

Stability Analysis: Where Do We Stand?¹

C. S. Lin, M. R. Binns, and L. P. Lefkovich²

ABSTRACT

To clarify the apparent confusion arising from the diversity of published stability statistics, and the relationship of these with the clustering of genotypes for similarity of response to environments, the interrelationship of nine stability statistics and nine similarity measures are investigated. The stability statistics fall into four groups depending on whether they are based on the deviations from the average genotype effect or on the genotype \times environment (GE) term, and whether or not they incorporate a regression model on an environmental index. These groups of stability statistics are shown to be related to three concepts: A genotype may be considered to be stable (i) if its among-environment variance is small, (ii) if its response to environments is parallel to the mean response of all genotypes in the trial, or (iii) if the residual mean square from a regression model on the environmental index is small. Unfortunately, these three concepts represent different aspects of stability and do not always provide a complete picture of the response. In the alternative approach of cluster analysis, the similarity measures define complete similarity in three different ways: i) equality of genotype's response across locations, ii) equality of all within location differences, and iii) equality of all within location ratios. The advantage of the non-parametric approach is that a cultivar's response characteristics can be assessed qualitatively, without the need for a mathematical characterization.

Additional index words: Genotype \times environment interaction, Cluster analysis, Cultivar assessment, Stability parameter.

INCREASED concern with the importance of homeostasis in living organisms has stimulated plant breeders' awareness for the need to develop well-buffered cultivars. This has led to a greater emphasis on phenotypic stability in breeding programs. However, the concept of stability is by no means unambiguous; it is defined in many ways depending on how the scientist wishes to look at the problem, while the statistics that parameterize these various concepts are also numerous. This leads many scientists to wonder which stability statistics should be used for their particular problem. Empirical comparisons of several methods have been reported (e.g., Gray, 1982; Kang and Miller, 1984; Ntare and Aken'Ova, 1985). Although such studies provide interesting information, the relationships among the methods can be explicitly explored only if comparisons are based on the structure of the stability statistics themselves. Several authors (e.g., Easton and Clements, 1973) have attempted to do this, but without identifying the basic structures from which the methods were derived.

The purpose of the present paper is to examine nine stability statistics currently in use, to show that they are derived from two components of a two-way classification of the data, and that there are three types of stability concepts. These concepts are discussed both from statistical and biological points of view. Some limitations of the parametric approach to stability are noted, and as an alternative, a nonparametric approach, namely to group genotypes according to their

similarity of response to a range of environments, is discussed. Several published similarity coefficients used in cluster analyses are summarized and their interrelationships are investigated.

STABILITY STATISTICS

For convenience, a two-way model is assumed. Let x_{ij} represent the observed mean value of genotype i ($i = 1, \dots, p$) in environment j ($j = 1, \dots, q$), and let $\bar{x}_{i.}$, $\bar{x}_{.j}$, and \bar{x} denote the marginal means of genotype i , environment j , and the overall mean, respectively. The nine stability statistics most frequently cited are given, with their formulae, in Table 1; a brief description of each follows.

(1) The variance of a genotype across environments (S_i^2), can be a measure of stability.

(2) Coefficient of variability (CV_i). Francis and Kanenberg (1978) used the conventional CV% of each genotype as a stability measure.

(3) Plaisted and Peterson's (1959) mean variance component for pairwise GE interaction ($\bar{\theta}$). The mean of the estimated variance components of the GE interaction for all pairs of genotypes that include genotype i is the stability measure for genotype i .

(4) Plaisted's (1960) variance component for GE interaction ($\theta_{(i)}$). One genotype (i) is deleted from the entire set of data and the GE interaction variance from this subset is the stability index for genotype i .

(5) Wricke's (1962) ecovalence (W_i^2). The GE interaction effects for genotype i , squared and summed across all environments, is the stability measure for genotype i .

(6) Shukla's (1972a) stability variance (σ_i^2). Based on the residuals in a two-way classification, the variance of a genotype across environments is the stability measure.

(7) Finlay and Wilkinson's (1963) regression coefficient (b_i). The observed values are regressed on environmental indices defined as the difference between the marginal mean of the environments and the overall mean. The regression coefficient for each genotype is then taken as its stability parameter.

(8) Perkins and Jinks' (1968) regression coefficient (β_i). Similar to (7) except that the observed values are adjusted for location effects before the regression.

(9) Eberhart and Russell's (1966) deviation parameter (δ_i^2). The residual mean square (MS) of deviation from the regression defined in (7) or (8) is the measure of stability for each genotype.

Other published stability statistics include Hanson's (1970) relative and comparative genetic stabilities, which will be considered in connection with clustering methods, and Tai's (1971) α_i and λ_i , which can be regarded as a special form of that of Eberhart and Russell (1966), when the environmental index is assumed to be random.

Note that Plaisted and Peterson's (1959), and Plaisted's (1960) statistics were originally defined in terms of replicated data, while in this paper the equations

¹ Contribution no. I-770 from the Engineering and Statistical Res. Centre, Res. Branch, Agric. Canada, Ottawa, Canada, K1A 0C6 Received 14 Nov. 1985.

² Research scientists, Engineering and Statistical Res. Centre, Agric. Canada.

Table 1. A summary of equations for the nine stability statistics.

| Group | Type | Equation | Authors or users |
|-------|------|--|-------------------------------|
| A | 1 | $S_i^2 = \sum_{j=1}^q (X_{ij} - \bar{X}_i)^2 / (q - 1)$ | Francis and Kannenberg (1978) |
| | 1 | $CV_i = S_i / \bar{X}_i \times 100$ | |
| B | 2 | $\theta_i = \frac{p}{2(p-1)(q-1)} \sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2 + \frac{SS(GE)\dagger}{2(p-1)(q-1)}$ | Plaisted and Peterson (1959) |
| | 2 | $\theta_{(i)} = \frac{-p}{(p-1)(p-2)(q-1)} \sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2 + \frac{SS(GE)}{(p-2)(q-1)}$ | Plaisted (1960) |
| | 2 | $W_i^2 = \sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2$ | Wricke (1962) |
| | 2 | $\sigma_i^2 = \frac{p}{(p-2)(q-1)} \sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2 - \frac{SS(GE)}{(p-1)(p-2)(q-1)}$ | Shukla (1972a) |
| C | 2 | $b_i = \sum_{j=1}^q (X_{ij} - \bar{X}_i)(\bar{X}_j - \bar{X}_{..}) / \sum_{j=1}^q (\bar{X}_j - \bar{X}_{..})^2$ | Finlay and Wilkinson (19563) |
| | 2 | $\beta_i = \sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})(\bar{X}_j - \bar{X}_{..}) / \sum_{j=1}^q (\bar{X}_j - \bar{X}_{..})^2$ | Perkins and Jinks (1968) |
| D | 3 | $\delta_i^2 = \frac{1}{(q-2)} \left[\sum_{j=1}^q (X_{ij} - \bar{X}_i)^2 - \beta_i^2 \sum_{j=1}^q (\bar{X}_j - \bar{X}_{..})^2 \right]$ | Eberhart and Russell (1966) |
| | | $\delta_i^2 = \frac{1}{(q-2)} \left[\sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2 - \beta_i^2 \sum_{j=1}^q (\bar{X}_j - \bar{X}_{..})^2 \right]$ | Perkins and Jinks (1968) |

† SS(GE) = $\sum_{i=1}^p \sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2$

are based on the cell means. The derivations of equations for $\bar{\theta}_i$ and $\theta_{(i)}$ are shown in Appendix. The formula used by Shukla (1972a) had previously been shown to be statistically equivalent to Grubbs' (1948) measurement error for calibration studies (Gaylor, 1956).

All of the formulae in Table 1 are based either on the deviation from the average genotype effect (DG), $(X_{ij} - \bar{X}_i)$, or on the GE interaction term, $(X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})$. Some are sums of squares (SS), others are regression coefficients or deviation MS from regression. They can thus be classified into four groups (A, B, C, D):

| | |
|--------------------|------------------------|
| Group A : DG | SS |
| Group B : GE | SS |
| Group C : DG or GE | regression coefficient |
| Group D : DG or GE | regression deviation |

The difference between the two statistics in Group A can be explained in terms of the logarithmic transformation. For small CV% (certainly good for CV < 20%), variance $[\log(y)] = \text{variance}(y) / [\text{mean}(y)]^2$ approximately. Therefore these two statistics are equivalent except for data preparation; analysis of proportional changes is equivalent to analysis of logarithms (for caution in using transformations, see Knight, 1970).

For ranking purposes, the four statistics in Group B are equivalent, because their second terms are constant and their first terms are proportional to each other. However, while $\bar{\theta}_i$, $\theta_{(i)}$, and W_i^2 are merely index numbers, σ_i^2 is an unbiased estimate of the variance for genotype i . Although the distribution of σ_i^2 is not clear, homogeneity of the estimates can be tested using Shukla's (1972b) approximate test.

The two statistics in Group C are equivalent ($\beta_i = b_i - 1$), and the two statistics in Group D are identical.

When the variability of the two components, $(X_{ij} - \bar{X}_i)$ and $(X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})$, can be expressed by a regression model, the regression coefficient can be conveniently used as a stability parameter. In these circumstances, the statistics of Group C are to be preferred over those of Group B because they give the shape of the response as well as its variation. A proper use of these two statistics requires that:

- (i) a large part of total variation is accounted for by the regression (high R_i^2 , where R_i is the multiple correlation coefficient); and
- (ii) the deviation MS from the regression are homogeneous across all genotypes.

If R_i^2 is small, the statistic b_i may not be truly representative of the response, while if the variances are heterogeneous, the estimates of b_i will have different precisions, making comparisons among the b_i 's tedious. Eberhart and Russell (1966) suggested using δ_i^2 as a second stability parameter (in addition to b_i or β_i), but reconciling two parameters for a single character is difficult. The inappropriateness of δ_i^2 as a stability parameter is discussed below.

THREE CONCEPTS OF STABILITY

The statistics investigated in the previous section represent the three different concepts of stability:

Type 1: A genotype is considered to be stable if its among-environment variance is small.

Type 2: A genotype is considered to be stable if its response to environments is parallel to the mean response of all genotypes in the trial.

Type 3: A genotype is considered to be stable if the residual MS from the regression model on the environmental index is small.

The statistics of Group A (Table 1) corresponds to Type 1 stability, those of Group B to Type 2, and those of Group D to Type 3. Group C can be interpreted as either Type 1 or Type 2, depending on how a standard stable genotype is defined. If stable genotypes are defined by having $b_i=1$ (Finlay and Wilkinson, 1963) or $\beta_i=0$ (Perkins and Jinks, 1968), Type 2 is implied, but if they are defined by $b_i=0$ ($\beta_i=-1$), Type 1 is implied.

Among these three types of stability, Type 1 is analogous to the concept of homeostasis. Becker (1981) termed this type of stability of biological concept and distinguished it from an agronomic concept of stability, which here is equivalent to Type 2. Although Type 1 is theoretically sound, breeders have used it infrequently. One reason seems to be that a breeder would like to find cultivars not only with good Type 1 stability but also with high yield. Type 1 stability is often associated with a relatively poor response and low yield in environments that are high yielding for other cultivars. In fact, b_i and yield are often positively correlated (Finlay and Wilkinson, 1963; see also Gray, 1982). Another reason is that although a high level of performance under a wide range of environments may be desirable, it is difficult to achieve in practice. Even if it can be achieved, the effort is not entirely necessary because several less widely adapted cultivars can be bred and then grown separately in different environments to achieve maximum production. Thus, the usefulness of Type 1 stability depends very much on the range of environments under which the experiment is conducted. If the range is very large, such as a collection of sites from a continental area, Type 1 stability may not be very meaningful, but if the geographical range can be restricted, then it could be important.

Type 2 stability is a relative measure depending on the genotypes included in the test so its scope of inference is confined to the test set and should not be generalized. Because the mean of all genotypes is used as standard response in each environment, drawing an inference from this type of stability requires extreme caution, unless the genotypes are a representative sample of those grown in the area in question. A genotype stable by this definition is so only with respect to the other genotypes in the test, without any assurance that it will appear stable if assessed with another set of genotypes. For example, in the set of genotypes (A, B, C, D, E), A may be assessed stable and B is unstable if A resembles C, D, E more closely than does B. However, in the set of genotypes (A, B, F, G, H), A may be considered unstable and B stable if A is less like F, G, H than is B.

Type 3 stability is a relatively new concept. Since Eberhart and Russell (1966) suggested δ_i^2 as a second parameter [Tai's (1971) second parameter λ_i can be included here], great attention has been focused on it. Breese (1969) strongly advocated its use for the following reasons. The variability of any genotype with respect to environment can be subdivided into a predictable part corresponding to regression and an unpredictable part corresponding to deviation MS. Be-

cause the regression part can be predicted and to some extent controlled (by selecting specific genotypes for specific locations), it is no longer profitable to consider this component of GE interaction as a measure of stability. He suggested that "The term 'stability' should now rather be reserved to describe measurements of unpredictable irregularities in the response to environment as provided by the deviation from regression (p. 33)".

Although this argument is appealing and has received a wide acceptance (as shown by numerous publications using this statistic), it is doubtful if the deviation MS from regression does represent the genotype's stability character. The main reason is that the regression model for GE interaction is a descriptive model based on the data being analysed, and not a prediction model as the above argument seems to assume. For a useful prediction model, the independent variable must be measurable prior to the experiment, and then the deviation MS from regression may have a deterministic property that can be associated with genotypes. However, for the linear descriptive model considered here, the independent variable (environmental index) cannot be measured prior to the experiment: It is no more than a data based device to represent the environment so that the genotypes' responses can be studied quantitatively. Because the model is purely empirical, the deviation MS for it does not have a deterministic property such as may be the case for a prediction model. Essentially, the deviation MS of this model indicates no more than how good is the fit, but has no direct bearing on the genotype's stability. A poor fit (i.e., R_i^2 small or δ_i^2 large), or heterogeneous deviation MS should be taken as an indication that the use of the regression model to estimate stability is not adequate and that other approaches should be investigated.

In all, among these three types of stability concept, Type 3 is the least attractive. Until such time as the environmental index can be replaced by actual environmental factors, such as temperature or rainfall, etc. (leading to a prediction model), the use of Type 3 is difficult to justify. Type 2 stability is useful for comparing a specific set of genotypes, but by being a relative measure, it does not have a sufficiently broad inferential base for general assessment. In contrast, Type 1 has a broad inferential base, because its stability definition does not depend on the other genotypes included in the test and is thus unambiguous. However, it does not provide information on the response pattern over the range of test environments that is so vital for cultivar recommendations.

CLASSIFYING GENOTYPES

The parametric approach described above gives only the individual aspects (Types 1, 2, 3) of stability but cannot provide an overall picture of the response. For example, a genotype may be assessed as stable in the Type 1 sense but could be found to be unstable in the Type 2 sense; or it may be judged as Type 3 unstable but simultaneously Type 2 stable, etc. It is difficult to reconcile all of these assessments into a unified conclusion. The basic reason for the difficulty is that a genotype's response to environments is multivariate,

yet the parametric approach tries to transform it to a univariate problem via a stability index. To escape from this dilemma, a different line of thought has emerged, namely to cluster genotypes according to their response structure. This represents a shift from ranking stability by a quantitative measure to assigning genotypes into qualitatively homogeneous stability subsets, i.e., to form a classification of them. The advantage of a classification, which we regard as the result of a nonparametric method, is that although genotypes are grouped based on a specific data set, the relative relationships among genotypes can be independent of any specific set of data analyzed. Using the example from the previous section, a cluster analysis may give the groups (B) and (A C D E) for the first set, (A) and (B F G H) for the second, and (A C D E) and (B F G H) for the combined set. In either case, Genotypes A and B are in different groups. The implication is that grouping genotypes by GE interaction structure can preserve the information of Type 2 stability, but without the inferential limitation of Type 2 stability statistics.

Several methods for clustering genotypes (or environments) based on similarity of response characteristics have appeared in the literature and are sum-

marized in Table 2. The comparison of cluster methods usually involves two aspects: One is the definition of similarity (or dissimilarity) measure, and the other is the algorithm that groups objects. Because the present interest is in the interrelationship between the parametric and nonparametric characterization of responses, attention is restricted to measures of pairwise resemblance. For convenience, all similarity measures are defined in terms of genotypes, using the same notation as in Table 1.

Basically there are two major classes of similarity measure: unicriterion and multicriterion. For the former, there are four groups: (i) Euclidean distance (d), (ii) standardized distance (d_s), (iii) dissimilarity index (D), and (iv) correlation coefficient (r). In each group, there are two types of index, identified as (A) and (B). The identification (A) indicates that the index is defined for the similarity using both genetic effect and GE interaction, while identification (B) indicates the similarity is based on GE interaction alone.

Hanson's (1970) Euclidean distance (Table 2), originally defined as comparative genetic stability, and the correlation coefficients proposed by Guitard (1960) and Habgood (1977), apparently have not been used for formal clustering, but a genotypic pattern can be iden-

Table 2. A summary of similarity coefficients between genotypes i and i' , other notations as in Table 1.

| Similarity measure defined | Authors or users |
|--|---|
| 1) Unicriterion approach | |
| Euclidean distance: | |
| $d^2(A)_{ii'} = \sum_{j=1}^q (X_{ij} - X_{i'j})^2$ | Hanson (1970, Mungomery et al. (1974), Johnson (1977) |
| $d^2(B)_{ii'} = \sum_{j=1}^q [(X_{ij} - \bar{X}_i) - (X_{i'j} - \bar{X}_{i'})]^2/q$ | Abou-El-Fittouh et al. (1969)† |
| Standardized distance: | |
| $d_s^2(A)_{ii'} = \sum_{j=1}^q \left(\frac{X_{ij} - \bar{X}_i}{S_i^{\ddagger}} - \frac{X_{i'j} - \bar{X}_{i'}}{S_{i'}^{\ddagger}} \right)^2$ | Fox and Rosielle (1982)† |
| $d_s^2(B)_{ii'} = \sum_{j=1}^q \left(\frac{X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..}}{w_i^{\ddagger}} - \frac{X_{i'j} - \bar{X}_{i'} - \bar{X}_j + \bar{X}_{..}}{w_{i'}^{\ddagger}} \right)^2$ | Abou-El-Fittouh et al. (1969)† |
| Dissimilarity index: | |
| $D(A)_{ii'} = \frac{SSD\$(V_i V_{i'}) - SSD(V_i) - SSD(V_{i'})}{\sum_i SSD(v_i)}$ | Lin and Thompson (1975) |
| $D(B)_{ii'} = \sum_{j=1}^q [(X_{ij} - \bar{X}_i) - (X_{i'j} - \bar{X}_{i'})]^2 / [2(q - 1)]$ | Lin (1982) |
| Correlation coefficient: | |
| $r(A)_{ii'} = \frac{\sum_j (X_{ij'} - \bar{X}_i)(X_{i'j} - \bar{X}_{i'})}{[\sum_j (X_{ij} - \bar{X}_i)^2 \sum_j (X_{i'j} - \bar{X}_{i'})^2]^{1/2}}$ | Guitard (1960)† |
| $r(B)_{ii'} = \frac{\sum_j (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})(X_{i'j} - \bar{X}_{i'} - \bar{X}_j + \bar{X}_{..})}{[\sum_j (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2 \sum_j (X_{i'j} - \bar{X}_{i'} - \bar{X}_j + \bar{X}_{..})^2]^{1/2}}$ | Habgood (1977)† |
| 2) Multicriterion approach | |
| Pattern distance‡: $d_{ii'}^2 = 2(1 - \cos \theta_{ii'})$ | Lefkovitch (1985) |
| Frechet distance‡: $\psi_{ii'}^2 = (\bar{X}_i - \bar{X}_{i'})^2 + (S_i - S_{i'})^2$ | |

† Originally proposed for grouping environments.

‡ $S_i^2 = \sum_{j=1}^q (X_{ij} - \bar{X}_i)^2 / (q - 1)$, $w_i^2 = \sum_{j=1}^q (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_{..})^2$.

§ $SSD(V_i V_{i'})$ and $SSD(V_i)$ represent the sum of squares of deviations from the joint regression and from individual regression respectively.

¶ See Lefkovitch (1985).

tified through inspection of the respective distance matrices. Johnson's (1977) distance is based on a model that combines both regression and multiplicative effects (derived from principal component analysis), but because the residuals in this model are small, this distance is approximately equal to $d(A)_{ii}$. Lin and Thompson's (1975) dissimilarity index [$D(A)_{ii}$], based on the test statistics for a joint regression of a pair of genotypes, requires that: (i) the data fit the regression model, and (ii) the deviation SS are small relative to the error term and are also approximately homogeneous. If these requirements are not met, the Lin (1982) method, which assumes no specific model for GE interaction, can be used as an alternative. A specific feature of $D(A)_{ii}$ and $D(B)_{ii}$ is that, because the index constructed at each cluster cycle is equivalent to the MS of the respective ANOVA, these measures, combined with a significance level, provide a convenient cutting point for the dendrograms.

In contrast to unicriterion clustering, the multicriterion procedure developed by Lefkovich (1985) uses a cluster algorithm that permits more than one measure of pairwise relationship. He defined dissimilarity of genotypes by three measures: The mean over all environments (\bar{X}_i), the variance across environments (S_i^2), and among environments pattern distance (d_{ii}), and pointed out that extension of this list to include additional measures involves no new principle. The use of more than one criterion is made possible by the conditional clustering procedure (Lefkovich 1980, 1982), and that genotypes can be grouped using all measures either sequentially or simultaneously to decide on group homogeneity. Unlike unicriterion clustering, this decomposition provides an analytical means to investigate each measure separately.

To identify relationships among the similarity measures, the indices defined in Table 2 (except $D(A)_{ii}$, which is not directly related) are expressed in terms of $(S_i - S_r)^2$, $r(A)_{ii}$, [$= \cos \theta_{ii}$ (Lefkovich, 1985)], and $\bar{X}_i - \bar{X}_r$) as follows

$$d^2(A)_{ii} = qd^2(B)_{ii} + q(\bar{X}_i - \bar{X}_r)^2 \dots\dots\dots [8]$$

$$d^2(B)_{ii} = [(S_i - S_r)^2 + 2(1 - r(A)_{ii}) S_i S_r] (q - 1) / q \dots\dots\dots [9]$$

$$d_s^2(A)_{ii} = 2(q - 1)[1 - r(A)_{ii}] \dots\dots\dots [10]$$

$$d_s^2(B)_{ii} = 2[1 - r(B)_{ii}] \dots\dots\dots [11]$$

Note that:

$$D(B)_{ii} = d^2(B)_{ii} / [2(q - 1) / q] \dots\dots\dots [12]$$

and

$$d_{ii}^2 = d_s^2(A)_{ii} / (q - 1) \dots\dots\dots [13]$$

Equations [8] and [9] show that Lefkovich's (1985) three measures are combined in $d^2(A)_{ii}$ but with different scales for each component. Equations [10] and [11] reveal that the standardized distance and the simple correlation coefficient are equivalent. Equation [12] shows that Lin's (1982) dissimilarity index is equivalent to $d^2(B)_{ii}$. Equation [13] shows that Lefkovich's (1985) pattern distance is equivalent to $d_s^2(A)_{ii}$.

These comparisons show that:

- (i) Parallelism of responses is measured by a single criterion, GE interaction MS, in the case of $D(B)_{ii}$,

but by two criteria, S_i^2 and $r(A)_{ii}$, in the Lefkovich (1985) decomposition. Under the latter approach, for there to be no GE interaction, both $S_i = S_r$ and $r(A)_{ii} = 1$ must be true. In other words, the presence of a GE interaction implies either heterogeneity of environmental variances, or low correlation between genotypes, or both.

- (ii) The mean and the GE interaction are combined into one coefficient in $d^2(A)_{ii}$ and $D(A)_{ii}$; in $D(B)_{ii}$ they are kept separate and grouped in two steps (i.e., a multiple range test is applied to within-group means); in Lefkovich's decomposition they are kept as three separate entities.

- (iii) The meaning of similarity differs in the following ways: for $d^2(A)_{ii}$ and $D(A)_{ii}$, it means complete equality of genotype responses across environments; for $d^2(B)_{ii}$ and $D(B)_{ii}$, it means equality of all within location differences (or parallelism of response); and for $d_s^2(A)_{ii}$, $r(A)_{ii}$, $r(B)_{ii}$, it means the equality of all within location ratios.

CONCLUSIONS

The parametric approach to stability has the advantage of computational simplicity. Many statistics have been proposed, but unless the concept of stability and the kind of environments that are included in the experiment are clearly understood, they are of little use and may be misleading. In general, if the scientist is concerned about stability over an entire range of environments, Group A is useful. If the scientist is interested in comparing relative stability among the group of cultivars included in the experiment and if the regression model fits the data, statistics of Group C are best. When the data do not fit, or if the residual MS from the regression are heterogeneous, statistics W_i^2 or σ_i^2 of Group B should be used. Parameters of Group D (Type 3) are not recommended for the reasons described above.

Although the parametric approach to stability is relatively simple, it does not provide information for the resolution of any conflicting Type 1 and Type 2 inferences. Under these circumstances, a better procedure is to forgo a quantitative mathematical characterization, and to describe genotypes qualitatively as like or unlike other genotypes, i.e., to adopt a nonparametric clustering procedure. The scientist can choose from among the methods given in Table 2 (or others) depending on: i) which components he wishes to group for similarity; ii) whether these components should be mixed or treated as separate entities; and iii) if separately, whether they should be grouped in one or more than one steps.

In whatever way it is done, clustering allows subsets of genotypes to be described by the characteristics of the separate groups, although not directly in terms of stability. If a well-known cultivar is included in the test, it can be used as a paradigm for the other genotypes in the same subset. These genotypes may be regarded as having the overall characteristics of this cultivar, and extrapolation for them to a much wider range of environments than those tested may be possible (Lin and Binns, 1985). In this way, some of the advantages apparently lost by using a nonparametric technique can be regained.

APPENDIX

1. Plaisted and Peterson's (1959) $\bar{\theta}_i$.

Let $\theta_{i'}$ be the GE interaction MS for a pair of genotypes i and i' . Following the notation of Table 1, $\theta_{i'}$ can be written as:

$$\theta_{i'} = \sum_{j=1}^q (X_{ij} - \bar{X}_i - X_{i'j} + \bar{X}_{i'})^2 / [2(q-1)]$$

and mean variance for genotype i as,

$$\bar{\theta}_i = \sum_{i' \neq i}^p \theta_{i'}/(p-1)$$

$$= \sum_{i' \neq i}^p \sum_{j=1}^q (X_{ij} - \bar{X}_i - X_{i'j} + \bar{X}_{i'})^2 / [2(p-1)(q-1)].$$

Since

$$\begin{aligned} & \sum_{i' \neq i}^p \sum_{j=1}^q (X_{ij} - \bar{X}_i - X_{i'j} + \bar{X}_{i'})^2 \\ &= \sum_{j=1}^q [p(X_{ij} - \bar{X}_i)^2 + \sum_{i'=1}^p (X_{ij} - \bar{X}_i)^2 \\ & \quad - 2p(X_{ij} - \bar{X}_i)(\bar{X}_j - \bar{X}_.))], \end{aligned}$$

and since

$$\begin{aligned} \sum_{i'=1}^p (X_{ij} - \bar{X}_i)^2 &= \sum_{i'=1}^p (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_.)^2 \\ & \quad + p(\bar{X}_j - \bar{X}_.)^2, \end{aligned}$$

by substitution,

$$\begin{aligned} \bar{\theta}_i &= \sum_{j=1}^q \left[\sum_{i'=1}^p (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_.)^2 \right. \\ & \quad \left. + p(X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_.)^2 \right] / [2(p-1)(q-1)]. \end{aligned}$$

Q.E.D.

2. Plaisted's (1960) $\theta_{(i)}$.

Suppose the i th genotype is excluded from the whole set ($p \times q$) of data. Then the marginal mean of the j th environment is $(p\bar{X}_j - X_{ij})/(p-1)$ and the grand mean is $(p\bar{X}_. - \bar{X}_i)/(p-1)$.

By definition:

$$\begin{aligned} (p-2)(q-1)\theta_{(i)} &= \sum_{i' \neq i}^p \sum_{j=1}^q \left[X_{i'j} - \bar{X}_{i'} \right. \\ & \quad \left. - (p\bar{X}_j - X_{ij})/(p-1) + (p\bar{X}_. - \bar{X}_i)/(p-1) \right]^2 \\ &= \sum_{i' \neq i}^p \sum_{j=1}^q \left[Z_{i'j} + Z_{ij}/(p-1) \right]^2, \end{aligned}$$

where $Z_{uv} = X_{uv} - \bar{X}_u - \bar{X}_v + \bar{X}_.$

$$\begin{aligned} &= \sum_{i' \neq i}^p \sum_{j=1}^q Z_{i'j}^2 + (p-1) \sum_{j=1}^q Z_{ij}^2 + 2 \sum_{j=1}^q Z_{ij} \sum_{i' \neq i}^p Z_{i'j}/(p-1) \\ &= \sum_{i=1}^p \sum_{j=1}^q Z_{ij}^2 - \sum_{j=1}^q Z_{ij}^2 + \sum_{j=1}^q Z_{ij}^2/(p-1) \\ & \quad - 2 \sum_{j=1}^q Z_{ij}^2/(p-1) \\ &= \sum_{i=1}^p \sum_{j=1}^q Z_{ij}^2 - p \sum_{j=1}^q Z_{ij}^2/(p-1). \end{aligned}$$

Q.E.D.

REFERENCES

Abou-El-Fittouh, H.A., J.O. Rawlings, and P.A. Miller. 1969. Classification of environments to control genotype by environment interactions with an application to cotton. *Crop Sci.* 9:135-140.

Becker, H.C. 1981. Correlations among some statistical measures of phenotypic stability. *Euphytica* 30:835-840.

Breese, E.L. 1969. The measurement and significance of genotype-environment interactions in grasses. *Heredity* 24:27-44.

Easton, H.S., and R.J. Clements. 1973. The interaction of wheat genotypes with a specific factor of the environment. *J. Agric. Sci. (Camb.)* 80:43-52.

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

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

Fox, P.N., and A.A. Rosielle. 1982. Reducing the influence of environmental main-effects on pattern analysis of plant breeding environments. *Euphytica* 31:645-656.

Francis, T.R., and L.W. Kannenberg. 1978. Yield stability studies in short-season maize. 1. A descriptive method for grouping genotypes. *Can. J. Plant Sci.* 58:1029-1034.

Gaylor, D.W. 1956. Equivalence of two estimates of product variance. *J. Am. Stat. Assoc.* 51:451-453.

Gray, E. 1982. Genotype \times environment interactions and stability analysis for forage yield of orchardgrass clones. *Crop Sci.* 22:19-23.

Grubbs, F.E. 1948. On estimating precision of measuring instruments and product variability. *J. Am. Stat. Assoc.* 43:243-264.

Guitard, A.A. 1960. The use of diallel correlations for determining the relative locational performance of varieties of barley. *Can. J. Plant Sci.* 40:645-651.

Habgood, R.M. 1977. Estimation of genetic diversity of self-fertilizing cereal cultivars based on genotype-environment interactions. *Euphytica* 26:485-489.

Hanson, W.D. 1970. Genotypic stability. *Theor. Appl. Genet.* 40:226-231.

Johnson, G.R. 1977. Analysis of genotypic similarity in terms of mean yield and stability of environmental response in a set of maize hybrids. *Crop Sci.* 17:837-842.

Kang, M.S., and J.D. Miller. 1984. Genotype \times environment interactions for cane and sugar yield and their implications in sugarcane breeding. *Crop Sci.* 24:435-440.

Knight, R. 1970. The measurement and interpretation of genotype-environment interactions. *Euphytica* 19:225-235.

Lefkovich, L.P. 1980. Conditional clustering. *Biometrics* 36:43-58.

———. 1982. Conditional clusters, musters, and probability. *Math. Biosci.* 60:207-234.

———. 1985. Multi-criteria clustering in genotype-environment interaction problems. *Theor. Appl. Genet.* 70:585-589.

Lin, C.S. 1982. Grouping genotypes by a cluster method directly related to genotype-environment interaction mean square. *Theor. Appl. Genet.* 62:277-280.

———, and M.R. Binns. 1985. Procedural approach for assessing cultivar-location data: Pairwise genotype-environment interactions of test cultivars with checks. *Can. J. Plant Sci.* 65:1065-1071.

———, and B. Thompson. 1975. An empirical method of grouping

- genotypes based on a linear function of the genotype-environment interaction. *Heredity* 34:255-263.
- Mungomery, V.E., R. Shorter, and D.E. Byth. 1974. Genotype \times environment interactions and environmental adaptation. I. Pattern analysis—application to soya bean populations. *Aust. J. Agric. Res.* 25:59-72.
- Ntare, B.R., and M. Aken'Ova. 1985. Yield stability in segregating populations of cowpea. *Crop Sci.* 25:208-211.
- Plaisted, R.L. 1960. A shorter method for evaluating the ability of selections to yield consistently over locations. *Am. Potato J.* 37:166-172.
- , and L.C. Peterson. 1959. A technique for evaluating the ability of selections to yield consistently in different locations or seasons. *Am. Potato J.* 36:381-385.
- Perkins, J.M., and J.L. Jinks. 1968. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity* 23:339-356.
- Shukla, G.K. 1972a. Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity* 29:237-245.
- . 1972b. An invariant test for the homogeneity of variances in a two-way classification. *Biometrics* 28:1063-1072.
- Tai, G.C.C. 1971. Genotypic stability analysis and its application to potato regional trials. *Crop Sci.* 11:184-190.
- Wricke, G. 1962. Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. *Z. Pflanzenzucht.* 47:92-96.