PRECISION AND QUALITY CONTROL IN VARIETY TRIALS[1]

Marcos Deon Vilela de Resende[2], João Batista Duarte[3]

ABSTRACT

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 (CVe) 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

PRECISÃO E CONTROLE DE QUALIDADE EM EXPERIMENTOS DE AVALIAÇÃO DE CULTIVARES

Este estudo teve como objetivo propor nova abordagem para a avaliação da qualidade dos ensaios de avaliação do valor de cultivo e uso (VCU) de cultivares, a qual, simultaneamente, considera três atributos: magnitude da variação residual, número de repetições e controle genético dos caracteres. Enfatizou a necessidade do uso de métodos de estimação/predição de valores genotípicos que promovam "shrinkage" sobre a média fenotípica do cultivar; isto é, levem em conta o coeficiente de determinação genética e a possível heterogeneidade de variâncias residuais entre cultivares. Concluiu-se que, para atingir 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 o número mínimo de repetições e o valor máximo para o coeficiente de variação experimental (CVe). Para caracteres de produção, com baixo coeficiente de determinação genética, os números de repetições usualmente empregados (entre dois e quatro) não permitem atingir essa meta de acurácia, mesmo quando se perseguem valores de CVe inferiores a 10% e a experimentação é realizada em vários locais e anos. Para isso, ao menos seis repetições são recomendadas. Demonstra-se, ainda, que os estimadores que promovem shrinkage garantem inferências mais precisas e realistas sobre as médias genotípicas das cultivares, devendo ser encorajados.

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

INTRODUCTION

Field trials are fundamental to plant breeding programs and play an essential role in the recommendation process of improved cultivars. High levels of experimental precision are desirable in such trials to ensure accurate inference regarding genotypic means; that is, the underlying genetic values of the treatments under evaluation. In Brazil, these genotypic means are officially referred to as "value for cultivation and use" (VCU) by the Ministry of Agriculture, serving as a legal requirement for the

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: marcos.resende@embrapa.br
3. Escola de Agronomia, Setor de Melhoramento de Plantas, Universidade Federal de Goiás. Caixa Postal 131, CEP 74001-970 Goiânia-GO. Email: jbduarte@ufg.br

recommendation, registration, and protection of cultivars (Brasil 2001).

Several statistical parameters have been proposed to assess the precision and reliability of agricultural experiments. The experimental coefficient of variation (CVe) has traditionally been recommended (Pimentel Gomes 1987), and remains widely adopted in practice (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). Acceptable CVe thresholds depend on the crop species and the trait under assessment. However, the CVe captures only the residual variation in proportion to the experimental overall mean.

Another statistical metric used for a similar purpose is the variation index proposed by Pimentel Gomes (1991), also referred to as the coefficient of experimental precision (CPe) by Storck *et al.* (2000). Unlike the CVe, the CPe incorporates both the residual variation and the experimental number of replications, offering a more comprehensive indicator of experimental precision.

The cultivar performance trials must be evaluated not only from a statistical perspective, but also from a genetic standpoint. In this context, one of the most informative parameters for assessing experimental quality is the selective accuracy; although still underutilized. This statistical metric has the property of incorporating information about the correct cultivar ranking for selection and also the effectiveness of inference on the breeding value of the cultivar, that is, its VCU (Resende 2002). Unlike CVe or CPe, the selective accuracy depends not only on the magnitude of the residual variance and the number of replications, but also on the ratio between genetic and residual variances for the trait under evaluation.

The objective of this study was to propose an integrative approach for evaluating the quality of cultivar performance trials by jointly considering three key elements: magnitude of residual variation, number of replications, and the genetic control of traits. Without loss of generality, our focus is on VCU trials planned and conducted under a randomized complete block design (RCBD). We further demonstrate that breeding value estimation/prediction methods that accommodate shrinkage of phenotypic means and heteroscedasticity among cultivars ensure more accurate and, therefore, more realistic inferences about genotypic evaluation.

METHODOLOGY

The evaluation of genetic treatments in field trials serves two main purposes: *i)* to infer the breeding values of the tested materials; and *ii)* to rank them based on these genetic or genotypic values for selection purposes. Thus, the primary interest is not in estimating their phenotypic means, but rather their genotypic values, i.e., their expected future performance when cultivated again under commercial crop conditions. In this case, even if cultivation takes place in the same site or region target by the assessment, block and plot effects are unlikely repeated. Since these effects are partially embedded in phenotypic means, such means are not appropriate for making inferences about genotypic values. Therefore, in scientific publications or in genetic evaluation catalogs, reporting simple phenotypic means is not advisable. Instead, the focus should be on reporting the genotypic means (values purged of environmental effects), that is, the actual value for cultivation and use (VCU).

In the estimation or prediction of genotypic values, the most critical aspect is the choice of the estimation/prediction method. This method must allow for the most precise and realistic inference possible, which should be assessed using appropriate statistics. According to Henderson (1984), within the context of genotypic evaluation, the most important statistical parameter is selective accuracy ($\hat{r}_{\hat{g}g}$). This parameter refers to the correlation between the true genotypic value of a given treatment and its estimated or predicted value based on experimental data. As a correlation, it ranges from 0 to 1, and the desirable accuracy values are those close to unity or 100%. Consequently, high accuracy is a key objective in cultivar evaluation trials. Accuracy also increases as the absolute deviations between the true (parametric) genotypic values and their corresponding estimates or predictions decrease. Such deviations can be quantified using the mean squared error (MSE). This statistic is equivalent to the average Euclidean distance between the estimates and the true values, and is given as: $\text{MSE} = (\text{Bias})^2 + \text{PEV}$; where, in addition to the term associated with the estimation bias, PEV is the variance of the prediction error.

The mean squared error (MSE) incorporates then both the bias and the precision in the estimation, two fundamental concepts that also underlie the concept of accuracy. An accurate estimator or predictor of the genotypic value exhibits negligible or no bias and high precision (i.e., low prediction error variance). Therefore, minimizing MSE implies maximizing accuracy, and the optimal estimation or prediction method is the one that achieves this objective. It has been demonstrated that such a method may exhibit some small bias, since the key objective is to minimize the sum: $(\text{Bias})^2 + \text{PEV}$ (Henderson 1984).

In the class of unbiased estimators/predictors, MSE quantifies precision, which is defined by the prediction error variance (PEV). This statistic is related to accuracy through the following equation (Henderson 1984): $\hat{r}_{\hat{g}g} = (1 - \text{PEV}/\sigma_g^2)^{1/2}$; which results in: $\text{PEV} = \left(1 - \hat{r}_{\hat{g}g}^2\right)\sigma_g^2$, where σ_g^2 represents the genotypic variance among the genetic treatments (at this stage, still free of any assumption regarding the nature of their effects). Accordingly, lower values of PEV imply both greater accuracy and higher precision. Thus, within the class of unbiased estimators or predictors, minimizing PEV also leads to maximizing accuracy. But more generally, if the requirement of unbiasedness is relaxed, what should be minimized is the mean squared error (MSE).

The phenotypic mean or simple arithmetic mean estimated by the ordinary least squares (OLS) method, although known to be unbiased, is not a minimum MSE estimator when more than two genetic treatments are being compared. The seminal study of Stein (1955), which represented a true paradox in statistical theory, demonstrated that the arithmetic mean is an inadmissible estimator, since there are others estimators that provide a lower mean squared error or, equivalently, lower risk, when more than two means must be estimated. In this context, James & Stein (1961) proposed an improved estimator for the population mean, given by: $M^* = k(\overline{Y}_{i..} - \overline{Y}_{...}) + \overline{Y}_{...}$, where k is a regressor or shrinkage factor applied to the sample mean of a given treatment i ($\overline{Y}_{i..}$), relative to the overall mean ($\overline{Y}_{...}$).

Methods that minimize MSE, whether biased or unbiased, generally lead to shrinkage-type estimators or predictors (Efron & Morris 1977). In general, such an estimator can be expressed as the product of a scalar (ranging between 0 and 1) and a vector of means estimated either by ordinary least squares or by maximum likelihood; that is, the phenotypic means are multiplied by a shrinkage factor that depends on the reliability of those estimates. This approach is sometimes applied empirically by economists, managers and breeders who have already observed that phenotypic means obtained in trials are often not reproduced in commercial cultivations, and that the average performance in these crop fields tends to be lower than that observed in the trials. As a result, they multiply phenotypic means by empirical reliability coefficients. Although, strictly speaking, these practitioners are using shrinkage estimators, they are not doing so in an optimal or formally structured way.

In the context of genetic trials, the optimal reliability factor is a function of the genotypic determination coefficient associated with the trait under evaluation, which, in the case of intrapopulation selection, corresponds to the heritability coefficient. Thus, the appropriate way to eliminate environmental effects of error embedded in phenotypic data is by multiplying the corrected phenotypic value by a shrinkage factor. This approach has been generalized to other types of experiments under the framework of so-called inter-effect information recovery (Federer 1996, Wolfinger *et al.* 1997, Federer 1998, Federer & Wolfinger 1998).

Several studies have emphasized the need to employ shrinkage-type estimators/predictors, even when the effects under analysis are considered as fixed according to traditional approaches (Hill & Rosenberger 1985, Stroup & Mulitze 1991, Piepho 1994, Resende *et al.* 1996, Piepho 1998, Smith *et al.* 2001, Duarte 2000, Duarte & Vencovsky 2001, Resende 1999, 2002, 2004). This recommendation derived from the algebraic equivalence between such shrinkage-based methods and those based on the assumption that the effects are random, implying that they belong to a conceptual population characterized by a common mean and variance (Henderson 1984, Wolfinger *et al.* 1997). In the case of genetic treatments, this population corresponds to that represented by the sample of cultivars under evaluation, whose intergenotypic variance is denoted here by σ_g^2.

Shrinkage-type estimators are employed in the best linear unbiased prediction (BLUP) method, as described by Henderson (1984). This method treats genotypic effects as random, and the BLUP is, additionally, an unbiased predictor. According to Stein (1955), when more than two treatments are involved, shrinkage-type estimators are always necessary, regardless of whether the effects are considered fixed or random.

The improved estimator proposed by James & Stein (1961) does not even require any assumptions regarding the nature of the effects or the distributions of the means to be estimated (Efron & Morris 1977); it only requires relaxation of the unbiasedness assumption. Although this estimator is biased, it yields a lower mean squared error than the least squares estimator. In the context of genotypic tests, the bias associated with the James-Stein estimator is small and becomes relevant only when the number of treatments is low (fewer than ten). As this number increases, the bias tends to decrease, thereby justifying it being referred to as an approximately unbiased estimator. Such estimators become particularly useful when unbiased methods produce estimates that fall outside the admissible parameter space.

In summary, in the context of genotypic evaluation, the main concern is not whether the treatments are considered as fixed or random effects, but rather the choice of more accurate estimators/predictors with minimum mean squared error. It is in this context that the estimators proposed by James & Stein (1961), also known as shrinkage estimators, have gained prominence. These estimators do not assume randomness, but as the number of treatments under evaluation increases, they enable a natural transition from purely fixed effects to entirely random effects. And this only depends on the sample size (the number of treatments). With a large number of treatments (typically more than ten), these estimators become equivalent to the BLUP method (Gianola 1990, Weigel *et al.* 1991). In such cases, however, BLUP offers the advantages of being easy to implement and applicable to unbalanced data scenarios.

On the other hand, under the approach advocated here, the VCU estimators should almost always be based on random or mixed models, and with random effects for genetic treatments. This is because the use of inter-genotypic information, resulting from the corresponding assumption of randomness, represents a statistical "concession" to exploit the dependence structure among cultivars, which may be related through a shared origin to the aforementioned conceptual population (Henderson 1984). In this context, the interaction of genotypes by environments (GE) should also be treated as a random effect, thus enabling inference across the entire target population of environments.

Therefore, assuming genetic treatments as random effects, the estimator of accuracy is given by (Resende 2002):

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

where:

b: number of replications or blocks (in the cases of typical randomized complete block design trials commonly used in VCU evaluations);

$h_i^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2}$: genetic determination coefficient among cultivars (inter-populational), at level of plots, associated with the trait under evaluation (the notation h^2 reflects its correspondence to the intra-populational broad-sense heritability coefficient at the individual level);

σ_g^2 : genotypic variance among cultivars;

σ_e^2 : residual variance, or within-cultivar variance.

It can thus be noted that accuracy essentially depends on the ratio between the mean residual variance and the genotypic variance, with the mean residual variance being a function of the number of replications.

The traditionally used precision parameters are:

$CV_e = \frac{\sigma_e}{m} \times 100$: the experimental coefficient of variation, expressed as a percentage, where "m" is the overall mean of the experiment; and

$CP_e = \frac{\sigma_e}{(b)^{1/2} m} \times 100 = \frac{CV_e}{(b)^{1/2}}$: the experimental precision coefficient, also expressed as a percentage.

In terms of such traditional parameters, the accuracy defined in Equation (I) can also be expressed as:

$$\hat{r}_{\hat{g}g} = \left[\frac{1}{1+CP_e^2/CV_g^2}\right]^{\frac{1}{2}} = \left[\frac{1}{1+(CV_e^2/CV_g^2)/b}\right]^{\frac{1}{2}} = \left[1 - \frac{1}{1+bCV_r^2}\right]^{\frac{1}{2}} \quad (II)$$

where:

$CV_g = \frac{\sigma_g}{m} \times 100$: genotypic coefficient of variation;

$CV_r = \frac{CV_g}{CV_e}$: relative coefficient of variation.

Therefore, based on these parameters, the selective accuracy is shown to depend on the ratio between the experimental and genotypic coefficients of variation, as well as on the number of replications. Thus, assessing the quality of cultivar competition trials solely on the basis of the experimental coefficient of variation (CVe) or the experimental precision coefficient (CPe) proves to

be an inadequate criterion, since it does not consider the magnitude of genotypic variation expressed in the trait. Furthermore, these criteria provide no information about the level of accuracy actually being achieved in the selection of cultivars. The use of the ratio between the genotypic and residual coefficients of variation was proposed by Vencovsky (1987) as a metric to infer the potential for success in breeding programs. However, even in that proposal, the relationship was not explicitly associated with the number of replications.

According to Resende (2002), the selective accuracy can alternatively be expressed as: $\hat{r}_{\hat{g}g} = (1 - 1/F)^{1/2}$, where F (Snedecor's F-statistic) refers to the variance ratio for treatment effects (cultivars), as obtained from the analysis of variance (ANOVA). Thus, an appropriate parameter for evaluating the quality of cultivar trials can be summarized in terms of a single statistic that simultaneously accounts for the experimental coefficient of variation, the number of replications, and the genotypic coefficient of variation. The following expression illustrates how the F-value encompasses all three parameters: $F = 1 + bCV_g^2/CV_e^2$.

The prediction error variance (PEV) can also be alternatively expressed as a function of the previously considered parameters:

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

In the context of estimation/prediction using mixed models, the traditional approach that treats cultivar effects as fixed implicitly corresponds to assuming that genotypic variance tends toward infinity (SAS Institute 1997), that is, that the genotypic determination coefficient approaches 1.0. In this case, the expression for the prediction error variance (PEV) becomes: $PEV = \sigma_e^2/b$, which is equivalent to the square of the standard error of the mean (SEM^2), or, alternatively, to the variance of the treatment mean in a fixed-effects model. Another way to express this statistic is:

$$PEV = (1 - 1/F)\sigma_e^2/b = \hat{r}_{\hat{g}g}^2\sigma_e^2/b = \hat{r}_{\hat{g}g}^2 SEM^2 \quad \text{(IV)}$$

This expression also reveals that the traditional approach of treating genotypes as fixed implicitly assumes that selective accuracy approaches unity. However, this assumption does not hold true for quantitative traits. Moreover, it becomes evident that the prediction error variance (PEV) is less than or equal to the square of the standard error of the mean, as a function of the reliability – or squared accuracy ($\hat{r}_{\hat{g}g}^2$), which ranges from 0 to 1. Therefore, the use of shrinkage-type estimators, or the assumption of genotypes as random effects, represents an approach that is, in most cases, more realistic and leads to greater precision (i.e., lower PEV)

In this context, treating genotypic effects as random does not necessarily require the explicit estimation of variance components. The genotypic determination coefficient required for the estimation/prediction process can be obtained directly from the F-value of the ANOVA, which is typically computed even when genotype effects are treated as fixed. In this case, the predicted genotypic value is given by Resende (1999, 2002) as:

$\hat{g} = VCU = m + (1 - 1/F)(\bar{Y}_{i.} - m)$, for i =1, 2, …,T genetic treatments (e.g., genotypes).

This approach implicitly incorporates the main genetic foundations of genotypic evaluation and simultaneously aligns with the four main scientific frameworks for estimating or predicting treatment effects (Weigel *et al.* 1991, Resende 2002): the traditional fixed-effects approach (allowing for the use of biased estimators); the mixed-model methodology (with random treatment effects); the James-Stein shrinkage estimators; and the Bayesian approach. It is worth mentioning that, under this latter approach, treatment effects are always considered random, thus naturally leading to the use of shrinkage estimators, even when the number of treatments is small (fewer than ten).

From the perspective of classical or frequentist statistics, the sample mean is the best linear unbiased estimator (BLUE) of the population mean. The estimator presented here, which introduces a regressor applied to the sample mean of the treatment, $(1 - 1/F)$, in relation to the overall mean, is biased in this case (Gianola & Fernando 1986). However, there are biased estimators and predictors that provide a lower mean squared error than BLUE estimators (Henderson 1984), and are therefore advantageous when compared to unbiased estimators (Efron 1975).

Stein (1955) also demonstrated that the arithmetic mean is an admissible estimator, but only when one or two means are being estimated. Thus, the regressor factor k proposed by James & Stein (1961) is given by: $k = 1 - \frac{(T-3)\,\sigma^2}{\sum(\bar{Y}_{i..} - \bar{Y}_{...})^2}$, with $T \geq 3$

treatments and σ^2 corresponding to the residual variance. Noting the similarity between $(T-3)$ and the degrees of freedom for treatments $(T-1)$, it becomes evident that there is also a strong resemblance between k and $(1-1/F)$. Hence, the correction is only necessary when more than three treatments are considered. Estimators of this kind, which involve shrinkage and minimize the mean squared error (MSE) without requiring any assumption regarding fixed or random effects, have received considerable attention from researchers (Theil 1971, Bibby & Toutenburg 1977, Vinod & Ullah 1981, Casella 1985, Judge *et al.* 1985).

The Bayesian estimation procedure also minimizes the expected squared error. Therefore, the James-Stein estimator is very similar to the Bayesian estimator, and they are even identical when the number of treatments is large (Efron & Morris 1977). As such, these estimators are also referred to as Bayes-Stein estimators, empirical Bayes estimators, or the Bayes empirical rule. In Bayesian inference, there is no distinction between fixed and random effects, and the parameters to be estimated are treated as random variables (Gianola & Fernando 1986). In the case of inferences about population means of treatments, under the Bayesian approach, Box & Tiao (1973) present as the regressor factor, precisely, the quantity $(1-1/F)$.

RESULTS AND DISCUSSION

Inference on experimental precision

Based on the derived expressions, tables of results were generated to simulate experimental scenarios and, consequently, assess the quality of the corresponding trials, according to the proposed approach. Table 1 presents values of the F-statistic (Snedecor's F) for treatment effects in the analysis of variance, in relation to the level of selective accuracy achieved or required: $F = 1/(1-\hat{r}^2_{\tilde{g}g})$. The F-values that must be attained in order to ensure high-accuracy estimation/prediction of VCU are highlighted in the table. For instance, it is noted that to achieve an ideal selective accuracy of 90% or higher – corresponding to values of determination coefficient above 80%, as recommended by Steel & Torrie (1980) for reliable statistical inference – the F-values for cultivar effects must be equal to or greater than 5.26. Therefore, this threshold may serve as a reference value for VCU trials. This reference is independent of the crop species and the trait under evaluation, making it a useful benchmark for assessing experiments of this nature.

Currently, maximum values of the variation coefficient have been established for VCU trials across different crops – for example, 20% for common bean and 30% for elephant grass (Brasil 2001). However, these thresholds are relatively empirical, as they do not provide information about the selective accuracy of the evaluation and, consequently, do not consider the magnitude of genotypic variation and the number of replications. In contrast, the use of the F-statistic is an alternative that simultaneously incorporates all these aspects in the evaluation of trials. Assuming, for instance, that in plant breeding programs a selective accuracy of at least 70% should be targeted, trials with F-values for cultivars below 2.0 (Table 1) should not be considered reliable. This is because they do not even allow the attainment of this minimum threshold of selective accuracy.

Another statistic used in genotypic evaluation is the relative coefficient of variation (CVr), which is related to selective accuracy through the latter part of Equation II. When the number of replications is fixed, the magnitude of CVr can be used to infer the level of accuracy and precision in genotypic evaluation (Table 2). This table highlights the CVr values required to achieve selective accuracy near 90% for different numbers

Table 1. Appropriate values of Snedecor's F-test for treatment (cultivar) effects in the analysis of variance, aiming to achieve specific levels of selective accuracy, and the corresponding categories of required precision in genotypic evaluation.

Accuracy	Precision class	F-value
0.99	Very high	50.25
0.975	Very high	20.25
0.95	Very high	10.26
0.90	**Very high**	**5.26**
0.85	High	3.60
0.80	High	2.78
0.75	High	2.29
0.70	**High**	**1.96**
0.65	Moderate	1.73
0.60	Moderate	1.56
0.55	Moderate	1.43
0.50	Moderate	1.33
0.40	Low	1.19
0.30	Low	1.10
0.20	Low	1.04
0.10	Low	1.00

of replication. These values vary, for example, from 1.50 in the case of two replications to 0.70 for ten replications. With thirty or forty replications, it becomes possible to reach this accuracy target even when the CVg/CVe ratio is lower than 0.40. Therefore, whether CVr values are considered adequate or not must be interpreted in conjunction with the number of replications.

Vencovsky (1987) reported that CVr values around unity are adequate for experimentation with maize. Indeed, results in Table 2 confirm that this value ensures high to very high levels of accuracy. However, with a number of replications greater than five, CVr values below unity can also provide high accuracy and precision. Such information can also be used in reverse, that is, based on the observed CVr value in VCU trials, it is possible to infer the appropriate number of replications to achieve a desired level of selective accuracy.

Still based on the accuracy expressions presented in Equation II, alternative values of the genotypic and experimental coefficients of variation were simulated, required to achieve approximately 90% accuracy in genotypic evaluation (Table 3). The results show that, with the number of replications commonly employed (between two and four), it is not possible to reach the desired levels of accuracy for most traits of interest in these evaluations. This is because, with a low number of replications, an accuracy of 90% could only be achieved for traits with a high coefficient of genetic determination ($h_i^2 \geq 0.60$), which is unlikely for quantitative traits. In such cases, the magnitudes of the experimental coefficient of variation are inadequate to reflect precision, even when low values such as 10% are targeted. Thus, for yield traits with $h_i^2 \leq 0.40$, at least six replications would be required. In this case, several CVe values are acceptable, provided that they maintain a ratio close to 0.82 with CVg (see last column of Table 3).

In summary, some currently widespread experimental guidelines (e.g., three replications and a maximum CVe of 20%) for VCU testing are not adequate to achieve selective accuracy goals of 90% or higher. This inadequacy applies both to the number of replications and the experimental coefficient of variation. The statistic that most effectively and comprehensively captures the key information regarding experimental quality is in fact the F-test value for treatment effects, which in this case must be greater than 5.0 (Table 1). This approach can be employed for any trait and crop species, although it should be noted that, for yield traits, at least six replications are needed to attain such an F value.

Table 2. Selective accuracy values according to different relative coefficients of variation (CVr) and numbers of replications (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]- The CVr values highlighted in bold correspond to selective accuracy levels of approximately 90% for the respective numbers of replications.

Table 3. Alternative percentage values of genotypic (CVg) and experimental (CVe) coefficients of variation required to reach around 90% accuracy in genotypic evaluation for different numbers of replications (b) (h^2_i is the genotypic determination coefficient at the plot level)[1].

b = 2			b = 3			b = 4			b = 5			b = 6		
CVr ≅ 1.5		h^2_i	CVr ≅ 1.2		h^2_i	CVr ≅ 1.0		h^2_i	CVr ≅ 0.9		h^2_i	CVr ≅ 0.8		h^2_i
CVg	CVe		CVg	CVe		CVg	CVe		CVg	CVe		CVg	CVe	
5.0	3.0	0.74	5.0	4.0	0.61	5.0	5.0	0.50	5.0	6.0	0.41	5.0	6.0	0.41
10.0	7.0	0.67	10.0	8.0	0.61	**10.0**	**10.0**	0.50	**10.0**	**11.0**	0.45	**10.0**	**12.0**	0.41
15.0	**10.0**	0.69	**15.0**	**13.0**	0.57	15.0	15.0	0.50	15.0	17.0	0.44	**15.0**	**18.0**	0.41
20.0	13.0	0.70	20.0	17.0	0.58	**20.0**	**20.0**	0.50	**20.0**	**22.0**	0.45	20.0	24.0	0.41
25.0	17.0	0.68	**25.0**	**21.0**	0.59	25.0	25.0	0.50	25.0	28.0	0.44	**25.0**	**30.0**	0.41
30.0	**20.0**	0.69	30.0	25.0	0.59	**30.0**	**30.0**	0.50	**30.0**	**33.0**	0.45	30.0	37.0	0.40
35.0	23.0	0.70	35.0	29.0	0.59	35.0	35.0	0.50	35.0	39.0	0.45	35.0	43.0	0.40
40.0	27.0	0.69	**40.0**	**33.0**	0.60	40.0	40.0	0.50	40.0	44.0	0.45	40.0	49.0	0.40
45.0	**30.0**	0.69	45.0	38.0	0.58	45.0	45.0	0.50	45.0	50.0	0.45	45.0	55.0	0.40
50.0	33.0	0.70	50.0	42.0	0.59	50.0	50.0	0.50	50.0	56.0	0.44	50.0	61.0	0.40
55.0	37.0	0.69	55.0	46.0	0.59	55.0	54.0	0.51	55.0	61.0	0.45	55.0	67.0	0.40
60.0	40.0	0.69	60.0	50.0	0.59	60.0	59.0	0.51	60.0	67.0	0.45	60.0	73.0	0.40
65.0	43.0	0.70	65.0	54.0	0.59	65.0	64.0	0.51	65.0	72.0	0.45	65.0	79.0	0.40
70.0	47.0	0.69	70.0	58.0	0.59	70.0	69.0	0.51	70.0	78.0	0.45	70.0	85.0	0.40
75.0	50.0	0.69	75.0	63.0	0.59	75.0	74.0	0.51	75.0	83.0	0.45	75.0	91.0	0.40
80.0	53.0	0.69	80.0	67.0	0.59	80.0	79.0	0.51	80.0	89.0	0.45	80.0	98.0	0.40
90.0	60.0	0.69	90.0	75.0	0.59	90.0	89.0	0.51	90.0	100.0	0.45	90.0	110.0	0.40
95.0	63.0	0.69	95.0	79.0	0.59	95.0	94.0	0.51	95.0	106.0	0.45	95.0	116.0	0.40
100.0	67.0	0.69	100.0	83.0	0.59	10.00	99.0	0.51	100.0	111.0	0.45	100.0	122.0	0.40

[1]- The CVg and CVe values highlighted in bold are those approximately associated with CVe values of 10%, 20%, and 30%, which are traditionally used as practical reference thresholds.

For perennial species, in which multiple harvests are conducted on the same experimental unit, the coefficients of genetic determination to be considered in Table 3 refer to those calculated at the level of means for the multiple harvests. In annual crops, when trials are repeated across multiple environments, the same reference F-value (5.26) can be used. However, it must be computed from the joint analysis of variance for the trials, that is: $F = MS_G/MS_{GE}$; with the mean squares (MS) for the effects of cultivars or genotypes (G) and genotype-by-environment interaction (GE).

In this context, it is worth noting that in VCU trials conducted with two or three replications and across two or three locations, the efficiency of genotypic evaluation also depends on the magnitude of the GE interaction. For traits with a genetic determination of 10%, even with three replications and evaluation in four environments (locations and/or years), the maximum accuracy obtained is 76%, assuming no GE interaction. However, such a condition is unlikely, indicating that for traits with genetic determination below 30%, using only two or three replications is insufficient to achieve optimal levels of selective accuracy, even when trials are conducted across up to four environments. However, for traits with genetic determination greater than 40%, using two or three replications per experiment may suffice to reach this level of accuracy, as long as the GE interaction remains moderate – that is, with genotypic correlation across environments above 80% (Resende 2007).

Inference about the genotypic mean (VCU)

a) Conditions of balanced data and homoscedasticity

In the approach proposed here, based on shrinkage estimators and with minimum mean squared error (MSE), the inferences on the genotypic means of cultivars (VCU) should be obtained from: $\hat{g} = VCU = m + (1 - 1/F)(\overline{Y}_{i.} - m)$; with standard error equal to: $SEP = (PEV)^{1/2}$ or $SEP = (1 - 1/F)^{1/2}\hat{\sigma}_e/(b)^{1/2}$. Thus, the confidence interval for inference about the VCU is given by: $m + (1 - 1/F)(\overline{Y}_{i.} - m) \pm t(1 - 1/F)^{1/2}\hat{\sigma}_e/(b)^{1/2}$; where "t" is the quantile of Student's t-distribution, for the desired significance level (α), with degrees of freedom associated to the estimate of the error variance (σ_e^2). By examining whether the confidence intervals of two cultivars overlap or not, it is possible to infer whether or not they differ significantly from each other, at the specified α probability level.

It is reiterated that, in this approach, the emphasis is not on whether the genetic treatments formally satisfy the traditional assumptions required to be considered random effects, but rather on the possibility of employing estimators with lower mean squared error, even assuming a small bias (when the number of treatments is very small). In this sense, the need to correct the phenotypic means is justified by using a reliability coefficient approximately given by $(1 - 1/F)$. This correction, depending on the number of treatments evaluated, can be applied as follows:

i) for three treatments, James & Stein (1961) recommended using the following regressor for the treatment mean: $1 - [(T - 2)/\ T]/F^*$, that is, $1 - 0.33/F^*$, where F^* is the Snedecor statistic centered at zero. This regressor is also centered at zero rather than at the overall mean and, therefore, it should directly multiply the treatment mean itself and not its deviation from the overall mean. This result coincides with the shrinkage estimator presented by Vinod (1976), in the context of multiple linear regression analysis.

ii) for four or more treatments, Efron & Morris (1977) also recommended a regressor centered at the overall mean: $1 - [(T - 3)/(T - 1)]/F$ (with values provided in Table 4). This regressor should directly multiply the deviations of the treatment means from the overall mean, in the same manner as is done for the Best Linear Unbiased Prediction (BLUP) procedure.

The transition from a fixed-effects model to a random-effects model for genotypes, as a function of the increasing number of treatments, can be seen in Table 4 (constructed considering the case of a model with fixed genotype effects, but relaxing the condition of unbiased estimation for such effects). It can be seen that, starting at ten treatments, there is virtually an equivalence between the fixed model using a shrinkage-type estimator and the random-effects model. When the number of treatments (T) exceeds five, the shrinkage estimator already aligns more closely with the random-effects model than with the fixed-effects model (without a regressor, or equivalently, with a regressor equal to unity), since

Table 4. Regressor values for the deviations of phenotypic means relative to the overall mean, adapted for variety trials targeting more accurate estimates of value for cultivation and use (VCU) under different numbers of treatments.

Number of treatments	Regressor[1]	Number of treatments	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*: Snedecor's F-statistic centered at zero (this regressor should be directly applied to the phenotypic mean of treatment); F: F-statistic centered at the overall mean.

the scalar to be divided by F is equal to or greater than 0.6; a value closer to one (random model) than to zero (fixed model). Therefore, if a choice between random and fixed models is required, it is advisable to adopt the random-effects model when $T > 5$, and to use the fixed-effects model only when $T < 5$. This proposal is consistent with Efron & Morris (1977), who reported that the use of shrinkage estimators substantially reduces the inferential risk when the number of treatments exceeds five. When $T = 5$, either model may be adopted, although the random-effects model is always more conservative than the fixed-effects alternative. Thus, the widespread use of the random-effects model, or, equivalently, the use of $(1 - 1/F)$ as the shrinkage factor, is a safe and robust choice for plant breeders.

In this context, traditional mean comparison tests are only recommended when the number of treatments is two, three, or four, as they were derived under the assumption of fixed effects for treatments (Steel & Torrie 1980). Accordingly, depending on the number of treatments being compared, the following may be adopted:

$T = 2$: Student's t-test;

$T = 3$: James-Stein estimator centered at zero;

$4 \leq T \leq 10$: James-Stein estimator centered on the overall mean; and

$T > 10$: BLUP.

In the BLUP procedure, genetic effects of treatments are modeled as random and macro-environmental effects (blocks, locations, etc.) may be treated as either fixed or random. When some of these environmental effects are assumed to be fixed, the best linear unbiased estimates (BLUE) of these effects are obtained, and these are used to calculate the BLUP predictions of the random effects (Searle *et al*. 1992, Littell *et al*. 1996). A question that arises is whether using biased estimators, instead of the BLUE estimators obtained from generalized least squares, could improve the BLUP method. Gianola (1990) considered precisely this as a way to enhance the REML/BLUP procedure, and Weigel *et al*. (1991) addressed the issue in greater detail, concluding through simulation that a slight improvement in the method does indeed occur.

The shrinkage effect on environmental deviations, which are sometimes treated as fixed effects, follows the same pattern shown in Table 4 (one can read the entries for the number of treatments as if they referred to the number of blocks, for instance). However, from the breeder's or geneticist's standpoint, what is conservative becomes inverted in relation to what was discussed for genotypic effects. The fraction $(1 - 1/F)$ penalizes environmental effects, which is no longer conservative from their perspective. Thus, in case of doubt about whether to treat environmental effects as fixed or random, the conservative choice is to treat them as fixed. Similarly to what was proposed for treatments, if the number "b" of blocks (replications in the usual randomized complete block design) is less than or equal to five, it is preferable to treat these effects as fixed; with $b > 10$, such effects may be treated as random; and with $6 \leq b \leq 10$, it would be better to use James-Stein estimators. However, since this is difficult to implement within the computational context of the REML/BLUP method, the conservative approach would recommend treating block effects as fixed. In summary, in randomized complete block designs, it seems appropriate to treat these effects as fixed when their number is less than or equal to ten, and as random when it exceeds ten.

b) Condition of unbalanced data and homoscedasticity

Under conditions of unbalanced data and variance homogeneity, there is a genotypic determination coefficient at the mean level ($\hat{r}^2_{\hat{g}g}$) and a selective accuracy for each genetic treatment. Hence, the estimation of genotypic values can be more easily performed using the BLUP procedure.

In such cases, traditional multiple comparison tests applied to phenotypic means are, once again, not recommended for the same reasons previously mentioned.

c) Conditions of heteroscedasticity

In the presence of variance heterogeneity among genetic treatments, in addition to different values of $\hat{r}^2_{\hat{g}g}$ and selective accuracies, each cultivar will also have its own genetic determination coefficient (h^2_{ei}), which is a function of the residual variance within this treatment (σ^2_{ei}). In this situation, a BLUP procedure accommodating heterogeneity of variances (BLUP-HET) should be preferred. This approach is already implemented in the Selegen-REML/BLUP software (Resende 2002). For the balanced case, an alternative approach is to follow the previously presented recommendation, but computing, for each cultivar, the multiplier: $\hat{r}^2_{\hat{g}g} = \frac{(F-1)}{(F-1)+\sigma^2_{ei}/\sigma^2_e}$, instead of $\hat{r}^2_{\hat{g}g} = 1 - 1/F$. This concept is similar to what is done in the Student's t-test when comparing two means under unequal variances.

Therefore, under heterogeneity of variances among genetic treatments, multiple comparison tests are once again not recommended; whether for cases of balanced or unbalanced data. These tests generally assume homogeneous (approximately equal) residual variances across all treatments. However, this condition may not hold, especially in VCU trials where there are different levels of genetic segregation within cultivars, and also due to the environmental sampling of the plots assigned to each treatment, particularly when the number of replications is small.

Practical example

Consider an experiment evaluating grain yield of six soybean cultivars, carried out in a completely randomized design with four replications. The raw data and the corresponding analysis of variance are presented in Tables 5 and 6. The results related to the coefficients of variation already defined (CVe, CPe, CVr, and CVg), as well as other relevant statistics in this study, are as follows:

$CV_e = \frac{\hat{\sigma}_e}{m} \times 100 = 14.13\%$;

Table 5. Data on grain yield units per plot from a completely randomized variety trial comparing six soybean cultivars with four replications.

Cultivar	Replications			
	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

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

Table 6. Analysis of variance for grain yield units of six soybean cultivars[1] assessed across four replications (Machado *et al.* 2005).

Source of variation	Degree of freedom (df)	Mean square (MS)	F-test	p-value
Cultivars	5	152.0	2.81	0.0476
Residual	18	54.0	-	

[1]- Overall mean = 52.0 units.

$CP_e = \frac{\hat{\sigma}_e}{m\,(b)^{1/2}} \times 100 = \frac{CV_e}{(b)^{1/2}} = 7.07\%$;

$CV_r = CV_g/CV_e = [(F-1)/b]^{1/2} = 0.67$;

$CV_g = CV_r \times CV_e \times 100 = 9.52\%$;

$h^2_i = \frac{(CVr)^2}{1+(CVr)^2} = 0.31$ (genetic determination coefficient among cultivars, at level of plots); and

$\hat{r}_{\hat{g}g} = (1 - 1/F)^{1/2} = 0.8030$ (selective accuracy).

Based on these results, the following inferences can be made:

i) The Snedecor F-statistic, equal to 2.81, provides a selective accuracy of approximately 80%. Although this level of accuracy can be classified as high (Table 1), it does not reach the minimum threshold (here assumed as 90%) considered adequate for VCU trials.

ii) The obtained relative coefficient of variation (CVr = 0.67) is considered low for VCU trials, since under four replications it would be necessary to reach a value of 1.0 to achieve the aforementioned target of selective accuracy (Table 2).

iii) The experimental coefficient of variation (CVe = 14.13%), although seemingly satisfactory under traditional recommendations, is relatively high for VCU trials with four replications. As shown in Table 3, the CVe value to provide 90% accuracy should be approximately equal to the CVg value, that is, around 10%.

iv) The values of the relative coefficient of variation (CVr = 0.67) and the genetic determination coefficient (h_i^2 = 0.31) indicate that, in this case, more than six replications would be necessary to reach the desired level of accuracy (90%); or, more precisely, nine or ten replications (Table 2).

The analysis of the heterogeneity of residual variance among treatments can be examined in Table 7. The highest variance (74.0) was observed for cultivar 6, and the lowest (32.7) for cultivar 3. The ratio between these variances is 2.26. This value, being lower than 3.0, generally leads to non-rejection of the hypothesis of homoscedasticity, which would support the application of a multiple comparison test, given the significance ($p < 0.05$) of the F-test for cultivars (Table 6).

However, from the perspective of genotypic evaluation, the residual variance differences are substantial, as the individual genetic determination coefficients ranged from 0.25 to 0.43 – corresponding to a relative difference of 72%. In other words, cultivar 6 exhibits a higher level of within-genotype variation and/or experienced more heterogeneous environmental conditions than cultivar 3. Thus, the phenotypic mean of cultivar 3 presents a considerably higher reliability coefficient (75%) compared to that of cultivar 6 (57%). This leads to different selective accuracies among genotypes and, consequently, to the need for distinct predictions of genotypic values using the BLUP-HET procedure. Therefore, a traditional multiple comparison approach would lead to less reliable inferences about VCU parameters. The phenotypic difference of 12.0 grain yield units between cultivars 6 and 3, as estimated by the traditional approach, is reduced to just over 8.0 units using the BLUP-HET method, which represents a substantial reduction of 33%.

Lastly, it should be emphasized that, although no change in the ranking of cultivars was observed in this example, according to the results of the estimators used (Table 7), the differences in the estimates or predictions and in the respective precisions – favoring those that apply shrinkage and/or accommodate heteroscedasticity – demonstrate that such estimators provide more realistic and reliable inferences about the genotypic evaluation.

CONCLUSIONS

1. Recommendations such as using three replications and setting a minimum experimental coefficient of variation (CVe) of 20% or 30% are inadequate for assessing the quality of trials whose purpose is to predict the value for cultivation and use (VCU) of genetic treatments.
2. The F-test value (Snedecor) for cultivar effects and the relative coefficient of variation (CVr = CVg / CVe), considered jointly with the number of replications, are more appropriate indicators for this purpose and should be used to infer the selective accuracy and precision in VCU trials.
3. To achieve a target accuracy of 90%, F values greater than 5.0 for treatment (cultivar) effects in the analysis of variance, along with CVr values ranging from 0.70 (with ten replications) to 1.50

Table 7. Genotypic evaluation results based on different estimators/predictors of "value for cultivation and use" (VCU) for grain yield of six soybean cultivars, also accounting for residual variance heterogeneity among genetic treatments.

Cultivar	Phenotypic mean ($\overline{Y}_i$)	Residual variance ($\hat{\sigma}^2_{ei}$)	Genetic determination (h^2_{ei})[1]	James-Stein genotypic mean	BLUP genotypic mean	BLUP-HET genotypic mean	BLUP-HET accuracy
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
Homoscedasticity[2]	52.00	54.00	0.312	-	52.00	-	0.803

[1]- Genetic determination coefficient at the plot level (analogous to the individual broad-sense heritability coefficient).

[2]- Values estimated under the assumption of homogeneity of residual variances among treatments.

(with two replications), should be obtained. Values of F around 2.0 only allow for selective accuracies on the order of 70%.

4. The number of replications currently employed in VCU trials (typically between two and four) does not allow for ideal levels of selective accuracy (i.e., 90% or higher), except for traits with high genetic determination ($h_i^2 > 60\%$). Therefore, at least six replications are necessary for yield-related traits ($h_i^2 < 40\%$).

5. With a low number of replications, the magnitude of the experimental coefficient of variation (CVe) is inadequate to inform on the precision of genotypic evaluation, even when low values such as 10% are targeted and trials are conducted across multiple locations and years.

REFERENCES

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 Agricola, 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 Agronômica, 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 Agronômica, 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. 434p.

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 Institue. 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, v. 1).

Storck, L., D.C. Garcia, S.J. Lopes & V. Estefanel. 2000. Experimentação vegetal. UFSM, Santa Maria. 198 p.

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.