CULTIVAR STABILITY ANALYSIS USING A DISCONTINUOUS BI-SEGMENTED MODEL: UNBALANCED EXPERIMENTS[1]

LINDOLFO STORCK[2]

ABSTRACT - Groups of cultivar experiments carried out at different environments are often unbalanced since some cultivars are not tested in all the locations. An analysis of variance and phenotypic stability by the discontinuous bi-segmented model with corrections due to errors in the variables, for the unbalanced cases, is presented in this paper. It was observed, by simulation, that the higher the level of imbalance the lower is the accuracy of the estimates. However, losses of up to 25% of environments are tolerable and do not preclude a good description of the behavior of the cultivars under environmental variation if the given cultivar is present in at least two favorable and two unfavorable environments.

Index terms: phenotypic stability, error correction, unbalanced analysis.

ANÁLISE DE ESTABILIDADE DE CULTIVARES PELO MODELO BI-SEGMENTADO DESCONTÍNUO: GRUPOS DE EXPERIMENTOS NÃO BALANCEADOS

RESUMO - Nos experimentos com cultivares em diferentes ambientes, muitas vezes algumas cultivares não ocorrem em todos os ambientes, resultando em grupos de experimentos não balanceados. A análise da variância e de estabilidade fenotípica pelo modelo bi-segmentado descontínuo, com correções devido aos erros nas variáveis, para o caso não balanceado, é apresentado neste trabalho. Observou-se, via simulação, que quanto maior o nível de desbalanceamento menor é a precisão das estimativas. No entanto, perdas de até 25%, desde que uma dada cultivar ocorra em pelo menos dois ambientes favoráveis e dois ambientes desfavoráveis, é tolerável e leva a uma boa descrição do comportamento das cultivares frente à variação ambiental.

Termos para indexação: estabilidade fenotípica, correção de erro, análise não-balanceada.

[1] Accepted for publication on December 8, 1997.

[2] Eng. Agr., Dr., Prof. Titular, Dep. de Fitotecnia, Universidade Federal de Santa Maria, CEP 97105-900 Santa Maria, RS. Bolsista do CNPq.

INTRODUCTION

Analysis of stability is an analytical procedure of data applied to groups of experiments involving cultivars in different environments. Plant breeders use this method to select cultivars with pre-established characteristics. A first attempt in assessing the individual behavior of cultivars is making a joint analysis of the experiments with decomposition of the effect of the environments plus the environmental interaction with cultivars in effect of the environments within each cultivar (Yates & Cochran, 1938). There have been contributions to this methodology (Plaisted & Peterson, 1959; Plaisted, 1960) but they were only initial procedures before applying more informative methods. In this respect, the greatest advance was obtained by the simple linear regression, using the logarithmic transformation, of the mean of a given cultivar on the mean of all cultivars in each given environment (Finlay & Wilkinson, 1963). Later, Eberhart & Russell (1966), modified the method obtaining a better interpretation of its results.

The biggest problem reported by Eberhart & Russell (1966) is the lack of independence of the errors of the dependent variable (mean of the cultivar) with the independent variable (and its error) which is the mean of all the cultivars in a given environment less the general mean (named the environmental index). Thus, various practical works (Fripp & Caten, 1971; Fripp, 1972; Perkins & Jinks, 1973; Wood, 1978; Carvalho et al., 1982; Fakorede & Opeke, 1986) were carried out to compare results of the analysis and interpretation of stability of the cultivars, using different types of independent variables. They led to the conclusion that the best way of estimating the environmental value (independent variable) is by the environmental index.

Verma et al. (1978) proposed an alternative regression technique that consists of adjusting two straight segments separately. One for the negative environments index (lower than the general mean) and the other for the positive environments index plus the lowest (absolute value) negative index, to give continuity in the regression line. From this, several other segmentation models were suggested (Cruz et al., 1989; Silva, 1995). However, the discontinuous bi-segmented model has better characteristics (Storck, 1995) because it has independence between the angles of the two segments; the tests on the hypothesis for the second segment are independent from the discontinuity; the estimated values fit the Gompertz growth curve well; the parameter estimates are more disperse allowing better discrimination among the cultivars; the model adjusts well in a wide range of situations, that is, adjusts to a complete growth curve, over the initial half and final half of the curve, and over the initial third, middle and end of it. Furthermore, for this model, the algorithms using corrections due to errors in the dependent and independent variables and their correlation are available for both parameter estimation and hypothesis testing (Storck & Vencovsky, 1994).

In practice some cultivars are not found in all the environments due to substitution of older cultivars by newly released ones.

The objective of this paper is to develop an adequate algorithm for the stability analysis, by the discontinuous bi-segmented model (Storck & Vencovsky, 1994) with correction due to the errors in the dependent and independent variables and the correlations among these errors for unbalanced experiments.

MATERIAL AND METHODS

The present work used yield data from a group of experiments with short cycle maize cultivars, carried out in Rio Grande do Sul State, RS, Brazil, in the agricultural year of 1992/93. The experiments, in each one of 14 environments, formed the state network. The experimental design was the randomized complete block with four replications and involved 27 cultivars. The locations were: 1) Pelotas; 2) Passo Fundo; 3) Encruzilhada; 4) Não-Me- -Toque; 5) Rio Grande; 6) Aratiba; 7) Nova Petrópolis; 8) Capão do Leão; 9) São Borja; 10) Santa Rosa; 11) Augusto Pestana; 12) Cruz Alta; 13) Ibirubá and 14) Vacaria.

The analysis of stability was carried out by the discontinuous bi-segmented model with corrections because of errors in the variables (Storck & Vencovsky, 1994), using all the data from the group of experiments. Losses (level of imbalance) of 5, 10, 15, 20 and 25% of cultivars in any environment over the whole group formed the unbalanced experimental groups. The simulations were truncated so that each cultivar would occur in at least half of the 14 environments. The least significance levels (LSL) of the tests of hypothesis of the parameters in the balanced group were compared with the means of the LSL of the 1,000 groups of losses simulated at each level of imbalance.

For the analysis of variance and stability of unbalanced experimental groups the algorithms (Storck & Vencovsky, 1994) were modified accordingly as follows: Considering the results of the J experiments of I cultivars in a randomized complete block design with K replications the mathematical model is given by

$$Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + b_{k(j)} + e_{ijk}$$

where

α_i is the fixed effect of the ith cultivar and i = 1, 2, ...I;

τ_j is the random effect of the jth environment (experiment) j = 1, 2, ...J;

γ_{ij} is the random effect of the ith cultivar with the jth environment;

$\mathbf{b}_{k(j)}$ is the random effect of the kth block within the jth environment; and

e_{ijk} is the random effect of experimental error.

The variables, $\mathbf{n}_{ijk}=1$ if the ith cultivar is present in the jth environment and kth replication and $\mathbf{n}_{ijk}=0$, if not; $\mathbf{w}_{ij}=1$ if the ith cultivar is in the jth environment and $\mathbf{w}_{ij}=0$ if not, are indicator variables.

For the joint analysis of variance of this group of experiments the notation R(.) was adopted for the reduction of the sum of the squares due to the effects added to the model (Searle, 1971). Therefore, $R(\alpha/\mu,\tau)$ is the partial effect of cultivars with N1=I-1 degrees of freedom; $R(\tau/\mu,\alpha)$ is the partial effect of the environment with N2=J-1 degrees of freedom; $R(\gamma/\mu,\alpha,\tau)$ is the effect of the interaction with $N3=\Sigma_{ij}w_{ij}$-N1-N2-1 degrees of freedom. Using the results of the analysis of variance of the experiments in each environment it was possible to obtain the sum of the squares of error by the SSError = $\Sigma_j SSE_j$ with N5 = $\Sigma_j DFE_j$ degrees of freedom (DF) where SSE_J and DFE_j are the errors sums squares and degrees of freedom at the jth environment, respectively, and if the design were randomized complete block then SSBl = $\Sigma_j SSB_j$ with N4 = $\Sigma_j DFB_j$ degrees of freedom where the SSB_j and DFB_j are the blocks sums square and degrees of freedom at the jth environments. The mean squares of interest are $V2=R(\tau/\mu,\alpha)/N2$; $V3=R(\gamma/\mu,\alpha,\tau)/N3$; V4=SSBl/N4; and, V5 = SSError/N5.

An important hypothesis to be tested is the significance of the interaction variance ($H_0:\sigma^2_\gamma=0$), which is tested by the F distribution of V3/V5. Another hypothesis is about significance of the environment variance ($H_0: \sigma^2_\tau=0$) which is tested by the F distribution (V2+V5)/(V3+V4) with g1=$(V2+V5)^2 / (V2^2/N2+V5^2/N5)$ and g2 = $(V3+V4)^2 / (V3^2/N3+V4^2/N4)$ degrees of freedom. If the variances of the interaction and the environment are significant, the partitioning of the source of variation due to "interaction plus environments" into "environments within cultivar" is carried out. The mathematical model is as follows:

$$Y_{ijk} = \mu + \alpha_i + \tau_{j(i)} + b_{k(j)} + e_{ijk}$$

where

$\tau_{j(i)}$ is the random effect of the jth environment within the ith cultivar.

The sum of the squares of environment within the ith cultivar $SSA/C_i = (1/n_{ij.})\sum_j w_{ij} Y^2_{ij.} - (1/n_{i..})Y^2_{i..}$ with the degrees of freedom given by $N2(i) = \sum_j w_{ij} - 1$ and $V2(i) = SSA|Ci/N2(i)$, for i=1, 2, ..., I. To test the significance of the variance of environment within the ith cultivar, it is taken that under the null hypothesis, $Ho(\tau): \sigma^2_{\tau(i)} = 0$, the statistic F = I.V2(i) / (N1.V5+V4) has F distribution with gi = N2(i) and g2 = (N1.V5+V4) / ($N1^2V5^2/N5 + V4^2/N4$) degrees of freedom. For the cultivars where the environmental variance is significant, the analysis of stability according to the discontinuous bi-segmented model (Storck & Vencovsky, 1994) is carried out, and the model is characterized by the functional and structural relationship given by:

$$\overline{Y}_{ij.} = \beta 0_i + \beta 1_i \tau_j + \beta 2_i \tau_j Z_j + \beta 3_i Z_j + \delta_{ij} + \varepsilon_{ij}$$

$$\hat{\tau}_j = \tau_j + \upsilon_j$$

$$\hat{\tau}_j Z_j = \tau_j Z_j + \upsilon_j Z_j$$

$Z_j = 1$, if $\hat{\tau}_j > 0$ or 0 if $\hat{\tau}_j \leq 0$, where for the ith cultivar the βo_i parameter is the value of the function at the point τ_j of the first line segment; the parameter $\beta 1_i$ is the slope of the first line segment; the parameter $\beta 2_i$ is the difference between the slopes of the two segments of the line such that $\beta 1_i+\beta 2_i$ is the slope of the second segment of the line; the parameter $\beta 3_i$ measures the discontinuity between the two segments of the line; the parameter δ_{ij} is the deviation of the jth observation of the ith cultivar from the model under the assumption of independence between the j and $E(\delta_{ij}) = 0$ and $E(\delta^2_{ij}) = \sigma^2_{\delta i}$ for any j; $\hat{\tau}_j$ is the estimated effect of the jth environment,

$$\hat{\tau}_j = (1/n_{.j.})\sum_i w_{ij} Y_{ij.} - (1/n_{...})\sum_{ij} w_{ij} Y_{ij.}$$

The error ε_{ij} associated with the $Y_{ij.}$ observation under the assumption of $E(\varepsilon_{ij}) = 0$, $E(\varepsilon^2_{ij}) = \sigma^2_e$ and ε_{ij} independent of δ_{ij} is estimated by the formula

$$\hat{\sigma}^2_\varepsilon = (1/Mh(I)Mh(K))\{V4 + (Mh(I) - 1)V5\},$$

$$Mh(K) = \sum_{ij} w_{ij} / \sum_{ij}(1/n_{ij.})$$ where and $$Mh(I) = J / \sum_j (1/w_{.j})$$

.

The number of degrees of freedom of $\hat{\sigma}^2_\varepsilon$ is obtained by the expression

$$n_\varepsilon = (V4 + Mh(I)V5)^2 / \{V4^2/N4 + (Mh(I) - 1)^2 V5^2/N5\}.$$

When υ_j is the error associated with the estimator $\hat{\tau}_j$ the variance of υ_j is estimated for the ith cultivar by $\hat{\sigma}_{\upsilon}^2 = (J-1)V4/(JMh_{(I)}Mh_{(K)})$ to provide a solution to the system.

Furthermore, for the ith cultivar, $\hat{cov}(\varepsilon_{ij};\upsilon_j) = \hat{\sigma}_{\upsilon}^2$; $\hat{cov}(\varepsilon_{ij};\upsilon_j Z_j) = p_i\hat{\sigma}_{\upsilon}^2$; $p_i = (1/\sum_j w_{ij})\sum_j w_{ij}Z_j$; $\hat{cov}(\upsilon_j;\upsilon_j Z_j) = p_i\hat{\sigma}_{\upsilon}^2$; $\hat{var}(\upsilon_j Z_j) = p_i\hat{\sigma}_{\upsilon}^2$.

An estimation of the parameters of the model, by the moment method, follows the solution of the following expressions:

$$[\hat{\beta}1_i;\hat{\beta}2_i;\hat{\beta}3_i]' = [Mxx_i - S(\upsilon\upsilon)]^{-1}[Mxy_i - S(\varepsilon\upsilon)];$$

$$\hat{\beta}o_i = \overline{Y}_i.. - \hat{\beta}1_i\overline{\tau} - \hat{\beta}2_i\overline{\tau z} - p_i\hat{\beta}3_i;$$

$$\hat{\sigma}_{\delta i}^2 = Svv_i - \hat{\underset{\sim}{\beta}}_i{}'S(\upsilon\upsilon)\hat{\underset{\sim}{\beta}}_i + 2S(\upsilon\varepsilon)\hat{\underset{\sim}{\beta}}_i - \hat{\sigma}_{\varepsilon}^2$$

where:

$$Mxx_i = \begin{vmatrix} S^2(\hat{\tau}) & S(\hat{\tau};\hat{\tau}z) & S(\hat{\tau};z) \\ & S^2(\hat{\tau};z) & S(\hat{\tau}z;z) \\ & \text{simetric} & S^2(z) \end{vmatrix}$$

$$S(\upsilon\upsilon) = \begin{vmatrix} \hat{\sigma}_{\upsilon}^2 & p_i\hat{\sigma}_{\upsilon}^2 & 0 \\ p_i\hat{\sigma}_{\upsilon}^2 & p_i\hat{\sigma}_{\upsilon}^2 & 0 \\ 0 & 0 & 0 \end{vmatrix} \quad S(\varepsilon\upsilon) = \begin{vmatrix} \hat{\sigma}_{\upsilon}^2 \\ p_i\hat{\sigma}_{\upsilon}^2 \\ 0 \end{vmatrix}$$

$$Mxy_i = [S(\overline{Y}_i..;\hat{\tau})\,S(\overline{Y}_i..;\hat{\tau}z)\,S(\overline{Y}_i..;z)]';$$

$$Svv_i = \sum_j w_{ij}(\overline{Y}_{ij}. - \hat{\beta}o_i - \hat{\beta}1_i\hat{\tau}_j - \hat{\beta}2_i\hat{\tau}_j Z_j - \hat{\beta}3_i Z_i)^2 / \left(\sum_j w_{ij} - 4\right);$$

$$S(\overline{Y}_i..;\hat{\tau}) = \left[\sum_j w_{ij}\overline{Y}_{ij}.\hat{\tau}_j - \left(\sum_j w_{ij}\overline{Y}_{ij}.\right)\left(\sum_j w_{ij}\hat{\tau}_j\right)/\sum_j w_{ij}\right]/\left(\sum_j w_{ij} - 1\right);$$

$$S(\overline{Y}_i..;\hat{\tau}z) = \left[\sum_j w_{ij}\overline{Y}_{ij}.\hat{\tau}_j Z_j - \left(\sum_j w_{ij}\overline{Y}_{ij}.\right)\left(\sum_j w_{ij}\hat{\tau}_j Z_j\right)/\sum_j w_{ij}\right]/\left(\sum_j w_{ij} - 1\right);$$

$$S(\overline{Y}_i..;z) = \left[\sum_j w_{ij}\overline{Y}_{ij}.Z_j - \left(\sum_j w_{ij}\overline{Y}_{ij}.\right)\left(\sum_j w_{ij}Z_j\right)/\sum_j w_{ij}\right]/\left(\sum_j w_{ij} - 1\right);$$

$$S^2(\hat{\tau}) = \left[\sum_j w_{ij}\hat{\tau}_j^2 - \left(\sum_j w_{ij}\hat{\tau}_j\right)^2/\sum_j w_{ij}\right]/\left(\sum_j w_{ij} - 1\right);$$

$$S(\hat{\tau};\hat{\tau}z) = \left[\sum_j w_{ij}\hat{\tau}_j^2 Z_j - \left(\sum_j w_{ij}\hat{\tau}_j\right)\left(\sum_j w_{ij}\hat{\tau}_j Z_j\right)/\sum_j w_{ij}\right]/\left(\sum_j w_{ij} - 1\right);$$

$$S(\hat{\tau};z) = \left[\sum_j w_{ij}\hat{\tau}_j Z_j - \left(\sum_j w_{ij}\hat{\tau}_j\right)\left(\sum_j w_{ij}Z_j\right)/\sum_j w_{ij}\right]/\left(\sum_j w_{ij} - 1\right);$$

$$S^2(\hat{\tau}z) = \left[\sum_j w_{ij}\hat{\tau}_j^2 Z_j - \left(\sum_j w_{ij}Z_j\right)^2 / \sum_j w_{ij}\right] / \left(\sum_j w_{ij} - 1\right);$$

$$S(\hat{\tau}z; z) = \left[\sum_j w_{ij}\hat{\tau}_j Z_j - \left(\sum_j w_{ij}\hat{\tau}_j Z_j\right)\left(\sum_j w_{ij}Z_j\right) / \sum_j w_{ij}\right] / \left(\sum_j w_{ij} - 1\right);$$

$$S^2(z) = \left[\sum_j w_{ij}Z_j - \left(\sum_j w_{ij}Z_j\right)^2 / \sum_j w_{ij}\right] / \left(\sum_j w_{ij} - 1\right);$$

$$\overline{Y}_{i..} = \sum_{jk} n_{ijk} Y_{ijk} / \sum_{jk} n_{ijk};$$

$$\overline{\tau Z} = \sum_j w_{ij}\hat{\tau}_j Z_j / \sum_j w_{ij};$$

$$\overline{\tau} = \sum_j w_{ij}\hat{\tau}_j / \sum_j w_{ij}.$$

The determination coefficient (R^2) for the ith cultivar is calculated by:

$$R_i^2 = 100\left\{SSA \mid C_i - \left(\sum_j w_{ij} - 4\right) Svv_i\right\} / SSA \mid C_i$$

The degrees of freedom of the estimate $\hat{\sigma}_{\delta i}^2$ is obtained by $fo_i = \left(\hat{\pi}_i - 1\right)^2 / \left\{\hat{\pi}_i / f1_i + 1/f2_i\right\}^2$, where

$\hat{\pi}_i = M1_i / \left(M2_i + M3_i + M4\right);$

$f1_i = \sum_j w_{ij} - 4;$

$f2_i = \left(M2_i + M3_i + M4\right)^2 / \left(M2_i^2 / N4 + M3_i^2 / N4 + M4^2 / n_\varepsilon\right);$

$M1_i = Svv_i;$

$M2_i = \hat{\beta}1_i \hat{\sigma}_\upsilon^2 \left(\hat{\beta}1_i + p_i \hat{\beta}2_i - 2\right);$

$M3_i = p_i \hat{\beta}2_i \hat{\sigma}_\upsilon^2 \left(\hat{\beta}1_i + \hat{\beta}2_i - 2\right);$

$M4 = \hat{\sigma}_\varepsilon^2.$

The hypotheses $Ho: \beta1_i = 1$; $Ho: \beta2_i = 0$ e $Ho: \beta3_i = 0$ for the ith cultivar are tested by the t test, by calculating the following statistics:

$$t1_i = \left(\hat{\beta}1_i - \beta1_i\right) / \left[S^2\left(\hat{\beta}1_i\right)\right]^{1/2};$$

$$t2_i = \left(\hat{\beta}2_i - \beta2_i\right) / \left[S^2\left(\hat{\beta}2_i\right)\right]^{1/2};$$

$t3_i = \left(\hat{\beta}3_i - \beta3_i\right) / \left[S^2\left(\hat{\beta}3_i\right)\right]^{1/2}$ with degrees of freedom equal to $gl_i = \Sigma_j w_{ij} - 4$.

The variance-covariance matrix of $\underset{\sim}{\hat{\beta}_i}$ is obtained by:

$$\left| S^2(\hat{\beta}1_i) \quad S(\hat{\beta}1_i; \hat{\beta}2_i \quad S(\hat{\beta}1_i; \hat{\beta}3_i \right|$$

$$\hat{V}(\hat{\underset{\sim}{\beta}}_i) = \begin{vmatrix} S^2(\hat{\beta}2_i) & S(\hat{\beta}2_i;\hat{\beta}3_i \\ \text{simetric} & S^2(\hat{\beta}3_i) \end{vmatrix}$$

$$\hat{V}(\hat{\underset{\sim}{\beta}}_i) = [Mxx_i - S(\upsilon\upsilon)]^{-1}Svv_i/(\Sigma_j w_{ij} - 1) +$$
$$[Mxx_i - S(\upsilon\upsilon)]^{-1}.[S(\upsilon\upsilon)Svv_i + \tilde{S}uv_i\tilde{S}vu_i].[Mxx_i - S(\upsilon\upsilon)]^{-1}/(\Sigma_j w_{ij} - 1) +$$

$$[Mxx_i - S(\upsilon\upsilon)]^{-1}.[S(\upsilon\upsilon)Srr_i + \tilde{S}uv_i\tilde{S}vu_i].[Mxx_i - S(\upsilon\upsilon)]^{-1}/N5$$, where:

$Srr_i = Gi.Sww.Gi'$; $G_i = [\,1 \;\; -\hat{\beta}1_i \;\; -\hat{\beta}2_i \;\; -\hat{\beta}3_i\,]$;

$$Sww = \begin{vmatrix} \hat{\sigma}^2_{\varepsilon} & S(\upsilon\varepsilon) \\ S(\varepsilon\upsilon) & S(\upsilon\upsilon) \end{vmatrix}$$

and $\tilde{S}uv_i = S(\varepsilon\upsilon) - S(\upsilon\upsilon)\hat{\underset{\sim}{\beta}}_i$

The hypothesis Ho: $\sigma^2_{\delta i}$ is tested by the F test where the statistic F = $(Mh_{(k)}\hat{\sigma}^2_{\delta i} + V5)/V5$ has g1 and g2 degrees of freedom, and g2 = N5 and

$$g1 = [Mh_{(K)}\hat{\sigma}^2_{\delta i} + V5]^2 / [(Mh_{(K)})^2(\hat{\sigma}^2_{\delta i})^2 / f_{0_i} + V5^2/N5].$$

RESULTS AND DISCUSSION

The algorithm developed for the stability analysis by the discontinuous bi-segmented unbalanced model and correction due to errors in the variables is the detailed in the methodology. It is evident that only with specific computer processes the application of these procedures is viable. Thus a software (BSDD) is available at the electronic address " ftp://ftp.ufsm.br/pub/pc/misc/bsdd ".

Using the software BSDD the set of data mentioned in the methodology was analyzed. Tables 1, 2 and 3 show the results of this analysis. The environmental variance and the interaction between environment and cultivars were significant (P<0.001), as was the variance of environments within all the cultivars (Table 1). This set of data is, therefore, adequate for the stability analysis. The wide range of the environment means variation, 7.506 t/ha (Table 2), also favors the study of cultivar behavior due to environmental variation. The chi-square and the maximum F tests allowed the conclusion that the variance of the mean squares of the error of the environments (Table 2) were not homogeneous. It was necessary, therefore, to correct the degrees of freedom for the test of hypothesis of the interaction (Pimentel-Gomes, 1985).

The cultivar assessment carried out by the estimated parameters (Table 3) shows that cultivars 3, 4, 20, 22, 24 and 25 performed differently from the general mean of cultivar behavior. This low number of different cultivars is due to the good fit of the model, that is, a very high determination coefficient (average of 98%) for all the cultivars (Table 3). Even so, 17 of the 27 cultivars had significant deviation variances, indicating that the determination coefficient in this case, because it was not corrected by the errors in the variables, should be replaced by the deviation variance as parameter for cultivar selection (Storck & Vencovsky, 1994).

Further, the cultivars are widely commercialized and assessed by growers and, therefore, the poorest performers had already been replaced by market forces which made them more similar.

TABLE 1. The variance analysis for grain yield (t/ha) with partitioning of the interaction.

Source of variation	DF	MS	F (under Ho)
Cultivar(C)	26	12.1615	4.20*
Environment (A)	13	573.2669	75.65*
Interaction CxA	338	2.8974	2.75*
Block/Environment	42	4.6941	4.46*
Error	1092	1.05322	
Environment within cultivar:			
1 (AG E 10502)	13	13.2271	11.13*
2 (AG E 10501)	13	13.8568	11.66*
3 (AG E 10401)	13	18.1514	15.28*
4 (AG 521)	13	19.7627	16.63*
5 (AG 223)	13	15.7377	13.25*
6 (AGROMEN 2007)	13	29.4090	24.75*
7 (AGROMEN 2010)	13	21.1039	17.76*
8 (AGROMEN 2014)	13	33.1207	27.88*
9 (AGROMEN 2016)	13	31.3422	26.38*
10 (EXP 9004)	13	30.6268	25.78*
11 (Xl 212-Exp 9101)	13	18.7349	15.77*
12 (C 506)	13	25.9644	21.85*
13 (BR 205)	13	19.0056	16.00*
14 (BR 206)	13	20.7452	17.46*
15 (DINA 70)	13	35.8123	30.14*
16 (DINA 170)	13	51.2394	43.13*
17 (DINA 771)	13	35.1417	29.58*
18 (G 85-S-G800)	13	25.8135	21.73*
19 (HATZ 1000)	13	26.1939	22.05*
20 (ICI 8447)	13	26.1171	21.98*
21 (ICI 8452)	13	18.7422	15.78*
22 (ICI 8418)	13	27.6203	23.25*
23 (CC EXP 7)	13	15.6887	13.21*
24 (CC 8993-7)	13	10.0763	8.48*
25 (AG 64 A)	13	19.8288	16.69*
26 (SAVE 394)	13	20.6112	17.35*
27 (XL 560)	13	24.9254	20.98*

* F test significant at 5% of probability.

The means of the cultivars in the inferior environments (LM), superior environments (HM) and the general mean (GM) (Table 3) coupled with the estimates of the other parameters can be used to select or discard cultivars. For example, cultivars 4 and 12 performed well in any environment while cultivars 3, 12 and 20 were suitable for above average environments and cultivars 5, 6 and 16 for higher environments.

Tables 4, 5, 6 and 7 show the least significance levels (LSL) of the tests of hypothesis for the 27 cultivars obtained in the analysis of the complete data (0% losses) and the means of the LSL obtained by simulation with different levels of losses in the data. It may be noted that for cultivars with low LSL (where Ho was rejected) the increase in the LSL means occurred with the increase in the level of data loss, which is a serious problem. The contrary happens where the LSL are high, which has no practical effect. Of the parameters analyzed, the lack of fit variance was the most sensitive to the increase in the data imbalance. Different parameters show different degrees of sensibility to imbalance. In the simulation, losses of up to 25% were possible to analyze by the algorithm developed, as these losses did not result in less than 50% of the 14 environments for a given cultivar. Thus, possible problems with undetermined solutions or nil degrees of freedom are avoided. With this 25% loss limitation for a determined cultivar, in the simulation of 1000 unbalanced groups only one case of indetermination was recorded. This happened when a cultivar existed only in a superior or in an inferior environment.

TABLE 2. Mean (t/ha) environmental index (), indicator variable (Z) mean squares (MS) and coefficient of variation (CV) of the different environments (ENV) where the experiments were carried out.

ENV	Mean	$\hat{\tau}$	Z	$\hat{\tau}$ Z	Error MS	Block MS	CV%
1	6.9801	-0.1655	0	0.0000	0.385788	0.724091	8.90
2	12.0313	4.8857	1	4.8857	0.968337	2.584802	8.18
3	4.5582	-2.5873	0	0.0000	1.687649	14.679380	28.50
4	10.1952	3.0497	1	3.0497	0.464020	1.051839	6.68
5	4.7228	-2.4227	0	0.0000	3.944384	9.160418	42.05
6	7.7499	0.6043	1	0.6043	2.429153	4.240406	20.11
7	8.9643	1.8188	1	1.8188	0.547456	4.000976	8.25
8	5.7979	-1.3477	0	0.0000	0.298736	0.242167	9.43
9	5.5325	-1.6130	0	0.0000	0.660098	3.223270	14.69
10	5.1510	-1.9945	0	0.0000	0.386085	0.470411	12.06
11	4.5250	-2.6205	0	0.0000	0.708830	17.398417	18.61
12	8.9755	1.8300	1	1.8300	0.921452	2.866071	10.69
13	6.7951	-0.3505	0	0.0000	0.597184	1.471682	11.37
14	8.0588	0.9133	1	0.9133	0.745933	3.604023	10.72
Mean	7.1460	0.0000	0.4286	0.9358	1.053222	4.694140	15.02

TABLE 3. Estimates of the parameters β0, β1, β2 and β3 of the discontinuous bi-segmented model, determination coefficient (R^2), lack of fit variance (VD) and means (t/ha) in the lower (LM) higher (HM) and in the general (GM) environments for the assessed cultivars.

Cultivar	βo	β1	β2	β3	R^2(%)	VD	LM	HM	GM
1	7.22	0.85	-0.10	-0.14	98.3	0.04085	5.83	8.70	7.06
2	7.53	0.91	-0.10	-0.43	98.0	0.09336	6.03	8.87	7.25
3	7.75	1.12	-0.48	0.32	98.9	0.00001	5.92	9.46	7.44
4	7.32	0.58	0.79*	-0.67	99.1	0.00001	6.38	9.64	7.78
5	7.88	1.04	-0.25	-0.40	97.9	0.16836*	6.17	9.27	7.47
6	6.83	0.85	0.27	0.56	97.6	0.63919*	5.45	9.84	7.33
7	7.15	1.18	-0.16	-0.94	95.7	0.91903*	5.22	8.42	6.59
8	7.73	0.89	0.54	-0.03	98.5	0.38472*	6.27	10.82	8.22
9	7.99	1.24	-0.09	0.01	99.0	0.14672*	5.95	10.52	7.91
10	6.33	0.84	-0.09	1.77	97.4	0.77135*	4.96	9.72	7.00
11	7.29	0.84	-0.01	0.31	97.9	0.26070*	5.92	9.43	7.42
12	7.80	1.06	-0.18	0.76*	99.7	0.00001	6.07	10.48	7.96
13	6.58	0.86	0.07	-0.26	95.4	0.88116*	5.17	8.35	6.54
14	7.32	1.15	-0.59	0.55	96.7	0.62515*	5.43	9.09	7.00
15	6.70	1.23	-0.08	0.34	97.9	0.71003*	4.69	9.54	6.77
16	7.21	1.41	0.29	-0.35	98.7	0.59647*	4.90	10.57	7.33
17	7.16	1.25	0.30	-1.03	98.0	0.65004*	5.12	9.50	7.00
18	7.07	1.13	-0.16	0.23	99.0	0.07394	5.22	9.42	7.02
19	6.07	0.98	0.02	0.44	98.7	0.20100*	4.46	8.70	6.28
20	7.59	1.23	-0.31	0.30	99.8	0.00001	5.58	9.90	7.43
21	7.35	0.99	-0.22	0.19	98.1	0.20369*	5.72	9.23	7.23
22	7.71	1.53*	-0.48	-0.64	99.3	0.00001	5.21	9.36	6.99
23	7.56	1.18	-0.16	-1.67	96.1	0.54567*	5.63	8.11	6.69
24	6.15	0.33*	0.39	0.44	97.7	0.03721	5.60	8.15	6.70
25	5.91	0.28	0.78	0.79	97.7	0.31742*	5.44	9.02	6.98
26	6.61	0.96	-0.25	0.62	97.4	0.44001*	5.04	8.77	6.64
27	7.11	1.09	0.27	-1.06	98.7	0.11047	5.33	9.03	6.91
Mean	7.15	1.00	0.00	0.00	98.0	0.32654	5.51	9.33	7.15

* Significant at 1% of probability.

TABLE 4. Least significance levels obtained in the tests of hypothesis of the parameter β1 in the simulation of different cultivar loss levels.

Cultivar	Loss levels (%)					
	0	5	10	15	20	25
1	0.5190	0.5585	0.5747	0.5906	0.5750	0.5816
2	0.7320	0.6918	0.6705	0.6449	0.6298	0.6338
3	0.5810	0.6074	0.6211	0.6339	0.6271	0.6153
4	0.0600	0.0714	0.0843	0.1032	0.1170	0.1385
5	0.8630	0.8246	0.7841	0.7528	0.7100	0.6792
6	0.7040	0.6851	0.6631	0.6471	0.6398	0.6226
7	0.6980	0.6937	0.6932	0.6736	0.6703	0.6579
8	0.7480	0.7376	0.7282	0.6986	0.6742	0.6611
9	0.6190	0.5503	0.5215	0.5030	0.5121	0.4969
10	0.7050	0.6969	0.6790	0.6716	0.6835	0.6607
11	0.6090	0.6277	0.6395	0.6535	0.6629	0.6653
12	0.6810	0.6749	0.6663	0.6567	0.6467	0.6138
13	0.7550	0.7453	0.7281	0.7285	0.7083	0.6905
14	0.7030	0.7047	0.7187	0.6978	0.6857	0.6923
15	0.5910	0.6073	0.6194	0.6331	0.6270	0.6437
16	0.3100	0.3815	0.4171	0.4423	0.4499	0.4334
17	0.5500	0.5722	0.5807	0.5866	0.5884	0.6083
18	0.5920	0.5990	0.6127	0.5953	0.6171	0.5896
19	0.9490	0.9312	0.9179	0.8963	0.8786	0.8663
20	0.0760	0.1045	0.1339	0.1759	0.2232	0.2443
21	0.9770	0.8932	0.8400	0.7882	0.7560	0.7125
22	0.0260	0.0389	0.0511	0.0723	0.0938	0.1291
23	0.6360	0.6451	0.6461	0.6478	0.6530	0.6279
24	0.0140	0.0231	0.0362	0.0435	0.0670	0.0998
25	0.0450	0.0581	0.0723	0.0891	0.1129	0.1379
26	0.8980	0.8311	0.7534	0.7042	0.6688	0.6204
27	0.7270	0.7378	0.7337	0.7338	0.7310	0.6914
Mean	0.5692	0.5664	0.5625	0.5579	0.5559	0.5487

TABLE 5. Least significance levels obtained in the tests of hypothesis of the parameter β2 in the simulation of different cultivar loss levels.

Cultivar	Loss levels (%)					
	0	5	10	15	20	25
1	0.7140	0.6859	0.6588	0.6398	0.6095	0.6286
2	0.7370	0.7259	0.7112	0.7005	0.6685	0.6685
3	0.0860	0.1139	0.1393	0.1801	0.2240	0.2578
4	0.0090	0.0267	0.0561	0.0768	0.0946	0.1229
5	0.5300	0.5296	0.5510	0.5380	0.5315	0.5325
6	0.5830	0.5994	0.6105	0.6103	0.6233	0.6139
7	0.7660	0.7402	0.7185	0.6703	0.6514	0.6350
8	0.2120	0.2582	0.2897	0.3250	0.3514	0.3965
9	0.7870	0.7764	0.7387	0.7006	0.6687	0.6393
10	0.8500	0.8061	0.7734	0.7269	0.7117	0.6955
11	0.9850	0.8891	0.8246	0.7807	0.7372	0.7178
12	0.2880	0.3642	0.3943	0.4256	0.4534	0.4710
13	0.8970	0.8509	0.8052	0.7837	0.7594	0.7228
14	0.2400	0.2856	0.3302	0.3578	0.3929	0.4338
15	0.8660	0.8020	0.7448	0.6860	0.6414	0.5978
16	0.5560	0.5895	0.6053	0.6140	0.6141	0.6181
17	0.5540	0.5710	0.5998	0.6028	0.6094	0.5858
18	0.5910	0.6080	0.6201	0.5970	0.6186	0.5901
19	0.9430	0.8836	0.8373	0.7995	0.7649	0.7408
20	0.0530	0.0751	0.0998	0.1437	0.1904	0.2231
21	0.5450	0.5876	0.6020	0.6019	0.6063	0.5925
22	0.0810	0.1179	0.1540	0.1958	0.2431	0.2900
23	0.7210	0.7221	0.7135	0.7045	0.6807	0.6526
24	0.1890	0.2395	0.2971	0.3141	0.3696	0.4123
25	0.0680	0.0904	0.1263	0.1502	0.1825	0.2269
26	0.5690	0.5851	0.5808	0.5851	0.5833	0.5841
27	0.6020	0.5303	0.5036	0.4929	0.5035	0.5176
Mean	0.5293	0.5205	0.5217	0.5187	0.5217	0.5247

TABLE 6. Least significance levels obtained in the tests of hypothesis of the parameter β3 in the simulation of different cultivar loss levels.

Cultivar	Loss levels (%)					
	0	5	10	15	20	25
1	0.8070	0.7803	0.7616	0.7324	0.7184	0.7088
2	0.5190	0.5553	0.5674	0.5964	0.5821	0.6151
3	0.5780	0.5978	0.6013	0.5992	0.6018	0.5756
4	0.2290	0.2997	0.3554	0.3896	0.4239	0.4442
5	0.5830	0.5676	0.5638	0.5542	0.5373	0.5242
6	0.5970	0.6220	0.6252	0.6353	0.6299	0.6397
7	0.5600	0.5328	0.5310	0.5104	0.5065	0.5112
8	0.9700	0.8830	0.8362	0.7907	0.7376	0.7101
9	0.9850	0.8681	0.7882	0.7336	0.6929	0.6766
10	0.1320	0.1857	0.2239	0.2786	0.3160	0.3283
11	0.6990	0.6891	0.6668	0.6556	0.6321	0.6148
12	0.0470	0.0661	0.0956	0.1265	0.1505	0.1928
13	0.8190	0.8093	0.7864	0.7604	0.7446	0.7375
14	0.6040	0.6233	0.6363	0.6514	0.6599	0.6496
15	0.7480	0.6951	0.6493	0.6074	0.5627	0.5285
16	0.7280	0.7217	0.7224	0.7114	0.7046	0.6624
17	0.3410	0.4557	0.4503	0.4589	0.4669	0.4781
18	0.7140	0.6933	0.6866	0.6638	0.6469	0.6551
19	0.5660	0.6018	0.6200	0.6174	0.6134	0.6132
20	0.6530	0.4364	0.4345	0.4392	0.4503	0.4861
21	0.7890	0.7815	0.7707	0.7471	0.7453	0.7170
22	0.2560	0.3326	0.3660	0.4056	0.4171	0.4401
23	0.1090	0.1432	0.1768	0.2011	0.2348	0.2541
24	0.5160	0.5357	0.5556	0.5461	0.5691	0.5815
25	0.6400	0.5293	0.4833	0.4788	0.4768	0.4795
26	0.5080	0.5464	0.5422	0.5534	0.5630	0.5398
27	0.1320	0.1653	0.1956	0.2353	0.2705	0.2981
Mean	0.5492	0.5451	0.5442	0.5437	0.5428	0.5430

TABLE 7. Least significance levels obtained in the tests of hypothesis of the parameter lack of fit of the model in the simulation of different cultivar loss levels.

Cultivar	Loss levels (%)					
	0	5	10	15	20	25
1	0.2720	0.2688	0.2714	0.2707	0.2865	0.2829
2	0.1020	0.1329	0.1450	0.1618	0.1740	0.1647
3	0.4960	0.4357	0.4063	0.3877	0.3698	0.3626
4	0.4960	0.4757	0.4567	0.4276	0.4242	0.4105
5	0.0200	0.0493	0.0667	0.0978	0.1326	0.1525
6	0.0000	0.0009	0.0033	0.0096	0.0166	0.0278
7	0.0000	0.0001	0.0003	0.0058	0.0066	0.0190
8	0.0000	0.0006	0.0044	0.0107	0.0240	0.0312
9	0.0320	0.0497	0.0764	0.0957	0.1098	0.1350
10	0.0000	0.0000	0.0002	0.0021	0.0045	0.0040
11	0.0020	0.0084	0.0208	0.0340	0.0485	0.0701
12	0.4960	0.4957	0.4957	0.4953	0.4944	0.4935
13	0.0000	0.0005	0.0004	0.0021	0.0040	0.0121
14	0.0000	0.0004	0.0013	0.0046	0.0106	0.0131
15	0.0000	0.0042	0.0130	0.0218	0.0404	0.0515
16	0.0000	0.0001	0.0015	0.0033	0.0061	0.0201
17	0.0000	0.0001	0.0054	0.0082	0.0130	0.0233
18	0.1500	0.1740	0.1874	0.2061	0.2058	0.2196
19	0.0090	0.0380	0.0578	0.0855	0.1065	0.1365
20	0.4960	0.4957	0.4957	0.4957	0.4954	0.4939
21	0.0080	0.0193	0.0293	0.0439	0.0611	0.0811
22	0.4960	0.4539	0.4073	0.3800	0.3534	0.3456
23	0.0000	0.0001	0.0012	0.0024	0.0080	0.0133
24	0.2880	0.2853	0.2721	0.2775	0.2703	0.2639
25	0.0010	0.0142	0.0276	0.0438	0.0611	0.0845
26	0.0000	0.0011	0.0033	0.0070	0.0148	0.0302
27	0.0710	0.0994	0.1240	0.1482	0.1692	0.2001
Mean	0.1272	0.1298	0.1324	0.1381	0.1448	0.1534

In practical situations the researcher, analyzing a set of unbalanced data, can check initially if each cultivar is found in at least five environments at least two of which are negative and two are positive. It should be remembered that, the higher the level of loss the higher the least significance level.

CONCLUSION

It is possible to carry out phenotypic stability analysis of unbalanced experimental groups if the loss of cultivars is less than 25% and distributed homogeneously in the different environments: each cultivar should be found in at least two favorable and two unfavorable environments.

ACKNOWLEDGEMENTS

To FEPAGRO/RS for providing the data analyzed in this paper and to CNPq.

REFERENCES

CARVALHO, F.I.F.; FEDERIZZI, L.C.; NODARI, R.O.; FLOSS, E.; GANDIM, C.L. Analysis of stability parameters and of genotype x environment interaction in oats grain yield in Rio Grande do Sul (Brazil). **Revista Brasileira de Genética**, Ribeirão Preto, v.3, p.517-532, 1982.

CRUZ, C.D.; TORRES, R.A.A.; VENCOVSKY, R. An alternative approach to the stability analysis proposed by Silva and Barreto. **Revista Brasileira de Genética**, Ribeirão Preto, v.12, n.2, p.567-580, 1989.

EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science***,* Madison, v.6, p.36-40, 1966.

FAKOREDE, M.A.B.; OPEKE, B.O. Environmental indices for the analysis of genotype X environment interaction in maize. **Maydica***,* Bergamo, v.31*,* p.233-243, 1986.

FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research***,* Melbourne, v.14, p.742-754, 1963.

FRIPP, J.Y. Genotype-environments interactions in *Schizophyllum commune*. II. Assessing the environment. **Heredity***,* London*,* v.28, p.223-238, 1972.

FRIPP, J.Y.; CATEN, C.E. Genotype-environmental interactions in *Schizophyllum commune*. I. Analysis and character. **Heredity**, London, v.27, p.393- -407, 1971.

PERKINS, J.M.; JINKS, J.L. The assessment and specificity of environmental and genotype- -environmental components of variability. **Heredity**, London, v.30, n.2, p.111-126, 1973.

PIMENTEL-GOMES, F.P. **Curso de Estatística Experimental**. 11.ed. São Paulo: Nobel, 1985. 466p.

PLAISTED, R.L. A shorter method for evaluating the ability of selections to yield consistently over locations. **American Potato Journal**, Washington, DC, v.37, p.166-172, 1960.

PLAISTED, R.L.; PETERSON, L.C. A technique for evaluating the ability of selections to yield consistently in different locations or seasons. **American Potato Journal**, Washington, DC, v.36, p.381-385, 1959.

SEARLE, S.R. **Linear models***,* New York: John Wiley, 1971, 531p.

SILVA, J.G.C. da. Análise da adaptabilidade através de regressão linear segmentada**. Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.4, p.435- -448, 1995.

STORCK, L. Comparison of models for analysis of cultivar phenotypic stability. **Revista Brasileira de Genetica**, Ribeirão Preto, v.18, n.1, p.75-80, 1995.

STORCK, L.; VENCOVSKY, R. Stability analysis on a bi-segmented discontinuous model with measurement errors in the variables. **Revista Brasileira de Genética**, Ribeirão Preto, v.17, n.1, p.75-81, 1994.

VERMA, M.M.; CHAHAL, G.S.; MURTY, B.R. Limitations of conventional regression analysis: a proposed modification. **Theorethical and Applied Genetics**, Berlim, v.53, p.89-91, 1978.

WOOD, J.T. The use of environmental variables in the interpretation of genotype-environment interaction. **Heredity,** London, v.37, n.1, p.1-7, 1978.

YATES, F.; COCHRAN, W.G. The analysis of groups of experiments. **Journal Agricultural of Science**, Cambridge, v.28, p.556-580, 1938.