Rev. Brasil. Genet. 17, 1, 75-81 (1994)
(Brazil. J. Genetics)

STABILITY ANALYSIS BASED ON A BI-SEGMENTED DISCONTINUOUS MODEL WITH MEASUREMENT ERRORS IN THE VARIABLES*

Lindolfo Storck[1] and Roland Vencovsky[2]

ABSTRACT

A bi-segmented discontinuous model is proposed for stability analysis of genetic materials. The model is defined as: $\mu_{ij} = \beta 0_i + \beta 1_i \tau_j + \beta 2_i \tau_j Z_j + \beta 3_i Z_j + \delta_{ij}$, such that: τ_j is the effect of environment j; $Z_j = 1$ if $\tau_j > 0$ and $Z_j = 0$ if $\tau_j \leq 0$; $\beta 1_i$ and $\beta 1_i + \beta 2_i$ are the slopes of the two segments; $\beta 3_i$ is the discontinuity parameter for cultivar i; and δ_{ij} is the deviation from the model. This model allowed the determination and measurement of components of variance and covariance of the errors inherent to the independent and dependent variables. A correction was consequently derived for estimation of parameters and tests of hypotheses.

INTRODUCTION

In agronomic experimentation, the make up of experimental groups where several cultivars are evaluated over several years and/or locations, called environments, is a common feature. In these experiments, the cultivars-environments interaction study involves the characterization of the phenotypic behaviour of cultivars under a variety of environmental values, and has been more adequately achieved by regression analyzes. This procedure, called stability analysis is used to partition the environmental component of variation for a cultivar into regression and deviation from regression, using an environmental index as an independent variable.

The procedure of linear regression, with a logarithmic scale of data, was introduced by Finlay and Wilkinson (1963). Eberhart and Russell (1966) improved this method using the real scale of the data, and included the variance of deviations from regression as a measure of the response predictability, which is estimated by a linear regression coefficient for a given cultivar. Eberhart and Russell (1966) also commented on a disadvantage of the method, namely the use of an environmental index as an independent variable calculated from the same set of data, which would result in biased estimation and hypothesis testing.

Verma *et al.* (1978), discarded the linear model, and defined as the ideal cultivar one which presents both a relatively high yield and a stable response in poor environments as well as an ability to positively respond to environmental improvements. They proposed an alternative regression technique which consists in separately adjusting two straight line segments, one for below-mean environments and the other for above-mean environments, using a common value as a linkage point. The adjustment of just one equation consisting of two straight line segments, with union at the zero environmental index value, was proposed by Silva and Barreto (1985). However, since in this model the estimated slopes are not independent, a new approach was proposed (Cruz *et al.*, 1989) which resulted in independence of the two straight line segments. In this case we generally have a discontinuity which is estimable but was not included in the model. Storck (1989) consequently incorporated an additional discontinuity parameter in the bi-segmented model, maintaining the zero correlation between estimates of the two regression coefficients. This model, for the i^{th} cultivar is characterizated as:

---

* Part of a thesis presented by L.S. to Departamento de Matemática e Estatística, ESALQ/USP, in partial fulfillment of the requirements for the Doctoral degree.

[1] Departamento de Fitotecnia, Universidade Federal de Santa Maria, 97119-900 Santa Maria, RS, Brasil. Send correspondence to L.S.

[2] Departamento de Genética, ESALQ-USP, 13418-260 Piracicaba, SP, Brasil.

$$\mu_{ij} = \beta 0_i + \beta 1_i \tau_j + \beta 2_i \tau_j Z_j + \beta 3_i Z_j + \delta_{ij} \quad (1)$$

in which:

$$Z_j = \begin{cases} 1, & \text{if } \tau_j > 0 \\ 0, & \text{if } \tau_j \leq 0 \end{cases}$$

In this model, $\beta 0$ is a constant; $\beta 1$ measures the slope of the first regression segment; $\beta 2$ is the difference in slope between the second and the first regression segment; $\beta 3$ measures the discontinuity between both linear regression segments; and, for the ijth combination, δ_{ij} is a measure of deviation from the model. The independent variable is estimated as $\hat{\tau}_j = \overline{Y}_{.j} - \overline{Y}_{..}$. To evaluate the applicability of this model a Gompertz growth curve [$E(Y) = A^{B^{\tau}}$] was introduced to explain the performance of a given cultivar over a wide range of environmental conditions. Adjustments were made for segments of the curve and also for its entire range. In all cases efficiency of adjustment was very high (up to 99%). Proper agronomic interpretations were given by Storck (1989) for all the parameters of his extended model (1), under different types of responses.

The objective of this study were: a) to propose a bi-segmented discontinuous regression model, with measurement error structure, for stability analysis; b) to derive corrections for the stability parameter estimates and for hypothesis testing, to compensate for errors in the dependent and independent variables and for their covariances.

METHODOLOGICAL DEVELOPMENT

For a joint analysis of J experiments (environments) with I genotypes (cultivars) and K replications (blocks), the adopted model was initially:

$$Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + b_{k(j)} + e_{ijk} \quad (2)$$

Y_{ijk} is the observed response of the variable at replication k (k = 1, 2, ..., K), of cultivar i (i = 1, 2, ..., I), in environment j (j = 1, 2, ..., J).

The following restrictions and postulations were adopted:

$E(\mu) = \mu$; $E(\mu^2) = \mu^2$;

$\Sigma_{i=1}^{I} \alpha_i = 0$; $E(\alpha_i) = \alpha_i$; $E(\alpha_i^2) = \alpha_i^2$, i = 1, 2, ..., I;
$\Sigma_{j=1}^{J} \tau_j \neq 0$; $E(\tau_j) = 0$; $E(\tau_j^2) = \sigma_\tau^2$, $\forall j$.; ($\forall$ = any)
$\Sigma_{i=1}^{I} \gamma_{ij} = 0$, $\forall j$ and $\Sigma_{j=1}^{J} \gamma_{ij} \neq 0$, $\forall i$
$E(\gamma_{ij}) = 0$; $E(\gamma_{ij}^2) = ((I-1)/I)\, \sigma_\gamma^2$, $\forall ij$;
$\Sigma_{k=1}^{K} b_{k(j)} \neq 0$, $\forall j$; $\Sigma_{j=1}^{J} b_{k(j)} \neq 0$, $\forall k$
$E(b_{k(j)}) = 0$, $E(b_{k(j)}^2) = \sigma_b^2$, $\forall jk$;
$E(e_{ijk}) = 0$, $E(e_{ijk}^2) = \sigma^2$, $\forall ijk$

A solution for model (2) was obtained by the least squares method, using the common restrictions, as follows:

$\Sigma_{i=1}^{I} \hat{\alpha} = 0$; $\Sigma_{j=1}^{J} \hat{\tau}_j = 0$; $\Sigma_{i=1}^{I} \hat{\gamma}_{ij} = \Sigma_{j=1}^{J} \hat{\gamma}_{ij} = \Sigma_{i=1}^{I} \Sigma_{j=1}^{J} \hat{\gamma}_{ij} = 0$;

$\Sigma_{k=1}^{K} \hat{b}_{k(j)} = 0$, $\forall j$.

Under the conditions above, the analysis of variance table with the expected values of mean squares (MS) and F statistics, under Ho hypotheses, is shown in Table I. In this table, the F_A^{**} statistics were obtained by: $F_A^{**} = (V_2 + V_5)/(V_3 + V_4)$ which, under Ho:$\sigma_\tau^2 = 0$, has an F distribution with $g_1 = (V_2 + V_5)^2/\{V_2^2/n_2 + V_5^2/n_5\}$ and $g_2 = (V_3 + V_4)^2/\{V_3^2/n_3 + V_4^2/n_4\}$ degrees of freedom according to Satterthwaite (1946).

The hypotheses tested in Table I were, respectively, Ho(α): $\alpha_i = 0$, $\forall i$; Ho(τ): $\sigma_\tau^2 = 0$; Ho(γ): $\sigma_\gamma^2 = 0$; Ho(b): $\sigma_b^2 = 0$.

In cases in which the hypothesis Ho(γ) is rejected, the common procedure is the partitioning of interactions, through analysis of the environmental effects within each cultivar, through a new characterization of model (2):

$$Y_{ijk} + \mu + \alpha_i + \tau_{j(i)} + b_{k(j)} + e_{ijk} \quad (3)$$

in which: $\tau_{j(i)} = \tau_j + \gamma_{ij}$ is a random effect for environment j within cultivar i, with the following conditions:

$\Sigma_{j=1}^{J} \tau_{j(i)} \neq 0$, $\forall i$; $E(\tau_{j(i)}) = 0$; $E(\tau_{j(i)}^2) = \sigma_{\tau(i)}^2$, and the other conditions of model (2).

Considering the common restrictions $\Sigma_{j=1}^{J} \hat{\tau}_{j(i)} = 0$, for i = 1, 2, ..., I, estimates $\hat{\tau}_{j(i)} = \overline{Y}_{ij.} - \overline{Y}_{i..}$, were obtained, according to model (3). The degrees of freedom, MS expectations and F statistics (under Ho) are included in Table I, in which the hypotheses of interest are that the environmental variance within cultivar i is null, that is, Ho(τ_i): $\sigma_{\tau(i)}^2 = 0$. These hypotheses were tested by $F_{2(i)}^* = I\, V_{2(i)}/\{n_1 V_5 + V_4\}$ statistics with $g_1 = n_2$ and $g_2 = \{n_1 V_5 + V_4\}^2/\{n_1^2 V_5^2/n_5 + V_4^2/n_4\}$ degrees of freedom.

For cultivars in which the hypothesis Ho($\tau_{(i)}$):$\sigma_{\tau(i)}^2 = 0$ is rejected the stability analysis is proceeded through a regression analysis, using τ_j, (j = 1, 2, ..., J), as an independent variable. To make this study possible, the model (1) proposed by Storck (1989) was adopted.

Table I - Analysis of variance table for a group of experiments, with partitioning of the interaction.

Variations	DF		MS	E(MS)$^+$	F (Under Ho)
Cultivars	I-1	$= n_1$	V_1	$\sigma^2 + K\sigma^2_\gamma + JK\varphi(\alpha)$	V_1/V_3
Environments	J-1	$= n_2$	V_2	$\sigma^2 + K\sigma^2_\gamma + I\sigma^2_b + IK\sigma^2_\tau$	F_A**
C x E	(I-1) (J-1)	$= n_3$	V_3	$\sigma^2 + K\sigma^2_\gamma$	V_3/V_5
Partition of interaction and environments					
Env./Cult.1	J-1	$= n_2$	$V_{2(1)}$	$\sigma^2 + \sigma^2_b + K\sigma^2_{\tau(1)}$	$F_{2(1)}$*
Env./Cult.2	J-1	$= n_2$	$V_{2(2)}$	$\sigma^2 + \sigma^2_b + K\sigma^2_{\tau(2)}$	$F_{2(2)}$*
...	...		...	...	...
Env./Cult.I	J-1	$= n_2$	$V_{2(I)}$	$\sigma^2 + \sigma^2_b + K\sigma^2_{\tau(I)}$	$F_{2(I)}$*
Blocks/Env.	J(K-1)	$= n_4$	V_4	$\sigma^2 + I\sigma^2_b$	V_4/V_5
Error	J(K-1) (I-1)	$= n_5$	V_5	σ^2	

+: $\varphi(\alpha) = (1/(I-1)) \Sigma^I_{i=1}\alpha^2_i$; * and ** se in text.

A model with errors in the variables

Errors and model structure identification

In model (1) we observe that, for the stability analysis in a practical situation, the variables are not observable and as a consequence parameter estimators cannot be obtained. However, considering model (2), a solution was obtained in which the observable variables $\overline{Y}_{ij.} = \hat{\mu} + \hat{\alpha}_i + \hat{\tau}_j + \hat{\gamma}_{ij} = \hat{\mu}_{ij}$ is an estimator of μ_{ij}. We have then:

$$\overline{Y}_{ij.} = \mu_{ij} + \varepsilon_{ij} \qquad (4)$$

being ε_{ij} the associated error of estimator $\overline{Y}_{ij.}$.

Considering that there are fixed (μ, α) and random (τ, γ) effects in μ_{ij}, the variance of $\overline{Y}_{ij.}$ was derived conditioned to μ_{ij}. Thus, from model (2), we have:

$$Var(\overline{Y}_{ij.}|\mu_{ij}) = Var(\varepsilon_{ij}) = \sigma^2_\varepsilon = E\{\varepsilon_{ij} - E(\varepsilon_{ij})\}^2 =$$

$$= E\{(1/K)\Sigma^K_{k=1}b_{k(j)} + (1/K)\Sigma^K_{k=1}e_{ijk}\}^2, \text{ and}$$

$$\sigma^2_\varepsilon = (1/K)\{\sigma^2 + \sigma^2_b\} \qquad (5)$$

Errors ε_{ij} are considered jointly independent and an estimator of σ^2_ε can be obtained from the analysis of variance (Table I) equating MS ≅ $\hat{E}$ {MS}, such that:

$$\hat{\sigma}^2_\varepsilon = (1/IK)\{V_4 + (I-1)V_5\} \qquad (6)$$

with $n\varepsilon = (V_4 + n_1V_5)^2/(V_4^2/n_4 + n_1V_5^2/n_4)$ degrees of freedom.

Substituting (1) in (4) we get:

$$\overline{Y}_{ij.} = \beta 0_i + \beta 1_i\tau_j + \beta 2_i\tau_jZ_j + \beta 3_iZ_j + \delta_{ij} + \varepsilon_{ij} \qquad (7)$$

Model (7) is still without a solution, since the independent variable τ_j is not observable. However, from model (2) an estimator for τ_j was derived as $\hat{\tau}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$. Therefore:

$$\hat{\tau}_j = \tau_j + \upsilon_j \qquad (8)$$

in which υ_j is the estimator ($\hat{\tau}_j$) associated error. If variables υ_j are uncorrelated and also independent of τ_j then, $\sigma^2_{\hat{\tau}} = \sigma^2_\tau + \sigma^2_\upsilon$.

The variance of $\hat{\tau}_j$ conditioned to μ_{ij} was derived as being:

$Var(\hat{\tau}_j \mid \mu_{ij}) = Var(\upsilon_j) = \sigma^2_\upsilon = E\{\upsilon_j - E(\upsilon_j)\}^2$. Considering model (2)

$$\upsilon_j = \frac{1}{K}\Sigma_{k=1}^{K} b_{k(j)} - \frac{1}{JK}\Sigma_{j=1}^{J}\Sigma_{k=1}^{K} b_{k(j)} + \frac{1}{IK}\Sigma_{i=1}^{I}\Sigma_{k=1}^{K} e_{ijk} +$$

$$-\frac{1}{IJK}\Sigma_{i=1}^{I}\Sigma_{j=1}^{J}\Sigma_{k=1}^{K} e_{ijk} = \bar{b}_{.j} - \bar{b}_{..} + \bar{e}_{.j.} - \bar{e}_{...},$$

and thus,

$$\sigma_\upsilon^2 = \frac{J-1}{IJK}\left\{\sigma^2 + I\sigma_b^2\right\} \qquad (9)$$

Equating MS ≅ $\hat{E}\{MS\}$, in the analysis of variance (Table I), the following estimator can be obtained

$$\hat{\sigma}_\upsilon^2 = \frac{J-1}{IJK} V_4 \qquad (10)$$

with n_4 degrees of freedom.

Using (7) and (8) one can write the functional and structural relations (Fuller, 1987) of the model as being:

$$\bar{Y}_{ij.} = \beta 0_i + \beta 1_i \tau_j + \beta 2_i \tau_j Z_j + \beta 3_i Z_j + \delta_{ij} + \varepsilon_{ij} \qquad (11a)$$

$$\hat{\tau}_j = \tau_j + \upsilon_j \qquad (11b)$$

$$\hat{\tau}_j Z_j = \tau_j Z_j + \upsilon_j Z_j \qquad (11c)$$

$$Z_j = \begin{cases} 1, & \text{if } \hat{\tau}_j > 0 \\ 0, & \text{if } \hat{\tau}_j \le 0 \end{cases} \qquad (11d)$$

We assume that deviations δ_{ij} are independent for any pair ij and that they are homogeneous over j within i. We admit that $E(\delta_{ij}) = 0$ and $E(\delta_{ij}^2) = \sigma_{\delta i}^2$. Also δ_{ij} is independent of ε_{ij} and υ_j, but ε_{ij} is not independent of υ_j because:

$$Cov(\varepsilon_{ij}; \upsilon_j) = E\{\varepsilon_{ij}\,\upsilon_j\} = \sigma_{\varepsilon\upsilon} = \sigma_\upsilon^2 \qquad (12)$$

$$Cov(\varepsilon_{ij}; \upsilon_j Z_j) = E\{\varepsilon_{ij}\,\upsilon_j Z_j\} = \sigma_{\varepsilon,\upsilon Z} = p\,\sigma_\upsilon^2 \qquad (13)$$

in which: $p = (1/J)\Sigma_{j=1}^{J} Z_j$. And also,

$$Cov(\upsilon_j; \upsilon_j Z_j) = E\{\upsilon_j\,\upsilon_j Z_j\} = \sigma_{\upsilon,\upsilon z} = p\,\sigma_\upsilon^2 \qquad (14)$$

$$Var(\upsilon_j Z_j) = E\{\upsilon_j Z_j\}^2 = \sigma_{\upsilon z}^2 = p\,\sigma_\upsilon^2 \qquad (15)$$

Estimation

In the functional and structural relations (11), we may obtain estimators of parameters $\beta 0_i$, $\beta 1_i$, $\beta 2_i$, $\beta 3_i$ and $\sigma_{\delta i}^2$ by the method of moments (Mood *et al.*, 1974), if parameters σ_ε^2, σ_υ^2, $\sigma_{\varepsilon\upsilon}$, $\sigma_{\varepsilon,\upsilon z}$, $\sigma_{\upsilon,\upsilon z}$ and $\sigma_{\upsilon z}^2$, obtained, respectively, through (6), (10), (12), (13), (14) and (15), are previously known or estimated.

Let us consider $A_i = {}_j[\underline{\bar{Y}_i}\ ;\ \underline{\hat{\tau}}\ ;\ \underline{\hat{\tau}Z}\ ;\ \underline{Z}]_4$ as the matrix of observable values and constants relative to the proposed model (11) for cultivar i. The first sample moment of A_i is the first moment of the vector contained by the matrix, and is given by:

$$\bar{A}_i = {}_1[\,\bar{Y}_i\,;\,0\,;\,\overline{\tau z}\,;\,p\,]_4 \qquad (16)$$

in which:

$$\bar{Y}_i = \left\{\Sigma_{j=1}^{J}\bar{Y}_{ij.}\right\}/J;\quad 0 = \Sigma_{j=1}^{J}\hat{\tau}_j = 0;\quad \overline{\tau z} = \left\{\Sigma_{j=1}^{J}\hat{\tau}_j Z_j\right\}/J.$$

The first population moment of model (11) was obtained as being:

$$E\{A_i\} = \mu_{A_i} = [\beta 0_i + \beta 1_i\mu_\tau + \beta 2_i\mu_{\tau z} + \beta 3_i\mu_z;\ \mu_\tau;\ \mu_{\tau z};\ \mu_z], \qquad (17)$$

from the right hand side of the equations, with:

$$\mu_\tau = \Sigma_{j=1}^{J}\tau_j/J;\quad \mu_{\tau z} = \Sigma_{j=1}^{J}\tau_j Z_j/J;\ \text{and}\ \ \mu_z = \Sigma_{j=1}^{J} Z_j/J.$$

The unbiased sample variance-covariance matrix is characterized by:

$$S(A_i) = \begin{bmatrix} S^2(\bar{Y}_i) & S(\bar{Y}_i,\hat{\tau}) & S(\bar{Y}_i,\hat{\tau}Z) & S(\bar{Y}_i,Z) \\ & S^2(\hat{\tau}) & S(\hat{\tau},\hat{\tau}Z) & S(\hat{\tau},Z) \\ & & S^2(\hat{\tau}Z) & S(\hat{\tau}Z,Z) \\ \text{(symmetric)} & & & S^2(Z) \end{bmatrix}, \qquad (18)$$

where

$$S^2(\bar{Y}_i) = \{\Sigma_{j=1}^{J}\bar{Y}_{ij.}^2 - J\bar{Y}_{i..}^2\}/(J-1);$$

$$S(\bar{Y}_i,\hat{\tau}) = \Sigma_{j=1}^{J}\bar{Y}_{ij.}\,\hat{\tau}_j/(J-1);\quad S^2(\hat{\tau}) = \{\Sigma_{j=1}^{J}\hat{\tau}_j^2\}/(J-1);$$

$$S(\bar{Y}_i,\hat{\tau}Z) = \{\Sigma_{j=1}^{J}\bar{Y}_{ij.}\hat{\tau}_j Z_j - J\bar{Y}_{i..}\overline{\tau z}\}/(J-1);$$

$$S(\bar{Y}_i,Z) = \{\Sigma_{j=1}^{J}\bar{Y}_{ij.}Z_j - pJ\bar{Y}_{i..}\}/(J-1);$$

$$S(\hat{\tau},\hat{\tau}Z) = \{\Sigma_{j=1}^{J}\hat{\tau}_j^2 Z_j\}/(J-1);\quad S(\hat{\tau},Z) = \{\Sigma_{j=1}^{J}\hat{\tau}_j Z_j\}/(J-1);$$

$$S^2(\hat{\tau}Z) = \{\Sigma_{j=1}^{J}\hat{\tau}_j^2 Z_j - J\overline{\tau z}^2\}/(J-1);$$

$$S(\hat{\tau}Z,Z) = \{\Sigma_{j=1}^{J}\hat{\tau}_j Z_j - pJ\overline{\tau z}\}/(J-1);\quad S^2(Z) = Jp(1-p)/(J-1).$$

Here $\Sigma_{(A_i)}$ is the first population variance-covariance matrix obtained from the right hand side of equalities of model (11), being characterized by:

$$\Sigma_{(A_i)} = \begin{bmatrix} \sigma_{\bar{Y}_i}^2 & \sigma_{\bar{Y}_i,\tau} & \sigma_{\bar{Y}_i,\tau z} & \sigma_{\bar{Y}_i,z} \\ & \sigma_{\hat{\tau}}^2 & \sigma_{\hat{\tau},\hat{\tau}z} & \sigma_{\hat{\tau},z} \\ & & \sigma_{\hat{\tau}z}^2 & \sigma_{\hat{\tau}z,z} \\ \text{(symmetric)} & & & \sigma_z^2 \end{bmatrix}, \qquad (19)$$

with

$\sigma^2_{\bar{Y}i} = \beta 1_i^2 \sigma^2_\tau + \beta 2_i^2 \sigma^2_{\tau z} + \beta 1_i \beta 2_i\ \sigma_{\tau,\tau z} + p(1\text{-}p)\ \beta 3_i^2 + \sigma^2_{\delta i} + \sigma^2_\varepsilon$;

$\sigma_{\bar{Y}i,\tau} = \beta 1_i\ \sigma^2_\tau + p\ \beta 2_i\ \sigma_{\tau,\tau z} + p\ \beta 3_i\ \sigma_{z,\tau} + \sigma_{\varepsilon\upsilon}$;

$\sigma_{\bar{Y}i,\tau z} = p\beta 1_i \sigma_{\tau,\tau z} + p\ \beta 2_i \sigma^2_{\tau z} + p\beta 3_i \sigma_{z,\tau z} + \sigma_{\varepsilon,\upsilon z}$;

$\sigma_{\bar{Y}i,z} = p\beta 1_i \sigma_{\tau,z} + p\beta 2_i \sigma_{\tau z,z} + p(1\text{-}p)\beta 3_i$;

$\sigma^2_T = \sigma^2_\tau + \sigma^2_\upsilon$; $\sigma_{T,TZ} = p\ \sigma_{\tau,\tau z} + \sigma_{\upsilon,\upsilon z}$;

$\sigma_{T,z} = p\ \sigma_{\tau,z}$; $\sigma^2_{TZ} = p\ \sigma^2_{\tau z} + \sigma^2_{\upsilon z}$;

$\sigma_{TZ,z} = p\ \sigma_{\tau z,z}$; and $\sigma^2_Z = p(1\text{-}p)$.

The parameter estimators of model (11) are obtained by equating $\bar{A}_i$ (16) with $\hat{E}\{A_i\}$ (17) and $S_{(Ai)}$ (18) with $\hat{\Sigma}_{(Ai)}$ (19). We have then:

$$\hat{\beta}_i = \{\hat{\beta}1_i\ \hat{\beta}2_i\ \hat{\beta}3_i\} = [Mxx - S(\upsilon\upsilon)]^{-1}\ [Mxy_i - S(\varepsilon\upsilon)] \qquad (20)$$

$$\hat{\beta}0_i = \bar{Y}_{i..} - \hat{\beta}2_i\overline{\tau z} - p\ \hat{\beta}3_i \qquad (21)$$

$$\hat{\sigma}^2_{\delta i} = Svv_i - \hat{\beta}_i'\ S(\upsilon\upsilon)\ \hat{\beta}_i + 2\ S(\upsilon\varepsilon)\ \hat{\beta}_i - \hat{\sigma}^2_\varepsilon \qquad (22)$$

where, from (20) and (22):

$$Mxx = \begin{bmatrix} S^2(\hat{\tau}) & S(\hat{\tau},\hat{\tau}Z) & S(\hat{\tau},Z) \\ & S^2(\hat{\tau}Z) & S(\hat{\tau}Z,Z) \\ (\text{symmetric}) & & S^2(Z) \end{bmatrix} ;$$

$$S(\upsilon\upsilon) = \begin{bmatrix} \hat{\sigma}^2_\upsilon & p\hat{\sigma}^2_\upsilon & 0 \\ p\hat{\sigma}^2_\upsilon & p\hat{\sigma}^2_\upsilon & 0 \\ 0 & 0 & 0 \end{bmatrix} ;$$

$M'_{xyi} = [S(\bar{Y}i,\hat{\tau})\ \ S(\bar{Y}i,\hat{\tau}Z)\ \ S(\bar{Y}i,Z)]$; ,

$S'_{(\varepsilon\upsilon)} = S_{(\upsilon\varepsilon)} = [\ \hat{\sigma}^2_\upsilon\ \ p\ \ \hat{\sigma}^2_\upsilon\ \ 0\ ]$, and

$$Svv_i = \Sigma^J_{j=1}\ \left\{\bar{Y}_{ij.} - \hat{\beta}0_i - \hat{\beta}1_i\hat{\tau}_j - \hat{\beta}2_i\hat{\tau}_j Z_j - \hat{\beta}3_i Z_j\right\}^2 / (J-4) \qquad (23)$$

The "Environments/Cultivar$_i$" sources of variation, are partitioned on a plot level as shown in Table II.

Table II - Analysis of variance with the environments within cultivar i source of variation partitioned into regression and deviations components.

Variations	DF	E(MS)*
Env.'s/Cult$_i$	J-1	$\sigma^2 + \sigma^2_\upsilon + K\sigma^2_{\tau(i)}$
Regression	3	$\sigma^2 + \sigma^2_\upsilon + K\sigma^2_{\delta i} + KC_i + K\varphi\{\beta\}_i$
Deviations	J-4	$\sigma^2 + \sigma^2_\upsilon + K\sigma^2_{\delta i} + KC_i$

* $C_i = \beta_i'\ E\{S(\upsilon\upsilon)\}\ \beta_i - 2\ E\{S(\upsilon\varepsilon)\}\ \beta_i$;

$\varphi\{\beta\}_i = (1/3)\ \{SS_{(\beta 0,\ \beta 1,\ \beta 2,\ \beta 3\ |\ \text{cult.}_i\ \text{mean})}$

The number of degrees of freedom of the variance estimate $\hat{\sigma}^2_{\delta i}$ are given, according to Searle (1971), by

$$\hat{f}0_i = \left\{\hat{\Pi}_i - 1\right\}^2 / \left\{\hat{\Pi}_i/f1 + 1/f2_i\right\}^2 \qquad (24)$$

with:

$\hat{\Pi}_i = M1_i/(M2_i + M3_i + M4)$; $f1$ = J-4;

$$f2_i = \frac{(M2_i + M3_i + M4)^2}{M2_i^2/n_4 + M3_i^2/n_4 + M4^2/f4} ;$$

$$f4 = \frac{(V_4 + n_1 V_5)^2}{V_4^2/n_4 + n_1 V_5^2/n_4} ;$$

$M1_i = Svv_i$; $M2_i = \hat{\beta}1_i\ \hat{\sigma}^2_\upsilon\ (\hat{\beta}1_i + p\ \hat{\beta}2_i - 2)$;

$M4 = \hat{\sigma}^2_\varepsilon$; $M3_i = p\ \hat{\beta}2_i\ \hat{\sigma}^2_\upsilon\ (\hat{\beta}1_i + \hat{\beta}2_i - 2)$.

Ignoring the correction due to error of the independent variable will be equivalent to taking $M2_i = 0$ and $M3_i = 0$.

Tests of hypotheses

The same principles of partial regression apply to testing of hypotheses with parameters β1, β2 and β3. The stability of cultivars is used in general to investigate if data fit a model chosen a priori, within a controlled margin of error (Type I error). For the determination of alternative hypotheses we consider two mutually exclusive situations, namely: the identification of superior or ideal cultivars to be selected or the identification of treatments to be discarded based on their inferiority. Hence, alternative hypotheses about these parameters should be bilateral with α levels of error and will depend on the purpose of the experiments.

If observations $\overline{Y}_{ij.}$ of model (11) are normal, independent and identically distributed, then, according to Fuller (1987), the variance-covariance matrix of the approximated distribution of $\hat{\underset{\sim}{\beta}}_i = [\hat{\beta}1_i\ \hat{\beta}2_i\ \hat{\beta}3_i]'$ is estimated by:

$$\hat{V}(\hat{\underset{\sim}{\beta}}_i) = \frac{1}{J-1} Mxx^{-1} Svv_i + \frac{1}{J-1} Mxx^{-1} [S(\upsilon\upsilon)Svv_i + \tilde{S}uv_i\, \tilde{S}vu_i]\, Mxx^{-1} +$$

$$+ \frac{1}{n_5} Mxx^{-1} [S(\upsilon\upsilon)\, Srr_i + \tilde{S}uv_i\, \tilde{S}vu_i]\, Mxx^{-1}\ , \qquad (25)$$

in which: $Mxx = Mxx - S(\upsilon\upsilon)$; $\tilde{S}uv_i = S(\varepsilon\upsilon) - S(\upsilon\upsilon)\hat{\underset{\sim}{\beta}}_i$, and

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

were obtained from (20) to (23); also Sww is composed of

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

The following hypotheses are pertinent. We give the corresponding tests and interpretation for a given cultivar i.

1) $Ho{:}\beta1_i = 1$. In this case we assume that the response is identical to that of the means of the whole set of cultivars, for the first segment of the model. The statistic

$$t_i = (\hat{\beta}1_i - \beta1_i)\ /\ \sqrt{\hat{V}(\hat{\beta}1_i)}\ , \qquad (26)$$

has a t distribution with (J-4) degrees of freedom under Ho. The alternative hypotheses are $H1{:}\beta1 < 1$ and $H1{:}\beta1 > 1$, to separate cultivars with low and high rate of response in the first segment, respectively.

2) $Ho{:}\beta2_i = 0$. Under this hypothesis the slope of the second linear segment is identical to that of the first, and corresponds to $Ho{:}\beta1_i + \beta2_i = 1$. The statistic

$$t_i = (\hat{\beta}2_i - \beta2_i)\ /\ \sqrt{\hat{V}(\hat{\beta}2_i)}\ , \qquad (27)$$

under Ho, follows a t distribution with (J-4) degrees of freedom. The alternative hypotheses are $H1{:}\beta2 < 0$ to discard inferior cultivars and $H1{:}\beta2 > 0$ to select the superior ones.

3) $Ho{:}\beta3_i = 0$ is the hypothesis under which the discontinuity between the two straight lines is null. Here also

$$t_i = (\hat{\beta}3_i - \beta3_i)\ /\ \sqrt{\hat{V}(\hat{\beta}3_i)}\ , \qquad (28)$$

has a t distribution with (J-4) degrees of freedom under Ho. As alternative hypotheses we take $H1{:}\beta3 < 0$ or $H1{:}\beta3 > 0$ to discard or select cultivars, respectively.

4) $Ho{:}\sigma^2_{\delta i} = 0$. Specifying that there are no deviations from the model this hypothesis is tested through

$$F = (K\ \hat{\sigma}^2_{\delta i} + V_5)/V_5 \qquad (29)$$

which under Ho has an F distribution with

$g_1 = \{K\ \hat{\sigma}^2_{\delta i} + V_5\}^2/\{K^2(\hat{\sigma}^2_{\delta i})^2/\hat{f}0_i + V_5^2/n_5\}$ and $g_2 = n_5$

degrees of freedom. Confidence intervals for $\sigma^2_{\delta i}$, for cultivar i, are given by

$$P\,(\hat{f}0_i\ \hat{\sigma}^2_{\delta i}\ /\ X_1^2 \le \sigma^2_{\delta i}, \le \hat{f}0_i\ \hat{\sigma}^2_{\delta i}\ /\ X_2^2\,) = 1 - \alpha \qquad (30)$$

in which X_1^2 and X_2^2 are the tabulated values of the X^2 distribution with $\hat{f}0_i$ degrees of freedom for $P[X^2 \le X_1^2\ ] = 1 - \alpha/2$ and $P[X^2 \le X_2^2\ ] = \alpha/2$.

DISCUSSION

The error of the dependent variable ε_{ij} (equation 5) is not equivalent to the term given by Eberhart and Russell (1966), which ignores the variance due to block effects. In our deviations we considered these effects as being random, and hence σ^2_ε includes σ^2_b in addition to σ^2.

Equation (9) shows that the error variance (σ^2_υ) of the independent variable (environmental index) is a function of the number of environments, replications/ environments and treatments. It also depends upon the experimental error and the blocks/environments component of variance. Derived expression of σ^2_υ indicates how this variance can be kept at low levels. When σ^2_υ is sizable but ignored, estimates of parameters of the model become biased. In fact, if we replace expressions 11b and 11c in 11a the following relation is obtained:

$$\overline{Y}_{ij.} = \beta0_i + \beta1_i(\hat{\tau}_j - \upsilon_j) + \beta2_i\,(\hat{\tau}_j - \upsilon_j)Z_j + \beta3_iZ_j + \delta_{ij} + \varepsilon_{ij}$$

$$\overline{Y}_{ij.} = \beta0_i + \beta1_i\hat{\tau}_j + \beta2_i\hat{\tau}_j\,Z_j + \beta3_iZ_j + \psi_{ij}, \text{ in which}$$

$$\psi_{ij} = \delta_{ij} + \varepsilon_{ij} - \beta1_i\upsilon_j - \beta2_i\upsilon_jZ_j$$

is the deviation from the model. Unbiased estimates of the parameters are obtained only when $\hat{\tau}_j$ and ψ_{ij} are independent. And this is not generally the case because

$$Cov(\hat{\tau}_j\,;\,\psi_{ij}) = E\{\{\hat{\tau}_j - E(\hat{\tau}_j)\}\{\psi_{ij} - E(\psi_{ij})\}\} = (1 - \beta1_i - p\beta2_i)\,\sigma^2_\upsilon.$$

There will be no bias only when $\beta1_i + p\beta2_i = 1$.

The statistic F=MSD/MSError is commonly used in stability analyses for testing the null hypothesis about deviations from the model for a given cultivar. This test is not entirely adequate since the hypothesis actually being is $Ho: \sigma^2_{\delta i} + C_i = 0$, in which

$$C_i = \{\underset{\sim}{\beta i}' E\{S(\upsilon\upsilon)\} \underset{\sim}{\beta i} - 2E\{S(\upsilon\varepsilon)\} \underset{\sim}{\beta i} + \sigma^2_\delta\}.$$

Estimates of the latter quantity may not be negligible, depending on the estimated values of $\beta 1$, $\beta 2$ and the variance due to deviations from the model. As they are a test for $Ho: \sigma^2_{\delta i} = 0$, interpretations of results will be misleading and not in complete agreement with the estimated value of the corresponding coefficient of determination (R^2). This occurs because R^2, calculated as $SSRegr_i/(SSEnv./C_i)$, for cultivar i, is overestimated since, as shown in Table II, E(SSRegr) contains the additional component $3KC_i$. Hence, unless R^2 is corrected, the appropriate way to evaluate the adjustment of data to the model should be the corrected estimate of $\sigma^2_{\delta i}$ and its corresponding test criterion (equation 29).

In expressions (20) and (22) the corrections due to errors of the variables are given by $S(\upsilon\upsilon)$ and $S(\varepsilon\upsilon)$. We observe that if these quantities are negligible or null, estimators of parameters $\beta 0$, $\beta 1$, $\beta 2$ and $\beta 3$ correspond to the least squares solution under the assumption of no error in the independent variable. This fact is relevant to the investigation of the importance of errors in this variable in parameter estimation.

The Software developed for the analysis proposed here is available under request from the main author.

ACKNOWLEDGMENTS

Publication supported by FAPESP.

RESUMO

Neste estudo foi proposto o modelo bi-segmentado descontínuo para ser utilizado nas análises de estabilidade de materiais genéticos. Este modelo foi definido como sendo: $\mu_{ij} = \beta 0_i + \beta 1_i \tau_j + \beta 2_i \tau_j Z_j + \beta 3_i Z_j + \delta_{ij}$, em que: τ_j é o efeito do ambiente j; $Z_j = 1$ se $\tau_j > 0$ e $Z_j = 0$ se $\tau_j \leq 0$; $\beta 1_i$ e $\beta 1_i + \beta 2_i$ são os coeficientes angulares dos dois segmentos e $\beta 3_i$ é o parâmetro de descontinuidade do cultivar i; e δ_{ij} é o desvio do modelo. Considerando este modelo, foram identificados e mensurados os componentes de variância e covariância dos erros das variáveis independentes e dependente e, por isto, foi deduzida uma correção para a estimação dos parâmetros e para os testes de hipóteses.

REFERENCES

Cruz, C.D., Torres, R.A.A. and **Vencovsky, R.** (1989). An alternative approach to the stability analysis proposed by Silva and Barreto. *Rev. Brasil. Genet. 12*: 567-580.

Eberhart, S.A. and **Russell, W.A.** (1966). Stability parameters for comparing varieties. *Crop Science 6*: 36-40.

Finlay, K.W. and **Wilkinson, G.N.** (1963). The analysis of adaptation in a plant-breeding programme. *Austr. J. Agr. Res. 14*: 742-754.

Fuller, W.A. (1987). *Measurement Error Models*. John Wiley, New York, pp. 450.

Mood, A.M., Graybill, F.A. and **Boes, D.C.** (1974). *Introduction to the Theory of Statistics*. 3rd edn. McGraw-Hill, New York, pp. 564.

Satterthwaite, F.C. (1946). An approximate distribution of estimates of variance components. *Biometrics 2*: 110-114.

Searle, S.R. (1971). *Linear Models*. John Wiley, New York, pp. 531

Silva, J.G.C. and **Barreto, J.N.** (1985). Aplicação da regressão linear segmentada em estudos da interação genótipo x ambiente. In: *Proceedings of Simpósio de Experimentação Agrícola,* 1, Piracicaba, 1985. Campinas, Fundação Cargill, pp. 49-50.

Storck, L. (1989). *Modelo de regressão bi-segmentado descontínuo com erros de medida aplicado na análise de estabilidade de cultivares*. Doctoral Thesis, ESALQ-USP, Piracicaba, SP.

Verma, M.M., Chahal, G.S. and **Murty, B.R.** (1978). Limitations of conventional regression analysis, a proposed modification. *Theoretical and Applied Genetics 53*: 89-91.

(Received December 13, 1993)