Assessing Stability of Crop Varieties with Incomplete Data

B.M.K. Raju[1*], V.K. Bhatia[2] and L.M. Bhar[2]

[1]*Central Research Institute for Dryland Agriculture, Hyderabad*
[2]*Indian Agricultural Statistics Research Institute, New Delhi*

SUMMARY

Joint regression has been very popular among plant breeders to evaluate stability of crop varieties tested under multi location-year trials. The plant breeders often finish their investigation for stability with Eberhart and Russell model (1966) though the component of deviation from linear regression is found significant for many varieties. Consequently, in such situations one cannot do ranking of all the genotypes tested with regard to stability. Eventually pair-wise comparisons with respect to stability can be made only in a subset of genotypes, whose deviations from linear regression are found not significant. This paper aims to emphasize the limitations of Eberhart and Russell model in evaluating stability of a set of varieties tested and suggest plant breeders alternative measures of stability when Eberhart and Russell model fails to comment on the stability of a sufficiently large number of varieties tested. Another problem of plant breeders that this paper also attempts, is dealing with the situation of the stability analysis when some cells in two-way table of genotype × environments are blank. This paper examines methods cited in literature to handle incomplete data situations and brings out their practical relevance in the current generation of computers. An attempt has been made to develop handy computational algorithms wherever required and compares various procedures with respect to their capabilities in evaluating stability of the varieties.

Keywords: Eberhart and Russell model, Multi location-year trials, Stability, Stability variance.

1. INTRODUCTION

In developing countries like India, where the number of small and marginal farmers with small holdings is very high, stable yields minimize risk and ensure sustainable food supply. One of the plant breeders' aim has been to develop cultivars that produce stable yields across a range of environments. Environments may be locations or years or combinations of both.

The existence of interaction reflecting differences among varieties in their ability to maintain performance over a wide range of environmental conditions has long been recognized to exist (Finlay and Wilkinson 1963). This ability, which is an important property of a variety, is usually referred to as the sensitivity or adaptability of a variety. The basic ANOVA model for two-way crossed classification with interaction serves to obtain a rough idea about the partition of variance over different terms. However, it identifies the interaction as a source but does not analyze it since the interaction here is modeled by a separate, additive parameter for each combination of genotype by environment coarsely and un-parsimoniously. No attempt is made at the interpretation of this interaction, leaving the causes of interaction.

As an alternative to linear formulations of interactions, multiplicative formulations may be chosen in an attempt to quantify the variety's contribution to genotype × environment interaction. These

---
Corresponding author : B.M.K. Raju
 E-mail addresses : bmkraju@yahoo.com, bmkraju@gmail.com

multiplicative formulations permit the interpretation of interaction as differential genotypic sensitivity to environmental variable(s). Whenever the information on external environmental characteristics such as weather parameters and soil characteristics are available, it may be accommodated in the basic ANOVA model for making an attempt to interpret the interaction. This technique consists of regression of the estimated interactions of each variety on to the measured external environmental variables to obtain the linear sensitivities. However, it is difficult to obtain data on natural environments that comply with the data structure of varieties and properly explain variations in agricultural value of crop varieties. In such cases one may consider the regression of observed yield on the observed environmental mean yield. Finlay and Wilkinson (1963) reasoned that the average yield of a large group of genotypes can be used to describe a complex natural environment, without the complexities of defining or analyzing the important edaphic and seasonal factors. Environment averages, or their deviations from the overall average, are generally used as environmental indices. The resulting regression coefficient may be interpreted as linear sensitivity of the variety to environmental change. This technique was first used by Yates and Cochran (1938) and later by Finlay and Wilkinson (1963) and Eberhart and Russell (1966). This technique is popularly known as Joint Regression, as the joint effect of all the genotypes is used as explanatory regression variable.

Though this technique is very popular among plant breeders, it has got certain limitations. There is a need to elaborate these limitations and suggest alternative measures of stability that do not suffer from such limitations. Another problem faced by plant breeders is non-availability of data on all locations-years and varieties, which makes the data set obtained from Multi Environment Testing (MET) unbalanced. It may be incidental or accidental. The list of varieties being tested changes over the years and not all varieties are tested in all the environments since the genotypes change from year to year. As new varieties become available, older ones become obsolete which makes the data set unbalanced. Some causes for the accidental imbalances are non-germination, damage of crop on account of pests and diseases and floods etc. Literature cites some methods that can handle incomplete data situations. But it is again an issue for the plant breeders to choose the best technique for a given situation.

Hence, there is a need to study the existing methodologies to bring out their practical relevance in the current generation of computers and to develop handy computational algorithms for evaluating stability of the varieties. Section 2 focuses on limitations of Eberhart and Russell (1966) model and highlights alternative measures of stability. The subsequent sections elicit on the comparison of various methodologies under incomplete data situations.

2. STABILITY ANALYSIS FOR BALANCED DATA

2.1 Eberhart and Russell (1966) model

Eberhart and Russell (1966) proposed an observational formulation for the Joint Regression context. The model proposed by Eberhart and Russell (1966) is written as

$$y_{ij} = \alpha_i + \beta_i\, e_j + \delta_{ij}$$

where

y_{ij} is the performance of i-th genotype at the j-th environment ($i = 1, \ldots, K; j = 1, \ldots, N$) averaged over R replications

α_i is the mean of i-th genotype over all the environments

e_j is the environmental index for the j-th environment which is obtained as the mean of all genotypes at the j-th environment minus the general mean.

β_i is the regression coefficient measuring the linear sensitivity of i-th genotype to environment change.

δ_{ij} is the 'deviation from regression' of the i-th genotype in the j-th environment

Testing for the significance of genotype environment interaction

The significance of either $G \times E$ (linear) mean squares or pooled deviation mean squares or both when tested against average error confirms the presence of GE interaction. If the latter alone is significant then no useful prediction is possible from this approach. If both are significant then the practical usefulness of the predictions depends on the significance of former relative to the latter.

Stability and adaptability

A genotype with unit regression coefficient i.e. $\beta_i = 1$ and the mean squared deviation not significantly

different from zero ($\overline{S}_{d_i}^2 = 0$) is said to be stable. Significance of $\overline{S}_{d_i}^2$ from zero invalidates the linear prediction. If $\overline{S}_{d_i}^2$ is not significantly different from zero, the performance of the genotype for a given environment may be predicted. Accordingly, a genotype whose performance can be predicted is said to be stable and it also helps in choosing genotypes for specific adaptation.

Eberhart and Russell model analysis for groundnut data

The data used in this study has been provided by Regional Agricultural Research Station (RARS), Palem of Acharya N.G. Ranga Agricultural University, Andhra Pradesh. The data was an outcome of multi location trials of released and pre-released varieties of groundnut conducted at research stations situated in different agro-climatic zones of Andhra Pradesh. The data was consisted of 15 varieties of groundnut viz., TPT-1, TPT-2, Girnar-1, ICG (FDRS)-4, ICG (FDRS)-10, K-134, SVGS-1, TCGS-1, TCGS-3, ICGV-86699, Kadiri-3, ICGS-11, ICGS-44, JL-24 and TMV-2. These are designated as G-1 to G-15 respectively. These varieties were grown in Kharif-1990 and Kharif-1991. The locations used for these trials were 14. The 6 locations, namely, RARS-Tirupati, ARS-Utukur, ARS-Darsi, RARS-Nandyal, ARS-Seethampet, RARS-Palem were used for Kharif-1990 as well as Kharif-1991. They are designated as E-1 to E-6 for Kharif-1990 and E-9 to E-14 for Kharif-1991. The 2 locations, namely, ARS-Kadiri and RARS-Jagitial were used in Kharif-1990 only. These were designated as E-7 and E-8. Remaining 6 locations, namely, ARS-Ananthapur, ARS-Peddipalli, ARS-Peddapuram, RARS-Yellamanchili, ARS-Ragolu, ARS-Vizayanagaram were used in Kharif 1991 only. These were designated as E-15 to E-20. The structure of environments is as under.

Location-Year combination is treated as environment and accordingly 20 environments are designated as E-1 to E-20. The experiments were laid in Randomized Block Design (RBD) with 3 replications. The pod yields were expressed as kg/ha. The mean data over the replicates for the 15 genotypes and 20 environments is given in Raju (2002).

The stability statistics of Eberhart and Russell's model are presented in Table 1. The results revealed that there was significant difference among the genotypes indicating wider genetic diversity among the genotypes. Genotype × Environment (linear) and pooled deviation were found to be significant when tested against pooled error. It indicated significant Genotype × Environment interaction. Genotype × Environment (linear) interaction was found to be not significant when tested against pooled deviation which implies that the genotypes do not differ for their regression on environmental index and overwhelming portion of G × E interaction is of non-linear type, which ultimately makes the behaviour of genotypes unpredictable. On examining the significance of deviation from linear regression for the 15 genotypes in Table 1, all the deviations are significant at 1% level except genotype-7 and genotype-14. The deviation for the genotype-14 is not significant and the regression coefficient β_i is around unity (0.921) and as such it is regarded as stable variety. Similarly, the deviation for genotype-7 is not significant at 1% level and the coefficient of linear sensitivity is very close to unity, hence this can also be regarded as stable variety. Genotype-6 tops with respect to the average yield over the environments. However, the significance of deviation from linear regression makes its behaviour unpredictable over the environments and one may not be able to comment on its stability from Eberhart and Russell model's point of view.

Location Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Kharif 1990	E-1	E-2	E-3	E-4	E-5	E-6	E-7	E-8	×	×	×	×	×	×
Kharif 1991	E-9	E-10	E-11	E-12	E-13	E-14	×	×	E-15	E-16	E-17	E-18	E-19	E-20

Limitation of Eberhart and Russell (1966) model with regard to making comments on the stability of each genotype

Eberhart and Russell (1966) model proves to be a good tool for understanding the nature and type of GE interaction exhibited by the data set in the sense that whether a dominating portion of it, is linear or non linear type. The scope of stability parameters of Eberhart and Russell (1966) model (β_i and $\overline{S}_{d_i}^2$) is limited as it may not be possible to comment on the stability of each of the genotypes tested. Whenever the component of deviation from linear regression is found to be significant for a genotype, then it is not possible to make any comments about the stability of that genotype. A genotype may possibly be stable, but due to the fact that its interaction with environments is not of linear type, one becomes handicapped to make any comments on its stability. Further, if there is no information about the stability of some genotypes, one cannot make any comparison among all the genotypes with respect to stability. The scope of the investigation

Table 1. Stability analysis results for balanced data

Eberhart and Russell Model Statistics				Stability Variance Statistics	
Source	df	MS	$\hat{\beta}_i$	Stability variance	Stability rank
Genotypes	14	254686 **	-	-	-
Env + Gen × Env	285		-	-	-
Env (linear)	1		-	-	-
Gen × Env (linear)	14	62925 [NS]	-	-	-
Pooled deviation	270	90839	-	-	-
G-1	18	58480 **	1.034	57030	6
G-2	18	40652 **	0.961	37716	3
G-3	18	165738 **	1.096	177785	13
G-4	18	211461 **	1.100	228151	14
G-5	18	227958 **	1.110	247177	15
G-6	18	51799 **	1.109	54365	5
G-7	18	35235 *	0.998	31153	2
G-8	18	57064 **	0.918	57951	7
G-9	18	83747 **	1.074	86689	10
G-10	18	127073 **	1.095	135561	12
G-11	18	52538 **	0.910	53564	4
G-12	18	59875 **	0.942	59582	8
G-13	18	73167 **	0.833	84863	9
G-14	18	30754 [NS]	0.921	28965	1
G-15	18	87042 **	0.899	92239	11
Average error	560	20168			

* Significant at 5% level of significance   ** Significant at 1% level of significance

will in that case be limited only to a subset of genotypes tested, whose deviations from linear regression are found to be not significant. Pair-wise comparisons with respect to stability can be made in that subset only. Ranking of all the genotypes tested with regard to stability will not be possible. In these situations one thus explores some stability measures on the lines of stability variance given by Shukla (1972). This, however, is not based on linear regression model. As a result, some limitations which are inherent in linear regression model, can be overcome. Nevertheless it becomes inevitable to look for alternatives when component of deviation from linear regression is found significant for more number of genotypes. From the results in Table 1, it is seen that the component of deviation from linear regression is significant at 1% level of significance for all genotypes except G-7 and G-14. For such data sets, it is cautioned that plant breeders may not end up their investigation for stability with Eberhart and Russell model alone.

2.2 Stability Variance — An Alternative Measure of Stability

Stability variance of *i*-th variety given by Shukla (1972) measures the variance of interaction residuals of i-th variety. The genotype with smallest stability variance was the most stable among the genotypes tested. The genotype with second smallest stability variance was the second most stable among the genotypes tested.

Stability variance of *i*-th variety given by Shukla (1972) is

$$\hat{\sigma}_i^2 = \frac{1}{(K-1)(K-2)(N-1)}\left[K(K-1)W_i - \sum_{s=1}^{K}W_s\right]$$

where $W_i = \sum_j (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} - \bar{y}_{..})^2$ is the Wricke's ecovalence for the i-th genotype. It is shown that for balanced data Shukla's estimator is a MINQUE (Minimum Norm Quadratic Unbiased Estimator) of σ_i^2. It is obvious that Shukla's estimator of stability variance is equivalent to Wricke's ecovalence W_i for ranking purposes.

Shukla model analysis for groundnut data

Stability variances were computed for the 15 genotypes of groundnut. The genotypes were ranked

with respect to their stability. The results are presented in Table 1. Genotype-14 is found to have maximum stability in pod yield whereas Genotype-5 is found to have least stability variance among the 15 varieties tested. Obviously there is no restriction with this stability variance measure while making stability comparisons among the varieties tested. This stability measure can capture nonlinear interactions too. This measure permits to make stability comparisons among the 15 varieties tested unlike the Eberhart and Russell (1966) model, where one can only make stability comparisons between varieties 7 and 14.

3. STABILITY ANALYSIS WITH INCOMPLETE GENOTYPE BY ENVIRONMENT DATA

The two stability approaches (i) Joint Regression, and (ii) Stability Variance for incomplete data situation are discussed as under.

3.1 Joint Regression Approach

When the yields of some of the genotypes are not available or are not reliable, then the orthogonality of the design is not satisfied and bias is introduced in the observed varietal means. The comparison based on these means is likely to favour the varieties which happen to be exposed to better than average environmental conditions. Hence before proceeding to evaluate stabilities, such compensation needs to be made in the means for the environments in which particular varieties are not present. This section describes two such procedures, namely, (i) Joint Regression with Fitcon estimates, and (ii) Modified Joint Regression. Though it is established that the latter one is a more generalized one, but to establish the superiority of the latter the details of methodology is outlined. An iterative algorithm has been developed for the latter procedure, which would be very handy for the programmers attempting to evaluate it.

3.1.1 Joint regression with Fitcon estimates

Fitcon Analysis: The usual method of obtaining the aforesaid compensation or adjustment is to use a fitting constants technique, described by Patterson (1978), for the additive model

$$y_{ij} = \alpha_i + e_j + \varepsilon_{ij} \qquad (3.1)$$

where

y_{ij} is the (average) yield of i-th variety in j-th environment.

α_i is the mean of i-th variety.

e_j is the effect of j-th environment.

ε_{ij} is the random error, distributed normally with mean zero and a constant variance.

For estimating the parameters α_i and e_j, we have to minimize the residual sum of squares $\sum_{i,j}\left(y_{ij} - \alpha_i - e_j\right)^2 \delta_{ij}$ with respect to both α_i and e_j; noting that the weights δ_{ij}, introduced to obtain the incomplete data set-up are such that

$$\delta_{ij} = \begin{cases} 1 \text{ if } y_{ij} \text{ is present in the data} \\ 0 \text{ if } y_{ij} \text{ is missing} \end{cases}$$

The iterative algorithm to solve for the parameters of (3.1) is as under

$$\hat{\alpha}_i = \overline{y}'_{i.} + \sum_{j*} \frac{e_j}{n_i} \tag{3.2}$$

$$\hat{e}_j = \overline{y}'_{.j} - \sum_{i} \frac{\delta_{ij}\alpha_i}{n_j} \tag{3.3}$$

where $\overline{y}'_{i.}$, $\overline{y}'_{.j}$ are the means based on the existing n_i and n_j observations for the i-th variety and j-th environment respectively. Adjustments for these estimates depend on each other's final estimates. Summation over j^* is for those environments where the i-th variety is found to be absent.

Firstly, start the iteration by considering the trial value $\overline{y}'_{i.}$ for $\hat{\alpha}_i$ in (3.3) giving rise to a set of e_j values. Substitution of these values in (3.2) gives rise to revised estimates of α_i's. These are then substituted in (3.3) to get the revised estimates of e_j's. This cycle is continued till we reach more or less stable values for α_i and e_j.

During 1970s, computation of inverse of matrices of higher dimension was indeed a difficult task. Majority of researchers, during that period, were busy in deriving numerical techniques that could yield an approximate solution to normal equations. In those days, the above iterative algorithm could prove to be very handy. But it becomes redundant as on today in the light of advanced software and hardware technology. The current generation of computers can invert a matrix of any dimension in no time. With incomplete data set considered in this study, subroutine of SAS statistical software namely *lsmeans*[*] produced almost same results as produced by Patterson's (1978) Fitting constants technique.

The stabilized Fitcon estimates for e_j can be used for Joint Regression. The linear sensitivity for each variety can be estimated by regressing the existing y_{ij}'s on the Fitcon estimates of e_j's as shown below

$$y_{ij} = \tau_i + \beta_i \hat{e}_j + \delta_{ij}$$

However, the estimates of τ_i are not the same as the estimates of α_i except for the varieties that are present in all the environments or that have unit sensitivities. The discrepancy arises from the fact that the adjustment given in (3.2) is made to be of same degree for every variety; i.e. there is no allowance for varieties differing in their response or sensitivity to environmental effect. When such differences are expected, it is better to incorporate the parameter β_i in the adjustment as

$$\hat{\alpha}_i = \overline{y}'_{i.} + \beta_i \left(\sum_{j*} \frac{e_j}{n_i} \right)$$

This leads to the consideration of the non-additive model

$$y_{ij} = \alpha_i + \beta_i e_j + \varepsilon_{ij}$$

Digby (1979) proposed this improved adjustment in his modified Joint Regression analysis for incomplete variety by environment data.

3.1.2 Modified Joint Regression

The model considered by Digby (1979) is

$$y_{ij} = \alpha_i + \beta_i e_j + \varepsilon_{ij} \tag{3.4}$$

Minimization of residual sum of squares $\sum_{i,j}\left(y_{ij} - \alpha_i - \beta_i e_j\right)^2 \delta_{ij}$ with respect to parameters α_i, β_i and e_j leads to the following normal equations

---

[*]   *Ismeans routine of SAS statistical software under ANOVA procedure produces the least square means of effects specified.*

$$\sum_j \delta_{ij} y_{ij} = \alpha_i \sum_j \delta_{ij} + \beta_i \sum_j \delta_{ij} e_j \qquad (3.5)$$

$$\sum_j \delta_{ij} e_j y_{ij} = \alpha_i \sum_j \delta_{ij} e_j + \beta_i \sum_j \delta_{ij} e_j^2 \qquad (3.6)$$

$$\sum_i \delta_{ij} \beta_i y_{ij} = \sum_i \delta_{ij} \beta_i \alpha_i + e_j \sum_i \delta_{ij} \beta_i^2 \qquad (3.7)$$

where δ_{ij} has the same meaning as given earlier.

Since equations (3.5), (3.6) and (3.7) are not linearly independent, they are to be solved subject to the constraint $\sum_j e_j = 0$. To solve for the parameters of (3.4) from (3.5), (3.6) and (3.7) subject to $\sum_j e_j = 0$, the following iterative algorithm is proposed

Step 1: Set the β_i's equal to one, which reduces the equations (3.5) and (3.7), the solutions of which ($\hat{\alpha}_i$ and $\hat{e}_j$) subject to constraint $\sum_j e_j = 0$ can be obtained from the iterative algorithm given in the equations (3.2) and (3.3)

Step 2: Substitute the estimates of e_j in equations (3.5) and (3.6) and obtain the estimates of β_i

Step 3: Substitute the estimates of β_i in equations (3.5) and (3.7). Treat β_i as fixed

Step 3a: Estimate α_i as $\hat{\alpha}_i = \overline{y}'_{i.} + \beta_i \left(\sum_{j*} \frac{e_j}{n_i} \right)$

Step 3b: Using the estimate of α_i obtained in step 3a, solve for the estimates of e_j as

$$\hat{e}_j = \frac{\sum_i \delta_{ij} \beta_i y_{ij} - \sum_i \delta_{ij} \beta_i \alpha_i}{\sum_i \delta_{ij} \beta_i^2}$$

Step 3c: Go to step 3a, till there is convergence in $\hat{\alpha}_i$ and $\hat{e}_j$

Step 4: Go to step 2, till there is convergence in $\hat{\alpha}_i$ $\hat{\beta}_i$ and $\hat{e}_j$

This algorithm is very handy for the programmers doing the analysis work.

3.2 Stability Variance Approach

Piepho (1994) proposed a procedure for estimating stability variance σ_i^2, when some cells in two-way table are empty. It is outlined as under

Let $\qquad x_{srj} = y_{sj} - y_{rj}$

and $V_{s-r}^2 = \frac{1}{N-1} \left[\sum_j x_{srj}^2 - \frac{\left(\sum_j x_{srj} \right)^2}{N} \right]$

where N is the number of environments in which the genotypes s, r are grown together.

We know that $E[V_{s-r}^2] = \sigma_s^2 + \sigma_r^2$ where $s = 1$, 2,…, $(K - 1)$ and $r > s$.

In order to estimate σ_i^2, the method of moments may be employed where the sample moments are equated to population moments. Replacement of $E[V_{s-r}^2]$ by V_{s-r}^2 may lead to the following system of equations, to be solved for σ_i^2

$$V_{s-r}^2 = \sigma_s^2 + \sigma_r^2 \qquad [s = 1, 2,\ldots, (K - 1) \text{ and } r > s]$$

There are $K(K - 1)/2$ different equations in K unknowns, so that for $K > 3$ there are more equations than there are unknowns. Grubbs' estimates are the least squares solutions of these equations (Jaech 1985).

Formally the system of equations can be represented in matrix notation as

$$\mathbf{Q}\sigma = \mathbf{V} \qquad (3.8)$$

where σ is a K dimensional vector of σ_i^2's, V is $K(K - 1)/2$ dimensional vector of V_{s-r}^2's and Q is a $K(K - 1)/2 \times K$ matrix with elements 0 and 1, that picks the appropriate σ_i^2's.

$Q'Q$ has full rank and thus can be inverted.
The solution of equation (3.8) is

$$\tilde{\tilde{\sigma}} = (Q'Q)^{-1} Q'V \qquad (3.9)$$

Grubbs' estimates are unbiased. If we take expectation on both sides of equation (3.9)

$$E[\tilde{\tilde{\sigma}}] = E[(Q'Q)^{-1} Q'V] = E[(Q'Q)^{-1} Q'Q\sigma]$$
$$= E[I\sigma] = \sigma$$

For 2 genotypes s and r, we can compute V_{s-r}^2 as long as they are grown together in atleast two environments. In this case, the 2 genotypes s and r are said to be connected. To obtain a unique solution of equation (3.8), we require that there be atleast K connected pairs of genotypes as we need atleast as many equations as there are unknowns. Also each genotype must be connected to atleast one other genotype.

Comparison of potential of various methods for incomplete data

The potential of a given method (for incomplete data) may be judged by its ability to reproduce the stability/sensitivity rank order given by the method with complete data. The Coefficient of Spearman's Rank Correlation between rank orders displayed by balanced and unbalanced data using that method may be used to quantify the potential. Comparison of various methods can be done with the help of their computed potentials.

Empirical study with unbalanced data

To evaluate the methodologies described for missing data, unbalancedness is created by eliminating 20 cells at random in the 15 × 20 matrix of genotype by environment yields. This unbalanced data has been analysed for Patterson's fitting constants, Joint Regression with Fitcon estimates and Digby's modified Joint Regression analysis with the iterative algorithms described earlier. The resulting parameter estimates are obtained and are given in Table 3.

On comparison of the unadjusted varietal means and the varietal means obtained from Patterson's Fitcon method, it is seen from equation (3.2) that if the variety of interest is absent in the positive environments, the adjustment is made upwards and vice versa. The sign and amount of adjustment is determined by the sum of environmental effects in which the variety of interest is absent. This sum of environmental effects can be obtained from Table 2. In this way the adjustment for variety-1 is made upwards which is absent in the environment-17 having the effect 457.83. Similarly, the correction for variety-2 is positive whereas the adjustment for variety-8 is negative, which is absent in E-5 and E-10. The adjustment for variety-6 is zero as it is present in all the environments.

Table 2. Estimated Environment effects with incomplete data

Environment	Fitcon	Modified Joint Regression
1	180.66	175.42
2	−359.02	−364.43
3	1361.18	1368.64
4	667.00	657.58
5	−308.73	−314.38
6	−138.20	−126.95
7	−187.75	−192.01
8	−912.38	−900.51
9	−47.85	−46.62
10	−918.90	−906.07
11	808.69	825.25
12	348.50	338.16
13	−446.28	−447.78
14	368.05	359.58
15	−83.74	−83.66
16	776.38	795.08
17	457.83	436.28
18	−535.52	−539.29
19	−657.34	−661.97
20	−372.60	−372.31

If one wants to compare the varietal means obtained from Joint Regression with Fitcon estimates and Patterson's Fitcon means, one has to study the adjustment given by Patterson's Fitcon method for the unadjusted means and the improvement offered by the Joint Regression with Fitcon estimates to Fitcon means by allowing the varieties to differ in their sensitivities to the environmental effect.

$$\hat{\alpha}_{i(f)} = \overline{y}'_{i.} + \sum_{j*} \frac{\hat{e}_j}{n_i}$$

$$\hat{\alpha}_{i(jf)} = \overline{y}'_{i.} + \beta_i \left(\sum_{j*} \frac{\hat{e}_j}{n_i} \right)$$

Table 3. Estimated variety parameters with incomplete data

Variety	Unadjusted mean	Fitcon mean	Joint Regression with Fitcon estimates		Modified Joint Regression analysis		Stability Variance Statistics	
			Mean	Sensitivity	Mean	Sensitivity	Stability variance	Stability rank
1	1478.68	1502.78	1503.32	1.023	1502.13	1.021	43976	6
2	1250.05	1321.69	1321.53	0.998	1321.95	0.998	39826	4
3	1422.79	1420.27	1420.03	1.094	1420.11	1.091	172119	13
4	1243.22	1296.39	1299.84	1.065	1301.22	1.076	235535	14
5	1325.37	1309.12	1307.24	1.116	1306.76	1.125	256158	15
6	1692.75	1692.75	1692.75	1.115	1692.75	1.115	57442	7
7	1481.58	1474.31	1474.29	1.003	1474.90	1.001	35030	3
8	1351.72	1283.52	1288.37	0.929	1288.90	0.927	27559	2
9	1390.53	1380.65	1379.82	1.083	1379.60	1.082	92553	10
10	1459.35	1401.81	1400.34	1.026	1399.55	1.030	109499	12
11	1338.50	1362.73	1361.63	0.955	1362.52	0.955	43863	5
12	1415.78	1472.20	1467.19	0.911	1465.98	0.908	59030	8
13	1480.58	1473.31	1474.50	0.836	1475.02	0.832	85740	9
14	1437.47	1456.85	1455.62	0.937	1455.16	0.935	26814	1
15	1324.79	1276.77	1282.45	0.882	1283.37	0.874	101612	11

where $\bar{y}'_{i.}$ is unadjusted mean; $\hat{\alpha}_{i_{(f)}}$ is Fitcon mean; and $\hat{\alpha}_{i_{(jf)}}$ is the mean obtained from the Joint Regression with Fitcon estimates.

One may identify the following 6 cases to study the adjustment offered to the Fitcon means by the Joint Regression with Fitcon estimates.

1. When $\sum_{j*}\hat{e}_j$ is negative (unadjusted means are corrected downwards) and $\beta_i < 1$: The adjustment to the Fitcon means are positive, e.g. variety-8.

2. When $\sum_{j*}\hat{e}_j$ is negative and $\beta_i > 1$: Fitcon means are corrected downwards, e.g. variety-5.

3. When $\sum_{j*}\hat{e}_j$ is positive (unadjusted means are corrected upwards) and $\beta_i < 1$: The adjustment to the Fitcon means is negative, e.g. variety-12.

4. When $\sum_{j*}\hat{e}_j$ is positive and $\beta_i > 1$: Fitcon means are corrected upwards, e.g. variety-4.

5. When $\sum_{j*}\hat{e}_j$ is zero, i.e. the variety of interest is absent in none of the environments: The adjustment given by Fitcon means as well as the adjustment given by the Joint Regression with the Fitcon estimates to the unadjusted means are zero e.g. variety-6.

6. When $\beta_i \cong 1$ i.e. the sensitivity is close to unity: The adjustment made to the Fitcon means by the Joint Regression with Fitcon estimates would be negligible, as Joint Regression with Fitcon estimates reduces to the Fitcon method, e.g. variety-7, variety-2.

One may also compare the means obtained from Digby's Modified Joint Regression with unadjusted (observed) varietal means, Fitcon means and the means obtained from Joint Regression with Fitcon estimates.

Unadjusted mean vs mean obtained from Digby's modified Joint Regression

If $(\sum\limits_{j*} \hat{e}_j)_D$ is positive, the unadjusted mean is corrected upwards and vice versa, where $(\sum\limits_{j*} \hat{e}_j)_D$ is the sum of environment effects in which the variety of interest is absent and obtained from Digby's modified Joint Regression analysis.

Fitcon means vs Digby's modified Joint Regression means

The improvement in varietal means offered Digby's modified Joint Regression over Patterson's Fitcon is determined by the quantity

$$\beta_i \left(\frac{\sum\limits_{j*} \hat{e}_j}{n_i} \right)_D - \left(\frac{\sum\limits_{j*} \hat{e}_j}{n_i} \right)_P = Q \text{ (say)}$$

where $\left(\dfrac{\sum\limits_{j*} \hat{e}_j}{n_i} \right)_P$ is adjustment offered by Patterson's

Fitcon method to the unadjusted mean. If $Q > 0$ then Fitcon means are corrected upwards and vice versa.

Joint Regression with Fitcon estimates vs modified Joint Regression

In Joint Regression with Fitcon estimates, the e_j's are merely unweighted means of $(y_{ij} - \alpha_i)$ for those varieties present in the j-th environment, whereas modified Joint Regression estimates e_j's as weighted means of $(y_{ij} - \alpha_i)$, the weights being proportional to the varietal sensitivities. Hence, the weighted means used by the iterative analysis are more appropriate.

For variety-15, $Q = -41.42 - (-48.02) = 6.6$. Hence the adjustment offered by modified Joint Regression to the Fitcon mean for variety-15 is +6.6.

Comparison between Joint Regression with Fitcon estimates and modified Joint Regression and stability variance

Joint Regression with Fitcon estimates and modified Joint Regression techniques are compared with respect to their ability to assess the sensitivity rank order obtained with balanced data. The association between the two has been quantified by Spearman's rank correlation. On observing the same sensitivity rank order with Joint Regression with Fitcon estimates and Digby's modified Joint Regression, the correlation of this rank order with the one obtained from balanced data using Eberhart and Russell (1966) model is 0.95. The result re-establishes that choice between Joint Regression with Fitcon estimates and modified Joint Regression is less critical when the varieties do not differ much in regard to sensitivities to environmental index.

Stability variance for the 15 genotypes of groundnut was evaluated using the methodology proposed by Piepho (1994) for incomplete two-way data. The genotypes were ranked with respect to their stability. The corresponding results are presented in Table 3. The rank correlation between the stability rank orders displayed by complete and incomplete data is found to be 0.9429, which is reasonably good concordance. So the researchers and plant breeders can use the measure of Stability Variance safely even in the case of incomplete data. This clearly indicated that stability variance measure is a robust measure of stability of crop variety.

As far as choice between Joint Regression and Stability Variance is concerned, it is to be kept in mind that stability rank order displayed by Joint Regression (based on deviation of sensitivity coefficient from 1) is not necessarily the true stability rank order of varieties. It becomes true stability rank order only when the components of deviation from linear regression are found to be not significant for all the tested varieties. This kind of problem, however, does not arise in case of Stability Variance measure.

4. CONCLUSION

In view of the above results, it is concluded that it is always better to employ the Stability Variance measure to evaluate the stability of a set of genotypes when the component of deviation from linear regression is found significant for sufficiently large number of varieties with Eberhart and Russell model. For incomplete data situation also it is preferable to use Stability Variance approach in place of Joint Regression as it gives stability rank order rather than conditional sensitivity rank order.

REFERENCES

Digby, P.G.N. (1979). Modified joint regression analysis for incomplete variety × environment data. *Cambridge J. Agril. Sci.,* **93**, 81-86.

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

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

Jaech, J.L. (1985). *Statistical Analysis of Measurement Errors.* Wiley, New York.

Patterson, H.D. (1978). Routine least squares estimation of variety means in incomplete tables. *J. National Instt. Agril. Botany,* **14**, 385-400.

Piepho, H.P. (1994). Missing observations in the analysis of stability. *Heredity,* **72**, 141-145.

Raju, B.M.K. (2002). A study on AMMI model and its biplots. *J. Ind. Soc. Agril. Statist.,* **55(3)**, 297-322.

Shukla, G.K. (1972). Some statistical aspects of partitioning genotype environmental components of variability. *Heredity,* **29**, 237-245.

Yates, F. and Cochran, W.G. (1938). The analysis of groups of experiments. *Cambridge J. Agril. Sci.,* **28**, 556-580.