Repeatability of traits evaluated in a split-plot or factorial experiment

Claudio Guilherme Portela de Carvalho[1] and Cosme Damião Cruz*[2]

[1] Embrapa Soja, Rod. Carlos João Strass, Caixa Postal 231, CEP 86001-970, Londrina, PR, Brazil; [2] Universidade Federal de Viçosa, Departamento de Biologia Geral, CEP 36571-000, Viçosa, MG, Brazil. (* Corresponding Author. E-mail: cdcruz.mail.ufv.br)

ABSTRACT

In this paper repeatability expressions are derived and their respective ANOVA estimators obtained by using split-plot and factorial models, both in a randomized complete block design. The paper also considers different fixed and random effect models and their assumptions and restrictions. Repeatability estimates, such as the correlation between successive measurements (over time) of the same genotype, always have the same value regardless of the model used, and this allows repeatability to be calculated using models based on the mean of the experimental units (mean of blocks) of each genotype in each time. This independence was not observed for repeatability estimates at the upper limit of broad-sense heritability based on the mean of successive measurements (over time) of the same genotype. The repeatability of traits evaluated in experimental trials of different designs is also discussed.

KEY WORDS: Experimental designs, factorial model, repeatability, split-plot model.

INTRODUCTION

When successive measurements of a trait are made on a group of individuals, the initial superiority or inferiority of each individual in relation to its peers is normally maintained in later measurements. This consistency of the relative positions of subjects in relation to each other during successive measurements is what is known as *repeatability* (Turner & Young, 1969). From a statistical point of view, this repeatability is the correlation between successive measurements made on a single individual submitted to repeated evaluations over time (Lush, 1945). Repeatability represents the proportion of the total phenotypic variance of a trait that is due to differences between individuals (Chapman, 1985). Such differences can be caused by genotypic variation and permanent changes in the common environment (Falconer, 1989). Estimates of repeatability allow the estimation of the number of measurements that need to be made on each individual in order to obtain selection with a specific degree of precision and the minimum of work. The estimated repeatability value defines the upper limit of broad-sense heritability of a trait at the individual level (Lush, 1945).

In experimental trials, the value of a specific genotype under selection is often inferred from the mean of the total experimental units having this genotype, not on the bases of measurements taken on a single individual. When each individual in each experimental unit is evaluated over time, repeatability can be thought of as the correlation between successive measurements of the same genotype; in this case successive measurements refers to the means of the experimental units obtained during the successive evaluations. In such a situation repeatability reflects the consistency of the relative position of the genotypes during successive measurements and has been considered to determine the number of measurements that should be made on each genotype in order to select precisely and with a minimum of work (Dias & Kageyama, 1998). Repeatability can also be thought of as the upper limit of broad-sense heritability of a trait based on the mean of successive measurements (over time) of the same genotype (Nyquist, 1991; Jahufer *et al.*, 1994). Thus repeatability can refer to both the constancy of measurements and the upper limit of broad-sense heritability, although the calculated value of repeatability is not the same for these two situations. Since repeatability is expressed by variance components, the repeatability value can be a function of the type of statistical model adopted, and may vary according to the different fixed and random effect models used and the assumptions and restrictions pertaining to the model.

In this paper we define the parametric values of repeatability as the correlation between successive

measurements (over time) of the same genotype and as the upper limit of broad sense heritability of a trait as represented by the mean of successive measurements (over time) of the same genotype. Split-plot and factorial models and the different fixed and random effect models and their assumptions and restrictions are used to derive suitable equations.

MATERIAL AND METHODS

Parametric repeatability value

In experimental trials in which successive evaluations are made over time for each individual of each experimental unit, various statistical models can be used to describe the trait measured in the i^{th} genotype at the k^{th} time.

In the present study, we considered evaluations made over several successive years in a trial with a randomized complete block design. The split-plot and factorial models (considering their different fixed and random effects, assumptions and restrictions) were used for the derivations of repeatabilities as a correlation between successive measurements (over time) of the same genotype (ρ_1), and as the upper limit of broad-sense heritability of a trait based on the mean of successive measurements (over time) of the same genotype (ρ_2).

A split-plot model in a randomized complete block experimental design can be described as follows:

$$y_{ijk} = \mu + g_i + b_j + \gamma_{ij} + a_k + ga_{ik} + \varepsilon_{ijk}$$

where y_{ijk} = the observation concerning the i^{th} genotype (i=1,...,h) in the j^{th} experimental block (j=1,...,r) and in the k^{th} year (k=1,...,n); μ = a constant inherent in all observations; g_i = the effect of the i^{th} genotype under the influence of the permanent environment, $NID(0,\sigma_g^2)$; b_j = the effect of the j^{th} experimental block, $\sum_{j=1}^{r} b_j = 0$; γ_{ij} = the error (error *a*) associated with the i^{th} genotype in the j^{th} experimental block, $NID(0,\sigma_\gamma^2)$; a_k = the effect of the k^{th} year; ga_{ik} = the effect of the interaction of the i^{th} genotype with the k^{th} year; and ε_{ijk} = the error (error *b*) associated with the observation of the i^{th} genotype in the j^{th} experimental block and in the k^{th} year.

The repeatability expressions ρ_1 and ρ_2 can be obtained from the following expressions:

$$\rho_1 = \frac{\text{cov}(\bar{y}_{i.k}, \bar{y}_{i.k'})}{\sqrt{V(\bar{y}_{i.k})V(\bar{y}_{i.k'})}}$$

and $$\rho_2 = \frac{\sigma_g^2}{V(\bar{y}_{i..})} = \frac{\sigma_g^2}{\frac{E(MSG)}{nr}}$$

where $\bar{y}_{i.k}$ and $\bar{y}_{i.k'}$ are the measurements (each measurement refers to the mean of blocks) made on the same genotype in the k^{th} and k'^{th} year of evaluation, $\bar{y}_{i..}$ is the mean of the measurements made on the same genotype, and E(MSG) is the expected mean square for the genotypes.

Based on the split-plot model and on the expected mean squares of the genotypes described in Table 1 when $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$ and in Steel & Torrie (1980) when $\varepsilon_{ijk} \sim NID(0,\sigma_\varepsilon^2)$, we can see that ρ_1 e ρ_2 can be expressed in different ways, depending on whether the a_k effect is random or fixed.

If the a_k effect is random and $a_k \sim NID(0,\sigma_a^2)$, $ga_{ik} \sim NID(0,\sigma_{ga}^2)$ and $\varepsilon_{ijk} \sim NID(0,\sigma_\varepsilon^2)$ are assumptions, then:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r}}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_{ga}^2}{n} + \frac{\sigma_\varepsilon^2}{nr}} \quad \text{(equation a1)}$$

If we consider that $a_k \sim NID(0,\sigma_a^2)$, $ga_{ik} \sim NID(0,\sigma_{ga}^2)$, $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$ and $\text{cov}(\varepsilon_{ijk},\varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$ (where the parameter θ is the correlation between the measurement errors of the same genotype in the same block) are assumptions and the remaining covariances between errors are equal to zero then:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\theta\sigma_\varepsilon^2}{r}}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_{ga}^2}{n} + \frac{\sigma_\varepsilon^2}{nr}\left[1+(n-1)\theta\right]}$$

(equation a2)

If the a_k effect is fixed and $ga_{ik} \sim NID(0,\sigma_{ga}^2)$ and $\varepsilon_{ijk} \sim NID(0,\sigma_\varepsilon^2)$ are assumptions, then:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r}}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_{ga}^2}{n} + \frac{\sigma_\varepsilon^2}{nr}} \quad \text{(equation b1)}$$

Considering $ga_{ij} \sim NID(0,\sigma_{ga}^2)$, $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$, $cov(\varepsilon_{ijk},\varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$ as assumptions and the remaining covariances between errors as equal to zero, we have:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\theta\sigma_\varepsilon^2}{r}}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_{ga}^2}{n} + \frac{\sigma_\varepsilon^2}{nr}\left[1+(n-1)\theta\right]}$$

(equation b2)

We can consider that $\sum_{k=1}^{n} ga_{ik} = 0$ for all values of i and $\sum_{k=1}^{n} a_k = 0$ are restrictions, while $ga_{ik} \sim N(0,\frac{n-1}{n}\sigma_{ga}^2)$; $cov(ga_{ik},ga_{i'k}) = 0$ for all values of k, i and i' ($i \neq i'$) and $\varepsilon_{ijk} \sim NID(0,\sigma_\varepsilon^2)$ are assumptions. By adopting $\sum_{k=1}^{n} ga_{ik} = 0$ for all values of i and $ga_{ik} \sim N(0,\frac{n-1}{n}\sigma_{ga}^2)$ we can consider the $cov(ga_{ik},ga_{ik'}) = -\frac{1}{n}\sigma_{ga}^2$ and constant for all values of i, k, k' ($k \neq k'$). Since the $cov(ga_{ik},ga_{ik'}) = -\frac{1}{n}\sigma_{ga}^2$ we have:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} - \frac{1}{n}\sigma_{ga}^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{n-1}{n}\sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_\varepsilon^2}{nr}} \quad \text{(equation b3)}$$

Considering that $\sum_{k=1}^{n} ga_{ik} = 0$ for all values of i and $\sum_{k=1}^{n} a_k = 0$ are restrictions, while $ga_{ik} \sim N(0,\frac{n-1}{n}\sigma_{ga}^2)$; $cov(ga_{ik},ga_{i'k}) = 0$ for all values of k, i and i' ($i \neq i'$); $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$; and $cov(\varepsilon_{ijk},\varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$ are assumptions

Table 1. Expected mean squares based on a split-plot model with a randomized complete block design with b_j fixed, g_i random, $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$, $cov(e_{ijk},e_{ijk'}) = \theta\sigma_\varepsilon^2$ and the remaining covariances between errors equal to zero.

Sources of variation	E(MS)[1/]	E(MS)[2/]	E(MS)[3/]
Blocks (B)	$\sigma_\varepsilon^2[1+(n-1)\theta]+n\sigma_\gamma^2+\frac{hn}{r-1}\sum_{j=1}^{r}b_j^2$	$\sigma_\varepsilon^2[1+(n-1)\theta]+n\sigma_\gamma^2+\frac{hn}{r-1}\sum_{j=1}^{r}b_j^2$	$\sigma_\varepsilon^2[1+(n-1)\theta]+n\sigma_\gamma^2+\frac{hn}{r-1}\sum_{j=1}^{r}b_j^2$
Genotypes (G)	$\sigma_\varepsilon^2[1+(n-1)\theta]+n\sigma_\gamma^2+r\sigma_{ga}^2+nr\sigma_g^2$	$\sigma_\varepsilon^2[1+(n-1)\theta]+n\sigma_\gamma^2+r\sigma_{ga}^2+nr\sigma_g^2$	$\sigma_\varepsilon^2[1+(n-1)\theta]+n\sigma_\gamma^2+nr\sigma_g^2$
Error (a)	$\sigma_\varepsilon^2[1+(n-1)\theta]+n\sigma_\gamma^2$	$\sigma_\varepsilon^2[1-(n-1)\theta]+n\sigma_\gamma^2$	$\sigma_\varepsilon^2[1-(n-1)\theta]+n\sigma_\gamma^2$
Years (A)	$\sigma_\varepsilon^2(1-\theta)+r\sigma_{ga}^2+hr\sigma_a^2$	$\sigma_\varepsilon^2(1-\theta)+r\sigma_{ga}^2+\frac{hr}{n-1}\sum_{k=1}^{n}(a_k-\bar{a}_.)^2$	$\sigma_\varepsilon^2(1-\theta)+lr\sigma_{ga}^2+\frac{hr}{n-1}\sum_{k=1}^{n}a_k^2$
G x A	$\sigma_\varepsilon^2(1-\theta)+r\sigma_{ga}^2$	$\sigma_\varepsilon^2(1-\theta)+r\sigma_{ga}^2$	$\sigma_\varepsilon^2(1-\theta)+lr\sigma_{ga}^2$
Error (b)	$\sigma_\varepsilon^2(1-\theta)$	$\sigma_\varepsilon^2(1-\theta)$	$\sigma_\varepsilon^2(1-\theta)$

[1/] Considering a_k random; [2/] Considering a_k fixed and $ga_{ik} \sim NID(0,\sigma_{ga}^2)$; [3/] Considering $\sum_{k=1}^{n} a_k = 0$ e $\sum_{k=1}^{n} ga_{ik} = 0$ for all i; when $ga_{ik} \sim N(0,\frac{n-1}{n}\sigma_{ga}^2)$, $l = 1$; and when $ga_{ik} \sim N(0,\sigma_{ga}^2)$, $l = \frac{n}{n-1}$.

and the remaining covariances between errors are equal to zero then:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} - \frac{1}{n}\sigma_{ga}^2 + \frac{\theta\sigma_\varepsilon^2}{r}}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{n-1}{n}\sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_\varepsilon^2}{nr}\left[1+(n-1)\theta\right]} \quad \text{(equation b4)}$$

We can consider that $\sum_{k=1}^{n} ga_{ik} = 0$ for all values of i and $\sum_{k=1}^{n} a_k = 0$ are restrictions, while $cov(ga_{ik}, ga_{i'k}) = 0$ for all values of k, i and i' ($i \neq i'$) and $\varepsilon_{ijk} \sim NID(0, \sigma_\varepsilon^2)$ are assumptions. By adopting $\sum_{k=1}^{n} ga_{ik} = 0$ for all values i and assuming that $ga_{ik} \sim N(0, \sigma_{ga}^2)$ we can consider the $cov(ga_{ik}, ga_{ik'}) = -\frac{\sigma_{ga}^2}{n-1}$ and constant for all values of i, k, k' ($k \neq k'$). Since $cov(ga_{ik}, ga_{ik'}) = -\frac{\sigma_{ga}^2}{n-1}$ we have:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} - \frac{\sigma_{ga}^2}{n-1}}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_\varepsilon^2}{nr}} \quad \text{(equation b5)}$$

Considering that $\sum_{k=1}^{n} ga_{ik} = 0$ for all values of i and $\sum_{k=1}^{n} a_k = 0$ are restrictions, while $ga_{ik} \sim N(0, \sigma_{ga}^2)$; $cov(ga_{ik}, ga_{i'k}) = 0$ for all values of k, i and i' ($i \neq i'$); $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ and $cov(\varepsilon_{ijk}, \varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$ are assumptions and the remaining covariances between errors are equal to zero, we have:

$$\rho_1 = \frac{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} - \frac{\sigma_{ga}^2}{n-1} + \frac{\theta\sigma_\varepsilon^2}{r}}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \sigma_{ga}^2 + \frac{\sigma_\varepsilon^2}{r}} \quad \text{and}$$

$$\rho_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\gamma^2}{r} + \frac{\sigma_\varepsilon^2}{nr}\left[1+(n-1)\theta\right]} \quad \text{(equation b6)}$$

The above equations should be considered in the light of a factorial model in a randomized complete block design where $y_{ijk} = \mu + g_i + b_j + a_k + ga_{ik} + \varepsilon_{ijk}$ and these are the terms of the equation which was defined earlier.

According to the fixed and random effect models and their assumptions and restrictions and the expected mean squares for genotypes described in Table 2 when $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ and in Steel & Torrie (1980) when $\varepsilon_{ijk} \sim NID(0, \sigma_\varepsilon^2)$, expressions for ρ_1 are obtained which are similar to those obtained for the split-plot model. In the factorial model, however, $\frac{\sigma_\gamma^2}{r}$ is not added to the numerators or the denominator of the expressions. The expressions for ρ_2 are also similar to those obtained for the split-plot model, although in the factorial model $\frac{\sigma_\gamma^2}{r}$ is not added to the denominator of the expression. Furthermore, ρ_2 cannot be derived when $cov(\varepsilon_{ijk}, \varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$, because it is not possible to estimate σ_g^2 without adopting certain restrictions, with an equal number of equations being presented with respect to the parameters to be estimated (Table 2).

ANOVA estimator

Traditionally, the analysis of variance has been used to estimate the repeatability coefficient. Regardless of the model or the assumptions and restrictions used for each model, the ANOVA estimators for ρ_1 and ρ_2 are given by the expressions presented below:

$$\hat{\rho}_1 = \frac{MSG - MSGA}{MSG + (n-1)MSGA} \quad \text{and} \quad \hat{\rho}_2 = \frac{\hat{\sigma}_g^2}{\frac{MSG}{nr}}$$

where $\hat{\rho}_1$ and $\hat{\rho}_2$ = ANOVA estimators for ρ_1 and ρ_2, MSG = genotype mean square, MSGA = genotype x year interaction mean square, and $\hat{\sigma}_g^2$ = σ_g^2 estimator.

The expected genotype and genotype x year interaction mean squares necessary to obtain $\hat{\rho}_1$ and $\hat{\rho}_2$ using the split-plot and factorial models in randomized complete block designs are presented in Tables 1 and 2 when $\varepsilon_{ijk} \sim NID(0, \sigma_\varepsilon^2)$ and in Steel & Torrie (1980) when $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$.

Table 2. Expected mean squares based on a factorial model with a randomized complete block design with b_j fixed, g_i random, $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, $cov(e_{ijk}, e_{ijk'}) = \theta\sigma^2_\varepsilon$ and the remaining covariances between errors equal to zero.

Sources of variation	E(MS)[1/]	E(MS)[2/]	E(MS)[3/]
Blocks (B)	$\sigma^2_\varepsilon[1+(n-1)\theta]+\frac{hn}{r-1}\sum_{j=1}^{r} b_j^2$	$\sigma^2_\varepsilon[1+(n-1)\theta]+\frac{hn}{r-1}\sum_{j=1}^{r} b_j^2$	$\sigma^2_\varepsilon[1+(n-1)\theta]+\frac{hn}{r-1}\sum_{j=1}^{r} b_j^2$
Genotypes (G)	$\sigma^2_\varepsilon[1+(n-1)\theta]+r\sigma^2_{ga}+nr\sigma^2_g$	$\sigma^2_\varepsilon[1+(n-1)\theta]+r\sigma^2_{ga}+nr\sigma^2_g$	$\sigma^2_\varepsilon[1+(n-1)\theta]+nr\sigma^2_g$
Year (A)	$\sigma^2_\varepsilon(1-\theta)+r\sigma^2_{ga}+hr\sigma^2_a$	$\sigma^2_\varepsilon+r\sigma^2_{ga}+\frac{hr}{n-1}\sum_{k=1}^{n}(a_k-\bar{a}_\cdot)^2$	$\sigma^2_\varepsilon(1-\theta)+lr\sigma^2_{ga}+\frac{hr}{n-1}\sum_{k=1}^{n} a_k^2$
G x A	$\sigma^2_\varepsilon(1-\theta)+r\sigma^2_{ga}$	$\sigma^2_\varepsilon(1-\theta)+r\sigma^2_{ga}$	$\sigma^2_\varepsilon(1-\theta)+lr\sigma^2_{ga}$
Error	$\sigma^2_\varepsilon(1-\frac{n-1}{hn-1}\theta)$	$\sigma^2_\varepsilon(1-\frac{n-1}{hn-1}\theta)$	$\sigma^2_\varepsilon(1-\frac{n-1}{hn-1}\theta)$

[1/] Considering a_k random; [2/] Considering a_k fixed and $ga_{ik} \sim NID(0, \sigma^2_{ga})$; [3/] Considering $\sum_{k=1}^{n} a_k = 0$ e $\sum_{k=1}^{n} ga_{ik} = 0$ for all i; when $ga_{ik} \sim N(0, \frac{n-1}{n}\sigma^2_{ga})$, $l = 1$; and when $ga_{ik} \sim N(0, \sigma^2_{ga})$, $l = \frac{n}{n-1}$.

A numerical example

In order to develop our argument we shall use the experiment described in detail by Carvalho (1999) as a numerical example. The experiment consisted of a randomized complete block trial with seven replicates in which the mean number of healthy fruits per plant (MNHFPP) was assessed for 20 cocoa plant hybrids obtained from crosses between different cocoa clones. Each plot consisted of 12 plants distributed in three rows and four columns with a 3.0 x 3.0 m plant spacing. Two rows of cocoa trees were planted around the experiment area as a border. Temporary shading was provided by 3.0 x 3.0 m spaced banana trees and by cassava plants, using four cassava plants per cocoa tree. Permanent shading was also provided by planting 24.0 x 24.0 m spaced *Erythrina glauca*, with one additional plant on the diagonal.

The analysis of variance was carried out using split-plot and factorial models and the expected mean squares given in Tables 1 and 2 when $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$ and in Steel & Torrie (1980) when $\varepsilon_{ijk} \sim NID(0, \sigma^2_\varepsilon)$. The ρ_1 and ρ_2 estimates were calculated using an ANOVA estimator as indicated in the last section.

RESULTS AND DISCUSSION

Correlation between successive measurements

The ρ_1 estimates for the mean number of healthy fruits per plant (MNHFPP) always had the same value (Tables 3 and 4), regardless of the assumptions and restrictions used for each model. This repeatability is a function of the covariances $cov(\bar{y}_{i.k}, \bar{y}_{i.k'})$, $V(\bar{y}_{i.k})$ and $V(\bar{y}_{i.k'})$ which can be estimated from the variance components associated with the particular model. Adopting different models or different assumptions leads to variable values for the variance components (Searle, 1971; Steel & Torrie, 1980). However, combining these components results in the same estimate when expressing the covariances mentioned, as can be verified from the fact that the only ANOVA estimator was obtained for ρ_1, which is a function of only MSG and MSGA, which do not vary in the statistical model adopted. From a practical point of view, this invariance increases the reliability of the use of this repeatability estimate for selection purposes.

The magnitude of ρ_1 depends on the nature of the trait, the genetic properties of the population and the environmental conditions under which the genotypes are maintained. It should be pointed out that even though in the present study ρ_1 was obtained in trials involving randomized complete blocks, the same expression for ρ_1 could be derived when completely randomized and latin-square designs are considered. In latin-square designs row and column effects should be considered to be fixed, so that the $cov(\bar{y}_{i.k}, \bar{y}_{i.k'})$ will not be expressed by variances due to these effects. In randomized complete block designs, the block effect is considered to be fixed ($\sum_{k=1}^{r} b_k = 0$) for analogous reasons; fixed blocks are quite common in agricultural research (Kempthorne, 1952; Steel & Torrie, 1980; Piepho, 1994).

Table 3. Analysis of variance and repeatability estimates for the mean number of healthy fruits per plant (MNHFPP) assessed in a cocoa hybrid experiment using a factorial model in a randomized complete-block design.

Sources	d.f.	Mean squares
		NMFSP
Blocks (B)	6	764.458
Genotypes (G)	19	674.995 **
Years (A)	5	7162.621 **
G x A	95	59.913 **
Error	714	39.686
$\hat{\rho}_1 \pm EP(\hat{\rho}_1)$		0.621 ± 0.091
$\hat{\rho}_2$ [1/]		0.911
$\hat{\rho}_2$ [2/]		0.941

[1/] ρ_2 estimates considering b_j fixed, g_i random, $ga_{ik} \sim NID(0,\sigma^2_{ga})$ e $\varepsilon_{ijk} \sim NID(0,\sigma^2_{\varepsilon})$; [2/] ρ_2 estimates considering b_j fixed, g_i random, $\sum_{k=1}^{n} ga_{ik} = 0$, for all i, and $\varepsilon_{ijk} \sim NID(0,\sigma^2_{\varepsilon})$; [3/] $P < 0.01$.

In addition to the fixed block effect, several other assumptions are needed for the derivation, definition and use of ρ_1. One assumption can be that the variance between genotype measurements is the same over successive evaluation periods (years etc.), i.e. $(V(\bar{y}_{i.k}) = V(\bar{y}_{i.k'}))$. If this assumption is true then the repeatability estimate, defined as the correlation between successive measurements (over time) of the same genotype, can be considered to be the regression coefficient of a measurement as a function of the other (Turner and Young, 1969). In addition, $V(\bar{y}_{i.k})$ can be subdivided into variance between and within genotypes.

The variance between genotypes reflects the permanent differences between genotype measurements caused by the permanent environment and by genetic differences, expressed as $cov(\bar{y}_{i.k}, \bar{y}_{i.k'})$. In the study, we considered not only the genetic variance component confounded with the permanent environmental component (σ^2_g) as permanent differences, but also $\frac{\sigma^2_\gamma}{r}$, $-\frac{1}{n}\sigma^2_{ga}$, $-\frac{\sigma^2_{ga}}{n-1}$ and $\frac{\theta\sigma^2_\varepsilon}{r}$, alone or in combination. Thus, regardless of the statistical model used, ρ_1 represents the proportion of $V(\bar{y}_{ij.})$ which is caused by permanent differences between genotype measurements.

The within genotype variance measures the variation in measurements of the same genotype caused only by differences in the temporary environment, expressing the difference between $V(\bar{y}_{i.k})$ and $cov(\bar{y}_{i.k}, \bar{y}_{i.k'})$. Depending on the assumptions used, differences in the temporary environment can be represented by $\sigma^2_{ga} + \frac{\sigma^2_\varepsilon}{r}$, $\sigma^2_{ga} + (1-\theta)\frac{\sigma^2_\varepsilon}{r}$, $\frac{n}{n-1}\sigma^2_{ga} + \frac{\sigma^2_\varepsilon}{r}$ or $\frac{n}{n-1}\sigma^2_{ga} + (1-\theta)\frac{\sigma^2_\varepsilon}{r}$.

Another assumption to be considered in estimating ρ_1 is that all the genes that affect a trait should be expressed in all measurements. In other words, the expression of the trait during the various measurements will depend on identical physiological and developmental processes (Falconer, 1989; Chapman, 1985). Furthermore, when the split-plot model is used to derive ρ_1, it is necessary to consider that the mean plot-error $(\bar{\gamma}_{i.})$ will be constant in successive measurements of a genotype, although this is not strictly valid. Even if some of these assumptions are not taken into account, repeatability can be valid for application purposes.

When the genes that affect a specific measurement are not entirely the same as those that affect other

Table 4. Analysis of variance and repeatability estimates for the mean number of healthy fruits per plant (MNHFPP) assessed in a cocoa hybrid experiment using a split-plot model in a randomized complete-block design.

Sources	d.f.	Mean squares
		MNHFP
Blocks (B)	6	764.458
Genotypes (G)	19	674.995 [3/]
Error (a)	114	72.574
Years (A)	5	7162.621 [3/]
G X A	95	59.913 [3/]
Error (b)	600	33.437
$\hat{\rho}_1 \pm EP(\hat{\rho}_1)$		0.621 ± 0.091
$\hat{\rho}_2$ [1/]		0.853
$\hat{\rho}_2$ [2/]		0.892

[1/] ρ_2 estimates considering b_j fixed, g_i random and $ga_{ik} \sim NID(0,\sigma^2_{ga})$. By adopting $cov(e_{ijk}, e_{ijk'}) = 0$ or $cov(e_{ijk}, e_{ijk'}) = \theta\sigma^2$, we shall have the same ρ_2 estimate; [2/] ρ_2 estimates considering b_j fixed, g_i random and $\sum_{k=1}^{n} ga_{ik} = 0$. By adopting $cov(e_{ijk}, e_{ijk'}) = 0$ or $cov(e_{ijk}, e_{ijk'}) = \theta\sigma^2$, we shall have the same ρ_2 estimate; [3/] $P < 0.01$;

measurements, the variation within genotypes will not be only caused by the temporary environment. The variance between the means of genotype measurements, i.e. $(V(\bar{y}_{i..}))$, will increase due to the additional variance produced by the genotype x environment interaction. This additional variance may be sufficient to counter the reduction in variance caused by the temporary environment and consequently offset the increase in the precision of inference about the true genotype value, which represents the major advantage to be obtained from multiple measurements (Falconer, 1989).

An other important factor is that ρ_1 estimates always have the same value regardless of the statistical model used, so that data processing can be simplified by adopting a model that uses the mean of the experimental units of each genotype during each year (Cruz & Regazzi, 1994; Dias & Kageyama, 1988). This model can be represented as:

$$y_{ik} = \mu + g_i + a_k + E_{ik}$$

where y_{ik} = the mean of the j[th] blocks (j=1,...,r) referring to the i[th] genotype (i=1,...,h) during the kh year (k=1,...,n), μ = a constant inherent in all means for each genotype during each year, g_i = the effect of the i[th] genotype under the influence of the permanent environment, a_k = the effect of the k[th] year, and E_{ik} is the residue including $ga_{ik} + \bar{\varepsilon}_{i.k}$. On the basis of this model, $\hat{\rho}_1$ is expressed by:

$$\hat{\rho}_1 = \frac{MSG - MSRe}{MSG + (n-1)MSRe}$$

where $\hat{\rho}_1$ = the estimator of ρ_1, MSG = genotype mean square, and $MSRe$ = residue mean square.

The same estimator of ρ_1 can also be obtained by considering the correlation between measurements of the same genotype made at different locations, but in this case successive measurements are not made on the same individuals. In general, the factorial model is adopted, which uses the effect of the j[th] block on the k[th] location $(b_{j(k)})$, instead of b_j, to describe the trait measured for the i[th] genotype at the k[th] location (Cruz & Regazzi, 1994). However, the σ_g^2 component only represents genetic variance because the effect of the i[th] genotype is not under the influence of the permanent environment as is the case for repeatability derivation.

Upper limit of heritability

In contrast to $\hat{\rho}_1$, repeatability estimates at the upper limit of broad-sense heritability based on the mean of successive measurements (over time) of the same genotype ($\hat{\rho}_2$ values), for MNHFPP, acquire different values depending on the fixed and random effect models and their assumptions and restrictions used (Tables 3 and 4). This is due to the fact that ρ_2 expresses the proportion of $V(\bar{y}_{i..})$ attributable to the genetic variance component confounded with the permanent environment. As we stated earlier, the use of different statistical models leads to different estimates of one or more variance components. Repeatability also varies with the nature of the trait, the genetic properties of the population and the environmental conditions under which the genotypes are maintained.

In the estimation of ρ_2 it is possible to remove the variance caused by the temporary environment, but the variance due to the permanent environment usually continues to be totally or partially confounding with genetic variance. If the part of the proportion of variance due to the permanent environment, $(\frac{\sigma_\gamma^2}{r})$, is not confounded with genetic variance, the value of $\hat{\rho}_2$ tends to be lower when the split-plot model is used than when the factorial model is used, considering the same restrictions and assumptions.

When the factorial model is used, ρ_2 estimates are not obtained when $cov(\varepsilon_{ijk}, \varepsilon_{ijk'})$ has the value $\theta\sigma_\varepsilon^2$, since it is not possible to estimate σ_g^2 (Table 2). However, the same ρ_2 estimate is obtained when $cov(\varepsilon_{ijk}, \varepsilon_{ijk'}) = 0$ or $cov(\varepsilon_{ijk}, \varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$ is used in the split-plot model because

the estimator of σ_g^2 is the same in the two situations (Table 5). We should emphasized that even considering $cov(\varepsilon_{ijk}, \varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$, as done by Danford *et al.* (1960), genetic variance continues to be partially confounded with the permanent environment in the σ_g^2 component. If this component represented only genetic variance with respect to $cov(\varepsilon_{ijk}, \varepsilon_{ijk'}) = \theta\sigma_\varepsilon^2$, ρ_2 would be heritability itself, which, in fact, is not the case.

Sometimes ρ_2 can be much higher than heritability, but never lower (Jahufer et al., 1994) because σ_g^2 does not represent genetic variance only, and the genetic and non-genetic components of σ_g^2 are assumed to be variances and therefore cannot be negative. However, the variance caused by the effects of the permanent environment confounded with genetic variance is a covariance and can be negative.

Table 5. Estimators of σ_g^2 for a split-plot model, with b_j fixed and g_i random.

Restriction or assumption[1/]	Estimator of σ_g^2 [2/]
$ga_{ik} \sim NID(0, \sigma_{ga}^2)$	$\frac{MSG + MSEb - MSEa - MSGA}{nr}$
$\sum_{k=1}^{n} ga_{ik} = 0$ for all i	$\frac{MSG - MSEa}{nr}$

[1/] By adopting $cov(e_{ijk}, e_{ijk'}) = 0$ or $cov(e_{ijk}, e_{ijk'}) = \theta\sigma^2$, the same σ_g^2 estimator is obtained; [2/] MSG = genotype mean square, $MSEa$ = error (a) mean square, $MSEb$ = error (b) mean square, $MSGA$ = genotype x year interaction mean square.

According to Chapman (1985), this variance is a measure of the extent to which these permanent environmental effects cause covariance between repeated measures of the same trait. In most cases this covariance is positive since the influence of these effects results in measurements that deviate from the mean in the same direction. However, Chapman reported that under unusual circumstances the influence of an environmental effect may cause an increase in mean performance during one year and a reduction of mean performance during the next. When this influence is important, the covariance caused by these permanent environmental effects may be negative and may cancel the contribution of σ_g^2 and cause ρ_2 to acquire a lower value than heritability, *i.e.* zero or even negative.

Another important point is that $\hat{\rho}_2$ increases with an increase in the number of measurements (n). As n increases the variance due to the temporary environment decreases (Lush, 1945), and consequently the variance between the means of genotype measurements decreases and $\hat{\rho}_2$ increases. When $ga_{ij} \sim NID(0, \sigma_{ga}^2)$ is considered, for example, the reduction in variance due to the temporary environment is $1/n$ of the value obtained with annual measurements.

Repeatability estimates such as upper heritability limits are useful in predicting the genetic gain obtained by selection (Turner & Young, 1969). The increase of $\hat{\rho}_2$ with increasing number of measurements tends to increase progress per generation. However, when several evaluations are performed, the interval between generations tends to increase, reducing the progress per year. This partially reduces the gain obtained by the genetic gain per generation mentioned above, and the consideration of multiple measurements causes a decrease in the selection differential (Lush, 1945). The use of $\hat{\rho}_2$ in the prediction of selection gain may prove to be unreliable because it only reflects only the upper limit of heritability data, although in some trials (e.g. those involving perennial plants) $\hat{\rho}_2$ is obtained instead of broad sense heritability due to the long time required to estimate this heritability.

ACKNOWLEDGMENT

The authors thank CAPES for financial support.

RESUMO

Repetibilidade de caracteres avaliados em ensaios experimentais

Derivação das expressões de repetibilidade e obtenção dos respectivos estimadores de ANOVA foram realizadas, utilizando-se os modelos parcela subdividida e fatorial em delineamentos em blocos completos casualizados, bem como diferentes restrições, naturezas e pressuposições para os efeitos de cada modelo estatístico. Independentemente do modelo e das restrições, das naturezas e das pressuposições utilizadas, estimativas de repetibilidade como correlação entre medidas sucessivas (no tempo) de um mesmo genótipo assumem sempre o mesmo valor. Isto possibilita a utilização do modelo que utiliza as médias das unidades experimentais (média dos blocos) de cada genótipo, em cada tempo, para o cálculo desta repetibilidade. Para as estimativas de repetibilidade como limite superior da herdabilidade em sentido amplo, em nível de média de medidas sucessivas (no tempo) de um mesmo genótipo, esta independência

não foi verificada. Uma discussão sobre repetibilidade de caracteres avaliados em ensaios com delineamentos experimentais foi realizada.

REFERENCES

Carvalho, C.G.P. de 1999. Repetibilidade e seleção de híbridos de cacaueiro. Ph.D. Diss. UFV, Viçosa.

Chapman, A.B. 1985. General and quantitative genetics. Elsevier Science Publishers, Amsterdam.

Cruz, C.D. and Regazzi, A.J. 1994. Modelos biométricos aplicados ao melhoramento genético. Imprensa Universitária, Viçosa.

Danford, M.B.; Hughes, H.M. and McNee, R.C. 1960. On the analysis of repeated – measurements experiments. Biometrics. 16:547-565.

Dias, L.A.S. and Kageyama, P.Y. 1998. Repeatability and minimum harvest period of cacao (*Theobroma cacao* L.) in Southern Bahia. Euphytica. 102:29-35.

Falconer, D.S. 1989. Introduction to quantitative genetics. Longman, London.

Jahufer, M.Z.Z.; Cooper, M. and Brien, L.A. 1994. Genotypic variation for stolon and other morphological attributes of white clover (*Trifolium repens* L.) populations and their influence on herbage yield in the summer rainfall region of new south wales. Australian Journal of Agricultural Research. 45:703-720.

Kempthorne, O. 1952. The design and analysis of experiments. Wiley, New York.

Lush, J.L. 1945. Animal Breeding Plans. Iowa State College Press, Ames.

Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences. 10:235-322.

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.

Searle, S.R. 1971. Linear models. John Wiley & Sons, New York.

Steel, R.G.D. and Torrie, J.H. 1980. Principles and procedures of statistics. A biometrical approach. McGraw–Hill, New York.

Turner, H.N. and Young, S.S.Y. 1969. Quantitative genetics in sheep breeding. Cornell University, New York.

Received: May 24, 2002;

Accepted: October 28, 2002.