501 p.
- Scapim, C.A., C.G.P. Carvalho & C.D. Cruz. 1995. Uma proposta para classificação dos coeficientes de variação para a cultura do milho. *Pesquisa Agropecuária Brasileira*, 30: 683-686.
- Smith, A. B. Cullis & A. Gilmour. 2001. The analysis of crop variety evaluation data in Australia. *Australian New Zealand Journal of Statistics*, 43: 129-145.
- Steel, R.G.D. & J.H. Torrie. 1980. Principles and procedures of statistics. 2th. ed. Mc Graw-Hill, New York. 633 p.
- Stein, C. 1955. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. p. 197-206. In Symposium on Mathematical Statistics and Probability, 3th. University of Berkeley, Berkeley, USA. (Proceedings, vol. 1).
- Storck, L., D.C. Garcia, S.J. Lopes & V. Estefanel. 2000. Experimentação vegetal. UFSM, Santa Maria. 198p
- Stroup, W.W. & D.K. Mulitze. 1991. Nearest neighbour adjusted best linear unbiased prediction. *American Statistician*, 45: 194-200.
- Theil, H. 1971. Principles of econometrics. John Wiley & Sons, New York. p.
- Vencovsky, R. 1987. Herança quantitativa. p. 137-214. In E. Paterniani & G.P. Viegas (Ed.). Melhoramento e produção de milho. 2.ed. Fundação Cargill, Campinas. v. 1. 795 p.
- Vinod, H.D. 1976. Simulation and extension of a minimum mean squared error estimator in comparison with Stein's. *Technometrics*, 18: 491-496.
- Vinod, H.D. & A. Ullah. 1981. Recent advances in regression methods. Marcel Dekker, New York. 362 p.
- Weigel, K.A., D. Gianola, R.J. Tempelman, C.A. Matos & I.H. C. Chen. 1991. Improving estimates of fixed effects in a mixed linear model. *Journal of Dairy Science*, 74: 3174-3182.
- Wolfinger, R.D., W.T. Federer & O. Cordero-Brana. 1997. Recovering information in augmented designs, using SAS Proc GLM and Proc Mixed. *Agronomy Journal*, 89: 856-859.