Indian J. Genet., 70(1): 80-84 (2010)

An index for simultaneous selection of genotypes for high yield and stability under incomplete genotype x environment data

A. R. Rao*, Shiv Kumar Choudhary[1], S. D. Wahi and V. T. Prabhakaran

Indian Agricultural Statistics Research Institute, New Delhi 110 012
[1]Rajendra Agricultural University, Bihar

(Received: October 2009; Revised: January 2010; Accepted: February 2010)

Abstract

A family of simultaneous selection indices is proposed here, which can be used for selecting genotypes simultaneously for high yield and stability under incomplete genotype x environment situations. Three indices are proposed by assigning different weights to yield (w_1) and stability (w_2) as w_1=0.8, w_2 =0.2 (I_1); w_1=0.7, w_2 =0.3 (I_2); w_1=0.6, w_2 =0.4 (I_3). These indices are tested for their performance based on Pearsonian correlations between yield based ranks and index based ranks, stability based ranks and index based ranks on groundnut data. It was found that the performance of I_1 index is best for selecting high yielders as well as stable performers to the extent of 0%–10% of incompleteness in genotype x environment data. Among the top 4 out of 15 varieties selected based on I_1, it is found that 2 – 3 are high yielders and high stable performers for 0%–10% of incompleteness.

Key Words: Genotype x environment interaction, stability, Incomplete data, Simultaneous selection for yield and stability

Introduction

Genotype x Environment Interaction (GEI) continues to be a challenging issue among plant breeders, geneticists and production agronomists who test the performance of genotypes across diverse environments. Whenever an interaction is significant, the use of overall genotype means across environments become questionable. Stability of performance is considered as an important aspect of yield trials. However, the stability measure alone is of limited use. To be of practical utility in a breeding or cultivar testing programme, both stability and yield of genotypes should be taken into account. Hence, integration of stability of performance with yield through suitable measures will reduce the effect of GEI and will help in selecting cultivars in a more useful manner.

Quite often, it is noticed that a large number of entries in the genotype x environmental table are missing. Incomplete data are primarily the result of a few genotypes having been not tested in all the environments due to various constraints like insufficient seed, non-germination and pest and disease attack. It may also be possible that the entries under test for each environment may change over a period of time and in that new entries continuously replace the adequately tested entries or genotypes. Statistical techniques appropriate to complete data do not directly apply to incomplete GEI data.

Several methods of simultaneous selection for yield and stability and relationships among them were discussed by Kang [1-3] and Kang and Pham [4]. Kang [5] discussed the reasons for emphasizing stability in the selection process and developed a Yield-Stability statistic (YS_i) that has enabled incorporation of stability in the selection process. Bajpai and Prabhakaran [6] observed that Kang's rank-sum method has an inherent weakness that it is weighing heavily towards yield performance. Accordingly they proposed three new indices, which were found to be superior to Kang indices. Rao and Prabhakaran [7] proposed a new family of simultaneous selection indices based on AMMI model and showed the superiority of their indices over the indices developed by Bajpai and Prabhakaran [6]. However, the above mentioned indices are useful only when genotype x environment data is complete.

In this study, we propose a new family of indices that takes care of both high yield and stability while selecting genotypes under incomplete data situations. The new family of indices is also tested for performance on groundnut data.

---

*Corresponding author's e-mail: arrao@iasri.res.in

Materials and methods

For statistical analysis of multi-location trial data, it is common to assume the following model:

$$y_{ij} = \mu + g_i + e_j + (ge)_{ij} + \varepsilon_{ij}\,(i = 1, ..., t;\, j = 1, ..., s) \quad (1)$$

where y_{ij} is the mean yield of genotype *i* in environment *j*, is the grand mean (fixed), is the effect of enotype *i* (fixed), is the effect of environment *j* (random), is the interaction effect of genotype *i* with environment *j* (random) and is the mean error of genotype *i* in environment *j* (random). The effects and are independently normally distributed with means zero and variances $Var(ge_{ij}) = \sigma_i^{'2},\ \ Var(\varepsilon_{ij}) = \sigma^2,$ respectively. The assumption of homogeneous error variances is reasonable since the test design is the same for all environments. In accordance with the concept of stability given by Shukla [8], genotype-environment interaction variance is allowed to differ among genotypes. Maximum stability of a genotype is attained if the interaction variance $\sigma^{'2} = 0.$ The larger the less stable is the corresponding genotype. In the model for means one cannot distinguish interaction from error. It is solely possible to estimate and hence to estimate the variance of genotype *i*. It can be further noted that $\sigma_i^2 = \sigma_i^{'2} + \sigma^2$.

In the unbalanced data, can be obtained by the Minimum Norm Quadratic Unbiased Estimation (MINQUE) principle of estimation [9]. It is noted that for balanced data Shukla's estimator is a MINQUE of [8]. Rao [9] provides a computational procedure for MINQUE in the general case, which can be used in data sets with empty cells.

The model in Eq. (1) can also be written in matrix notation as

$$\mathbf{Y = Xb +} \quad (2)$$

where ***Y*** is the vector of observations, b is the parameter vector of main effects, is a vector of -effects and ***X*** is the design matrix. Denote by *n* the number of filled cells, by **M** = $\{m_{pq}\}$ (*p*, *q* = 1, ..., *n*) the projection matrix ***I – X (X'X)⁻X'***, by **t** vector of squares of the residuals (***I* – *X (X'X)⁻X'***)**Y**, and by vector of variances . ***(X'X)⁻*** stands for a g-inverse of ***(X'X)***. Here, the dimension of is equal to *n*, the number of filled cells. For a cell *p*, , where i is the subscript of the genotype in that cell. Defining $F = \{m_{pq}^2\}$ and considering the equations given by $F\theta = \tau$, without loss of generality, let the first n_1 variances of be equal to the second n_2 to , ... By adding up the first n_1 equations of one can get the first equation in , , ... Similarly, adding up the next n_2 equations the second equation in , , ... can be obtained. Denoting the reduced equations by , where is the vector [illegible] variances (assumed different), and using results obtained by Rao [9] the MINQUE of , ..., , can be obtained from the equation provided ***G*** is non-singular.

$$(=(1/t)\sum Y^2) \quad \left(\frac{1}{\hat{\sigma}_i^2}\right) \quad \frac{\bar{Y}_{Di}}{\bar{\bar{Y}}_D}\ ^2 \quad \frac{1}{t}\sum_i\left(\frac{1}{\hat{\sigma}_i^2}\right)$$

For the incomplete data, we propose simultaneous selection indices belong to the family

(3)

where σ_{Mi}^2 is the stability variance of the i^{th} variety obtained by MINQUE method, is the adjusted yield of the i^{th} variety obtained from the incomplete data using modified regression analysis [10], is the mean adjusted yield, w_1

and w_2 are the weights given to yield and stability components in the index such that $w_1 + w_2 = 1$.

A higher value of $\overline{Y}_{Di}$ is always desirable. If a genotype performs better than the average it will contribute a value > 1 to the Index. In contrast a higher value of (indicating lesser stability) is not desirable and that is why the inverse ratio of this parameter has been used in the index. The values of w_1 and w_2 are decided by the breeder depending on the importance he would want to attach to the stability component in the light of his requirement. At the moment we are not in a position to suggest any objective criterion for deciding w_1 and w_2. By assigning different values of weights w_1=0.8, w_2 =0.2; w_1=0.7, w_2 =0.3; w_1=0.6, w_2 =0.4, a new family of indices consisting of three indices I_1, I_2 and I_3 are proposed. Under the proposed method, the cultivars showing higher index values are selected. The potential of the selected cultivars for general/specific adaptation is decided on the basis of high yield and low

The index given by Bajpai and Prabhakaran [6] will become a particular case of the proposed index when the data is complete.

For comparing the performance of the proposed simultaneous selection indices, the yield data of multi-location varietal trials of released and pre-released varieties of groundnut conducted in different locations (RARS-Tirupati, ARS-Utukur, ARS-Darsi, RARS-Nandyal, ARS-Seethampet, RARS-Palem, ARS-Kadiri, RARS-Jagitial, ARS-Ananthapur, ARS-Peddipalli, ARS-Peddapuram, RARS-Yellamanchili, ARS-Ragolu, ARS-Vizayanagaram) of Andhra Pradesh over years 1990 & 1991 [11] is used. The data consists of pod yield (kg/ha) of 15 groundnut varieties raised over 20 environments (location-year combinations). The experiments were laid in Randomized Complete Block Design (RCBD) with 3 replications. Incompleteness in the above data has been created by randomly deleting the observations, to the extent of 1%, 5% and 10% of the total sample size (300).

Results and discussion

The analysis of variance with GEI is done on the complete data obtained from the groundnut varietal trials and it is found that genotype, environment and genotype x environment interactions are highly significant (Table 1). Various selection measures for complete data i.e. yield, stability, and indices (I_1, I_2 and I_3) are worked out. The indices I_1, I_2 and I_3 are worked out for three incomplete data sets (with 1%, 5% and 10% missing observations) by computing the adjusted mean yield ( ) and MINQUE estimates of stability variance ( ) of different varieties. The estimates of mean yield, stability variance and values of all the three indices along with the ranks of different varieties under 0%-10% missing observations are obtained to work out the correlations between index rank and yield rank as well as stability ranks and are presented in Table 2.

Top 4 varieties selected out of 15 varieties on the basis of Index values, i.e. 25% of the total varieties, are arranged from rank 1 to rank 4 and are shown in Table 2. The superscripts and subscripts of each selected variety indicate yield rank and stability rank respectively. Rank one indicates the variety with highest yield / low stability variance / high simultaneous selection index value. The correlations between yield based ranks and $\overline{\Phi}^2_{Mi}$ $\overline{Y}_{Di}$ index based ranks, stability based ranks and index based ranks were tested by t-test with t–2 degrees of freedom and found to be significant at 5% level of significance. Further, it was observed that the index based ranks are highly correlated with yield ranks and stability ranks thereby indicating the performance of index for selecting genotypes simultaneously for yield and stability. The correlations between Index rank and stability rank are of same order for all the three indices whereas the correlation between index rank and yield rank is higher in I_1 followed by I_2 and I_3 respectively. The performance of I_1 and I_2 are comparable for selecting stable performers. The proportion of selected stable performers in I_3 is slightly higher as compared to I_1 and I_2. This is true even when the observations are

Table 1. Analysis of variance for G x E Interaction on groundnut pod yield (kg/ha)

Source	d.f.	Sum of squares	Mean sum of square	F-ratio
Genotypes	14	3565604.00	254686.00	12.63**
Environments	19	107622796.00	5664357.70	280.86**
G x E interaction	266	25408293.00	95519.90	4.74**
Average error	560	11294080.00	20168.00	

** - (P<0.01)

Table 2. Correlations between index rank and yield, stability ranks and top four varieties selected on the basis selection index values along with their yield and stability ranks at different levels of missing observations

Index		Index rank			
		0% missing	1% missing	5% missing	10% missing
I_1	Yield rank	0.64**	0.60**	0.65**	0.63**
	Stability rank	0.94**	0.95**	0.93**	0.94**
	Selected varieties#				
I_2	Yield rank	0.55**	0.55**	0.56**	0.50**
	Stability rank	0.96**	0.96**	0.96**	0.98**
	Selected varieties#				
I_3	Yield rank	0.51**	0.51**	0.54**	0.50**
	Stability rank	0.97**	0.97**	0.98**	0.98**
	Selected varieties#				

** - ($P<0.01$); * - ($P<0.05$); Source of data: Raju (2002)

- subscripts indicating stability rank and superscript indicating yield rank

[illegible]

missing to the extent of 10 per cent of the total sample size.

The proportion of high yielders and most stable performers out of top 4 varieties selected on the basis of index values are presented in Table 3. It is observed that among the top 4 varieties selected based on the proposed indices, around 2 - 3 are high yielders and high stable performers. It is evident from the results that among the 3 indices, the indices I_1 and I_2 are almost robust against missing observations for selecting high yielders and stable performers. In case of I_3, the proportion of high yielders selected increases and the proportion of high stable performers selected decreases under missing observations.

Since, the proposed index I_1 show significant correlations with both high yield and high stability as

Table 3. Number of high yielders (HY) and high stable performers (HSP) present out of top 4 genotypes selected on the basis of proposed indices

Index		Index rank			
		0 % missing	1% missing	5 % missing	10% missing
I_1	HY	3	2	3	2
	HSP	2	3	2	3
I_2	HY	2	2	2	2
	HSP	3	3	3	3
I_3	HY	1	2	2	2
	HSP	4	3	3	3

well as selects large proportion of high yielders and stable performers, this index can be used by the breeders and production agronomists for selecting genotypes simultaneously for high yield and stability, even when the data are incomplete up to the extent of 10% of total observations.

References

1. **Kang M. S.** 1988. A rank-sum method for selecting high-yielding, stable corn genotypes. Cereal Res. Commun., **16**: 113-115.
2. **Kang M. S.** 1990. *Genotype-by-environment interaction and plant breeding*. Louisiana State Univ., Agric. Center, Baton Rouge, LA.
3. **Kang M. S.** 1993a. Issues in GE interaction. p. 67-73. *In:* V. Rao, I. E. Hanson and N. Rajanaidu (eds.). Genotype-environment interaction studies in perennial tree crops. Palm Oil Research Institute of Malaysia, Kuala Lumpur.
4. **Kang M. S. and Pham H. N.** 1991. Simultaneous selection for high yielding and stable crop genotypes. Agron. J., **83**: 161-165.
5. **Kang M. S.** 1993b. Simultaneous selection for yield and stability in crop performance trials: Consequences for growers. Agron. J. **85**: 754-757.
6. **Bajpai P. K. and Prabhakaran V. T.** 2000**.** A new procedure of simultaneous selection for high yielding and stable crop genotypes. Ind. J. Genet.*,* **60**: 141-146.
7. **Rao A. R. and Prabhakaran V. T.** 2005. Use of AMMI in simultaneous selection of genotypes for yield and stability. Jour. Ind. Soc. Ag. Statistics, **59**: 76-82.
8. **Shukla G. K.** 1972. Some statistical aspects of partitioning genotype-environmental components of variability. Heredity, **29**: 237-245.
9. **Rao C. R.** 1970. Estimation of heteroscedastic variances in linear models. J. Am. Stat. Assoc., **65**: 161-172.
10. **Digby P. G. N.** 1979. Modified joint regression analysis for incomplete variety x environment data. J. Agric. Sci., **93**: 81-86.
11. **Raju B. M. K.** 2002. A study on AMMI model and its biplots. Jour. Ind. Soc. Ag. Statistics, **55**: 297-322.