

## **Comparison of Various Measures of Stability with Respect to Ranking Ability under Varying Situations**

B.M.K. Raju and V.K. Bhatia

*Indian Agricultural Statistics Research Institute, New Delhi-110012*

(Received : October, 2002)

### **SUMMARY**

Though large number of stability measures are available in literature, the problem of plant breeder has been to decide which of the stability measure is to be chosen for the purpose of selecting stable genotypes. In other words, the problem is of identification of a reliable stability measure. Further, there is some evidence that for a variety, the interaction effect with environment may be non-normally distributed. If such is the case, what are the robust measures of stability that a plant breeder may rely upon for his purpose? To identify the robust and reliable measures of stability, interaction effects having normal as well as different non-normal distributions are generated using Monte Carlo simulation. Thirteen measures of stability (7 parametric and 6 nonparametric) have been considered in total for the study. Defined values of  $\sigma_i^2$  have been assigned to each genotype so that the true stability rank order is known a priori. The measures of stability taken for the study are then compared with respect to their ability to assess the true rank order, which further has been quantified by Spearman's rank correlation averaged over 1000 runs. The results suggest that, given a normal distribution of interaction effects, it is best to estimate stability by the MINQUE of  $\sigma_i^2$  (or equivalently by  $W_i/W_{i(\text{AMMI})}$ ). The situation does not change dramatically under mild departures from normality. However with extreme and longer tailed distributions, significantly departing from normality for the interaction effects, it may be worthwhile to use one of the more robust measures.

*Key words* : Genotype  $\times$  Environment interaction, Non-normal distribution, Robust measures of stability, Monte Carlo simulation, Rank correlation.

### *1. Introduction*

Recently, yielding stability as a selection trait in plant breeding programmes and evaluation trials is gaining importance over yielding ability especially in developing countries like India, where the number of small and marginal farmers with small holding is very high. A main strategy among small-scale subsistence farmers, particularly in marginal areas, is risk minimization. In

these areas, stable yields are the key to sustainable food supplies. The interest of plant breeder in stability stems from the need to develop well-buffered cultivars. The term stability refers to the behaviour of a crop in varying environments. The breeders' aim is to develop cultivars that are stable across a range of environments. Environments may be locations or years or combinations of both.

Two different approaches to assessing stability may be distinguished: the static concept and the dynamic concept (Becker and Leon [1]). According to the static concept (Type 1 statistics in Lin *et al.* [9]), maximum stability occurs when the yield of the genotype under consideration is constant across the environments, i.e., the stability considered in the sense of homeostasis. According to the dynamic concept (Type 2 statistics in Lin *et al.* [9]), a genotype is regarded as stable if its performance in different environments is close to what can be expected from the potentials of those environments. Maximum stability occurs if the difference between a genotype's yield and the environmental index, commonly defined, measured by the mean of all tested genotypes in that environment, is constant across the environments. Whenever this difference is not the same across all environments, the corresponding genotype is said to interact with environments. This interplay of genetic and non-genetic effects causing differential relative performances of different varieties in different environments is called genotype-environment interaction. 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 [4]). This ability, which is an important property of a variety, is usually referred to as the sensitivity or adaptability of a variety.

Though large number of stability measures (parametric as well as non-parametric) are available in literature, the problem of plant breeders has been, which out of these large number of stability measures may be chosen for the purpose of selecting stable genotypes. In other words, the problem is of identification of the reliable stability measure. Further, Piepho [12] has shown that interaction effects for a variety may be non-normally distributed. If such is the case, what are the robust measures of stability that a plant breeder may rely upon for his purposes. Piepho [14] made a simulation study and identified some robust and reliable measures of stability. The present study is in this direction where many more measures of stability and non-normal distributions are considered for arriving at robust and reliable measures of stability.

## 2. Model -- Measures of Stability

### 2.1 Model

The measures of stability are usually based on the statistical analysis of two-way linear model for the mean of  $i$ -th genotype grown in  $j$ -th environment.

$$y_{ij} = \mu + \alpha_i + \beta_j + v_{ij}$$

where

$\mu$  is the grand mean

$\alpha_i$  ( $i = 1, 2, \dots, K$ ) is the fixed additive genetic effect of the  $i$ -th genotype

$\beta_j$ , ( $j = 1, 2, \dots, N$ ) is the random additive environmental contribution of the  $j$ -th environment

$v_{ij}$  is the random effect comprising the genotype environment interaction and the mean error associated with the  $i$ -th genotype in the  $j$ -th environment

A genotype  $i$  is said to be stable if  $\sigma_i^2$ , the variability of effects  $v_{ij}$  associated with that genotype is small. We have a large number of stability statistics at our disposal, which may be interpreted as measures of spread of  $v_{ij}$ 's. Now, the aim of this study is to suggest the measures of stability that best reflect the "true" stability rank order and to identify the measures that are robust to changes in the parent distribution of the  $v_{ij}$  effects, which is often assumed normal, but may at times, depart from normality (Piepho [12]). Further one can work out different measures of stability rank order from the basic model as well.

The basic model for two-way crossed classification with interaction is

$$Y_{ijr} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijr}$$

$$i = 1, 2, \dots, K; j = 1, 2, \dots, N; r = 1, 2, \dots, R$$

where

$Y_{ijr}$  is the yield of  $r$ -th replicate of  $i$ -th variety in  $j$ -th environment

$\mu$  is the overall mean

$\alpha_i$  is the fixed effect of  $i$ -th variety

$\beta_j$  is the random effect of  $j$ -th environment

$(\alpha\beta)_{ij}$  is the random interaction effect of  $i$ -th genotype in  $j$ -th environment

$e_{ijr}$  is the pure error associated with  $Y_{ijr}$

The underlying assumptions are

$\beta_j$ 's are independent and identically distributed as  $N(0, \sigma_\beta^2)$

$(\alpha\beta)_{ij}$  are independent and distributed as  $N(0, \sigma_{\alpha\beta}^2)$

$e_{ijr}$  are independent and identically distributed as  $N(0, \sigma_e^2)$

The assumption of homogeneous error variances is reasonable if the test design is same for all the environments.

Usually stability statistics are based on means model (atleast the measures considered in this study) which looks as follows

$$y_{ij} = \mu + \alpha_i + \beta_j + v_{ij}$$

where  $y_{ij} = \sum_{r=1}^R y_{ijr} / R$  and  $v_{ij} = (\alpha\beta)_{ij} + \bar{e}_{ij}$  with  $\bar{e}_{ij} = \sum_{r=1}^R e_{ijr} / R$

$v_{ij}$  is random effect comprising the GE interaction and mean error associated with i-th genotype and j-th environment.

With the means model it is no longer possible to distinguish between GE interaction and mean error. Now the error related to  $G \times E$  interaction in means model is  $\bar{e}_{ij}$ , mean of pure error over replications. This can not be extracted from  $v_{ij}$  as it is inseparable from interaction  $(\alpha\beta)_{ij}$ . Only  $v_{ij}$  can be estimated. Further,  $\bar{e}_{ij}$  (the error related to  $G \times E$  interaction) does not influence the comparison of varieties with respect to stability at least for the stability measures considered in the study. This will be clear from the following.

Stability comparisons may be made from  $\text{Var}((\alpha\beta)_{ij}) = \sigma_i^2$  such that lesser the  $\sigma_i^2$ , more is the stability. Let  $\text{Var}(\bar{e}_{ij}) = \frac{\sigma_e^2}{R} = \sigma_0^2$  (say) and  $\text{var}(v_{ij}) = \sigma_i^2$

Now  $\sigma_i^2 = \sigma_i^2 + \sigma_0^2$  [since  $(\alpha\beta)_{ij}$  and  $\bar{e}_{ij}$  are independent]

If the assumption of homogeneous error variance is correct, the genotype rank order given by stability variance  $\sigma_i^2$  will exactly be equal to that given by the interaction variance  $\sigma_i^2$ . The most stable genotype, i.e., the genotype with smallest interaction variance  $\sigma_i^2$ , will then also have the smallest stability variance  $\sigma_i^2$ , the genotype with second smallest  $\sigma_i^2$  will have the second smallest  $\sigma_i^2$ , and so on. Now it hardly matters if we compare the varieties based on  $\sigma_i^2$  or  $\sigma_i^2$ ; the stability rank order remains same. Therefore, the ranking ability of stability measures does not depend on the error related to the interaction genotype\* environment i.e. mean pure error.

With this examination of the invariant of true error component, we consider mean model for various parametric and non-parametric stability statistics that may be interpreted as measures of spread of  $v_{ij}$ 's.

### 2.2 Measures of Phenotypic Stability

The statistics to be described are based on the observed residuals

$$\hat{v}_{ij} = y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..}$$

where  $\bar{y}_{.i} = \sum_j \frac{y_{ij}}{N}$ ,  $\bar{y}_{.j} = \sum_i \frac{y_{ij}}{K}$  and  $\bar{y}_{..} = \sum_{i,j} \frac{y_{ij}}{KN}$

The first measure of this type of stability is the ecovalence proposed by Wricke [19]

$$W_i = \sum_{j=1}^N \hat{v}_{ij}^2$$

whereas Shukla's [18] unbiased estimator of  $\sigma_i^2$  is

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

This estimator is MINQUE (Minimum Norm Quadratic Unbiased estimator) of  $\sigma_i^2$  (Rao [17]). It should be noted that MINQUE is equivalent to  $W_i$  for ranking purposes.

Jaech [6] proposed a maximum likelihood estimator (MLE) for the variance of measurement errors, which is given by the following equations for ( $K > 3$ )

$$\hat{\sigma}_i^2 = \frac{(N-1)(b_{0i}b_{li} - b_{2i})}{Nb_{0i}^2} - \frac{1}{b_{0i}}$$

where

$$b_{0i} = \sum_{s \neq i}^K \frac{1}{\sigma_s^2}, \quad b_{li} = \sum_{s \neq i}^K \frac{v_{i-s}^2}{\sigma_s^2}, \quad b_{2i} = \sum_{s \neq i}^K \sum_{r > s}^K \frac{v_{s-r}^2}{\sigma_s^2 \sigma_r^2}$$

where

$$v_{s-r}^2 = \frac{1}{(N-1)} \left[ \sum_{j=1}^N x_{srj}^2 - \frac{\left( \sum_{j=1}^N x_{srj} \right)^2}{N} \right], \quad \text{where } x_{srj} = y_{sj} - y_{rj}$$

An iterative solution is obtained by assigning starting values to  $\sigma_1^2, \dots, \sigma_{K-1}^2$  (e.g. the MINQUE's) and solving the above system of equations for  $\sigma_K^2$ . Then with current estimates of  $\sigma_s^2$  ( $s$  unequal to  $i$ ) one computes estimates of the other  $\sigma_i^2$  ( $i < K$ ). This procedure is repeated until all estimates converge to a predetermined level of accuracy. This estimate has been suggested as an estimate of the stability variance in the analysis of stability (Piepho ([12], [13])).

In analogy to Wricke's ecovalence, Piepho and Lotito [15] proposed the sum of absolute values of residuals as a measure of stability

$$L_i = \sum_{j=1}^N |\hat{v}_{ij}|$$

Along with the above four measures, we may also consider the measures of stability derived from AMMI model by Raju [16] in a separate study, which may be computed from the  $K \times N$  matrix of residuals.

Let  $X$  be the matrix of interaction residuals obtained from ANOVA,  $X = [(v_{ij})]$  where  $v_{ij}$  is the interaction residual of  $i$ -th genotype in  $j$ -th environment,  $i = 1, \dots, K$  and  $j = 1, \dots, N$ .  $\lambda_1^2, \lambda_2^2, \dots, \lambda_m^2, \dots, \lambda_M^2$  be the positive eigenvalues of  $XX'$ ; where  $M = \text{rank}(XX')$  and  $\gamma_1, \gamma_2, \dots, \gamma_m, \dots, \gamma_M$  be the corresponding eigen vectors, where  $\gamma_M$  is a  $K \times 1$  vector, containing the PCA scores for the  $K$  genotypes corresponding to the axis  $M$ .

The measure based on the first PCA axis is  $FP_i = \lambda_1^2 \gamma_{1i}^2$ , where  $\gamma_{1i}$  is the  $i$ -th element of  $\gamma_1$ . This will be equivalent to the biplot with first PCA axis for ranking purposes.

The measure accommodating the first two PCA axes is  $B_i = \sum_{m=1}^2 \lambda_m^2 \gamma_{mi}^2$

This will be equivalent to biplot with first two PCA axes for the purpose of ranking of genotypes with respect to their stability.

The measure equivalent to Wricke's ecovalence is

$$W_{i(\text{AMMI})} = \sum_{m=1}^M \lambda_m^2 \gamma_{mi}^2$$

In addition to the above parametric measures of stability, there are some nonparametric measures for this type of study.

To compute Huhn's non-parametric stability measures, the residuals are ranked within environments. Let the resulting ranks (ranging from 1 to  $K$  within each environment) be denoted by  $r_{ij}$ . Then Gini's mean difference of ranks  $S_i^{(1)}$  and the rank variance  $S_i^{(2)}$  are computed as reported by Nassar and Huhn [10] and Huhn [5]

$$S_i^{(1)} = \frac{2 \sum_{j < j'}^N |r_{ij} - r_{ij'}|}{N(N-1)}, \quad S_i^{(2)} = \frac{\sum_{j=1}^N (r_{ij} - \bar{r}_i)^2}{(N-1)} \quad \text{where } \bar{r}_i = \frac{1}{N} \sum_{j=1}^N r_{ij}$$

Instead of ranking within environments, one may rank the residuals in the whole data set, which implies assigning ranks from 1 to  $NK$ . Gini's mean difference and variance of these ranks are denoted by  $P_i^{(1)}$  and  $P_i^{(2)}$  respectively (Piepho [12]).

Furthermore, the absolute values of residuals may be transformed to ranks. The rank sum is then a measure of variability of the residuals. If ranking is done within environments, the rank sum of  $i$ -th genotype is denoted by  $R_i$  (Piepho and Lotito [15]). For ranking over all residuals across environments, the rank sum is designated by  $P_i^{(3)}$  (Piepho [12]).

For clarity among the non-parametric measures, it may be useful to distinguish between the two groups

Group 1: Measures based on ranking of  $v_{ij}$  values within environments i.e.,  $S_i^{(1)}$ ,  $S_i^{(2)}$  and  $R_i$

Group 2: Measures based on ranking of  $v_{ij}$  values across the whole data set i.e.,  $P_i^{(1)}$ ,  $P_i^{(2)}$  and  $P_i^{(3)}$

### 3. Ranking Ability – Simulation

Since all the stability statistics described above measure the variability of the  $v_{ij}$ 's, they lend themselves to inferences on the parameter values of the stability variances  $\sigma_i^2$ . In practical situations, namely in plant breeding programmes, selection of the best genotypes is the prime objective. Selection involves ranking of genotypes. One may view the stability rank order of a set of genotypes as given by the rank order of the stability variances  $\sigma_i^2$ . To assess this stability rank order, the genotypes are grown in different environments. Here environments could be effects of different regions, different seasons, different years with further classification as wet and dry years, different rainfalls, different irrigations, different cultural practices etc. This will ensure more randomness in the component of environment with the final result of  $v_{ij}$ 's as random component. In statistical terms, we thus obtain realizations of the random effects  $v_{ij}$ 's, which are estimated to compute the stability measures, described above. For practical purposes, we seek the stability measure that best reflects the "true" stability rank order given by the rank order of the  $\sigma_i^2$ . This implies we seek the measure, for which the estimated rank order is close to the true rank order in a given set of genotypes. If we assign the rank order given by stability variance  $\sigma_i^2$  as true, we can compare the measures of stability described earlier with respect to their ability to assess the true stability rank order.

In practice,  $\sigma_i^2$  is unknown. In a Monte Carlo Simulation, however, we may assign defined values of  $\sigma_i^2$  to each genotype so that the true stability rank

order is known a priori. Now the concordance between the true stability rank order and the rank order displayed by the stability measure under consideration may be quantified by Spearman's rank correlation ( $r_s$ ) and the average of this rank correlation ( $r_s$ ) over 1000 runs of the Monte Carlo experiment is referred to as the ranking ability of the stability measure. Now, the ranking abilities of various stability measures discussed above may be compared under varying number of environments.

A common assumption in mixed linear models is that random effects are normally distributed. In the particular case at hand, this leads to the assumption that the distribution of  $v_{ij}$  effects is normal. In practice, however, it is difficult to verify this assumption, and there is some evidence that  $v_{ij}$ 's may at times be non normally distributed (Piepho [12]). In fact, statistical procedures for the stability variance are very sensitive to violations of the normality assumption (Brindley and Bradley [2]).

It is expected that the distribution of  $v_{ij}$  effects has an influence on the ranking ability of different measures of stability. This influence may be studied by means of Monte Carlo experiment, in which  $v_{ij}$  effects are generated from different distributions Piepho ([12], [14]). Now, on the basis of ranking ability, we may be able to identify the measures of stability that are robust to changes in parent distribution of  $v_{ij}$  effects.

#### 4. Non Normal Distributions used in the Study

The departure from normality may be quantified by the shape parameters  $\sqrt{\beta_1}$  and  $\beta_2$ , i.e. the skewness and kurtosis (Kendall and Stuart [8], 106). For normal distribution we have  $\sqrt{\beta_1} = 0$  and  $\beta_2 = 3$ . Skewed distributions are characterized by values of  $\sqrt{\beta_1} < 0$  or  $\sqrt{\beta_1} > 0$ . For long tailed distributions we have  $\beta_2 > 3$ . The departures of  $\sqrt{\beta_1}$  from zero and  $\beta_2$  from 3 may lead to many non-normal distributions.

Various non-normal distributions are taken for the study from Johnson's  $S_u/S_b$  system of distributions (Johnson [7]). The system is given by the following equations

$$z = \tau + \delta * \ln[y/(1-y)] \text{ for } S_b \quad \text{and} \quad z = \tau + \delta * \sinh^{-1}(y) \text{ for } S_u$$

If  $z$  is an  $N(0,1)$  variable,  $y$  follows an  $S_b$  or  $S_u$  distribution, respectively. Skewness  $\sqrt{\beta_1}$  and kurtosis  $\beta_2$  of the distribution depend on the parameters  $\tau$  and  $\delta$ . For a grid of  $\sqrt{\beta_1}$  and  $\beta_2$  values,  $\tau$  and  $\delta$  are tabulated in Pearson and Hartley [11].



The non-normal variate following  $S_b/S_u$  distribution may be obtained from the above equations as following

$$y = \frac{e^{\left(\frac{z-\tau}{\delta}\right)}}{1 + e^{\left(\frac{z-\tau}{\delta}\right)}} \text{ follows } S_b \text{ distribution; } 0 \leq y \leq 1$$

and

$$y = \sinh\left(\frac{z-\tau}{\delta}\right) \text{ follows } S_u \text{ distribution; } -\infty \leq y \leq \infty$$

One may also consider various contaminated normal distributions for generating the  $v_{ij}$  effects; whose probability density will be a mixture of two super-imposed normal distributions

$$f(x) = Wf_1(x) + (1-W)f_2(x)$$

where  $W$  is proportion associated with  $f_1(x)$ ;  $0 \leq W \leq 1$

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}, \text{ and } f_2(x) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \quad (\text{Cohen [3]})$$

The following contaminations may be distinguished

#### Location Contaminated Normal Distribution

This distribution may be characterized by  $\mu_1 \neq \mu_2$ ,  $\sigma_1 = \sigma_2$ . The expected value of the location contaminated variate  $X$  may be obtained as follows

$$\begin{aligned} E(X) &= \int xf(x)dx \\ &= \int x[Wf_1(x) + (1-W)f_2(x)] dx \\ &= \int [xWf_1(x) + x(1-W)f_2(x)] dx \\ &= \int xWf_1(x)dx + \int x(1-W)f_2(x) dx \\ &= W \int xf_1(x)dx + (1-W) \int xf_2(x) dx \\ &= W\mu_1 + (1-W)\mu_2 \end{aligned}$$

For a chosen  $W$  and  $\mu_1$  one can solve for  $\mu_2$  so that the contaminated variate has the zero expectation.

#### Scale Contaminated Normal Distribution

This distribution is characterized by  $\mu_1 = \mu_2$ ,  $\sigma_1 \neq \sigma_2$ . To ensure zero expectation for the contaminated variate, it is necessary to have  $\mu_1 = \mu_2 = 0$ . The variance of a scale-contaminated variate may be obtained as follows

$$\begin{aligned}
 V(X) &= E(X^2) = \int x^2 f(x) dx && (\because E(X) = 0) \\
 &= \int x^2 [Wf_1(x) + (1 - W)f_2(x)] dx \\
 &= \int [x^2 Wf_1(x) + x^2(1 - W)f_2(x)] dx \\
 &= \int x^2 Wf_1(x) dx + \int x^2(1 - W)f_2(x) dx \\
 &= W \int x^2 f_1(x) dx + (1 - W) \int x^2 f_2(x) dx \\
 &= W\sigma_1^2 + (1 - W)\sigma_2^2 && (\because \mu_1 = \mu_2 = 0)
 \end{aligned}$$

If we scale  $\sigma_1^2$  and  $\sigma_2^2$  such that  $W\sigma_1^2 + (1 - W)\sigma_2^2 = 1$ , the resulting contaminated variate will have zero mean and unit variance.

Contamination of a Normal Variate with a Variate following Altogether a Different Distribution.

Instead of contaminating a normal variate with another normal variate having different parameters, one may also consider altogether a different distribution (e.g. uniform) to contaminate normal variate. Furthermore, one may consider here  $W = 0$ ; i.e., a completely different distribution for  $v_{ij}$  effects. To visualize this, let us consider the uniform distribution for generating  $v_{ij}$  effects. If we generate a random variate following  $U(-\sqrt{3}S, \sqrt{3}S)$  its mean will be equal to zero and its variance will be equal to one.

The contamination of normal distribution with uniform distribution may be visualized as follows.

$$\text{Let } X \sim \begin{cases} N(0, S^2) \text{ with proportion } W \\ U(-S, S) \text{ with proportion } (1-W) \end{cases}$$

Obviously the mean of  $X$  will be

$$E(X) = W\mu_1 + (1-W)\mu_2 = 0$$

and the variance of  $X$  will be

$$V(X) = WS^2 + (1-W)S^2/3$$

One may scale this variance by using  $WS^2 + (1-W)S^2/3 = 1$

This relation may be used to determine  $S^2$  for a given proportion,  $W$ .

### 5. Design of Monte Carlo Experiment

Considering one at a time, the normal and various non-normal distributions described above for interaction effects, the true stability rank order for a set of genotypes (say  $K$ ) may be assigned as follows

$$\sigma_i^2 = D * i \quad (i=1, 2, \dots, K)$$

where D is a constant.

This assignment corresponds to symmetric distribution of  $\sigma_i^2$ 's.

One may also consider non-symmetric distributions for the  $\sigma_i^2$ 's by assigning

$$\sigma_i^2 = D * i^2 \quad (i = 1, 2, \dots, K)$$

The  $v_{ij}$  effects were simulated as random variates with an expected value of zero. The distribution of  $v_{ij}$ 's varied according to the list given in Table 1. For symmetric distribution of the  $\sigma_i^2$ 's, D was taken to be 4 and for non-symmetric distribution of the  $\sigma_i^2$ 's, D was taken to be 1.

**Table 1.** List of distributions used for generating effects,  $v_{ij}$  in the Monte Carlo experiment

S.No.	Distribution	$\sqrt{\beta_1}$	$\beta_2$	W	$\mu_1 - \mu_2$	$\sigma_1^2$	$\sigma_2^2$
1	Normal	0.00	3.0				
2	S <sub>b</sub>	1.15	3.0				
3	S <sub>b</sub>	0.00	1.1				
4	S <sub>u</sub>	2.00	15.0				
5	S <sub>u</sub>	1.00	7.0				
6	S <sub>u</sub>	0.50	5.0				
7	LCN			0.99	100		
8	LCN			0.9	10		
9	SCN			0.9		0.2	8.2
10	SCN			0.8		0.2	4.2
11	NCU			0.9		1.072	0.357
12	Uniform						

LCN = Location contaminated normal distribution.

SCN = Scale contaminated normal distribution.

NCU = Normal distribution contaminated with uniform distribution.

### 6. Results and Discussion

Ranking abilities of various measures of stability, given the normal distribution ( $\sigma_2^2 = 0$  and  $\beta_2 = 3$ ) of  $v_{ij}$  effects and symmetric distribution of  $\sigma_i^2$ 's, are presented in Table 2, from which one may conclude that the stability measure  $W_i / \text{MINQUE}/W_{i(\text{AMM})}$  is best in the terms of ranking ability which topped in all the 12 (K, N) combinations. Further critical examination of Table 2 leads to the following findings.

**Table 2.** Ranking abilities of various measures of stability given the normal distribution ( $\sqrt{\beta_1} = 0, \beta_2 = 3$ ) of  $v_{ij}$  effects and symmetric a priori distribution of  $\sigma_i^2$

K	N	$W_i$	MINQUE	MLE	$L_i$	$FP_i$	$B_i$	$W_{i(AMM)}$	$S_i^{(1)}$	$S_i^{(2)}$	$P_i^{(1)}$	$P_i^{(2)}$	$R_i$	$P_i^{(3)}$
10	5	0.670	0.670	0.593	0.658	0.487	0.607	0.670	0.441	0.445	0.531	0.540	0.606	0.608
10	10	0.795	0.795	0.764	0.777	0.566	0.713	0.795	0.600	0.617	0.699	0.715	0.716	0.723
10	15	0.848	0.848	0.827	0.835	0.606	0.751	0.848	0.702	0.714	0.783	0.792	0.774	0.784
10	20	0.880	0.880	0.865	0.868	0.631	0.785	0.880	0.757	0.765	0.828	0.834	0.812	0.822
20	5	0.701	0.701	0.667	0.688	0.474	0.610	0.701	0.502	0.514	0.550	0.563	0.631	0.637
20	10	0.811	0.811	0.799	0.797	0.532	0.676	0.811	0.673	0.689	0.718	0.731	0.736	0.741
20	15	0.856	0.856	0.847	0.846	0.557	0.718	0.856	0.754	0.765	0.792	0.802	0.793	0.799
20	20	0.881	0.881	0.876	0.871	0.584	0.740	0.881	0.795	0.803	0.829	0.835	0.823	0.828
30	5	0.702	0.702	0.681	0.691	0.463	0.603	0.702	0.517	0.531	0.552	0.566	0.636	0.640
30	10	0.810	0.810	0.803	0.797	0.514	0.661	0.810	0.691	0.708	0.720	0.735	0.742	0.747
30	15	0.856	0.856	0.851	0.847	0.533	0.694	0.856	0.767	0.780	0.791	0.802	0.797	0.801
30	20	0.886	0.886	0.882	0.877	0.569	0.722	0.886	0.817	0.824	0.837	0.843	0.831	0.835

As the number of environments and/or the number of genotypes increase, the difference in ranking ability between  $W_i/\text{MINQUE}/W_{i(\text{AMMI})}$  and MLE comes down and MLE approaches the former; which implies that with larger number of genotypes and/or environments, the selection between  $W_i/\text{MINQUE}/W_{i(\text{AMMI})}$  and MLE is not that much critical. With smaller number of genotypes or with smaller number of environments,  $L_i$  is found to be better than MLE standing at 2<sup>nd</sup> position after  $W_i/\text{MINQUE}/W_{i(\text{AMMI})}$ , which is evident from the table when we consider the case of  $K = 10$  or the case of  $N = 5$ . Now let us consider the stability measures derived from AMMI model. Except for the combination ( $K = 10, N = 5$ ) the measure  $FP_i$  was found to be inferior with least ranking ability to all the measures considered here for study. This implies that one should be cautious while using the biplot with first PCA axis for making stability conclusions. Also, the ranking abilities improved from  $FP_i$  to  $B_i$  and  $B_i$  to  $W_{i(\text{AMMI})}$  for all the combinations of  $K$  and  $N$ .  $W_{i(\text{AMMI})}$  is nothing but Wricke's ecovalence,  $W_i$  in terms of AMMI parameters and MINQUE is equivalent to  $W_i$  for ranking purposes. This is the reason behind the clubbing of  $W_{i(\text{AMMI})}$  with  $W_i$  and MINQUE and referring it as  $W_i/\text{MINQUE}/W_{i(\text{AMMI})}$ . Now let us compare various non-parametric stability measures among themselves given the normal distribution of  $v_{ij}$  effects.  $S_i^{(2)}$  is found to be superior to  $S_i^{(1)}$  and  $P_i^{(2)}$  is found to be superior to  $P_i^{(1)}$  in terms of ranking abilities with all the combinations of  $K$  and  $N$  which implies that rank variance is a better measure of stability as compared to Gini's mean difference of ranks. One may also be interested to compare the measures based on ranking within environments against the measures based on ranking across the environments. With all the combinations of  $K$  and  $N$ ,  $P_i^{(1)}$  is superior to  $S_i^{(1)}$ ,  $P_i^{(2)}$  is superior to  $S_i^{(2)}$  and  $P_i^{(3)}$  is superior to  $R_i$  in terms of ranking ability implying that the measures based on ranking across the environments are superior to the measures based on ranking within environments. We may also observe the following preference order among the measures as

$$S_i^{(1)} < S_i^{(2)} < P_i^{(1)} < P_i^{(2)}$$

The order of preference among the measures based on ranking within environments was found to be

$$S_i^{(1)} < S_i^{(2)} < R_i$$

Further to examine the difference between various stability measures for the situation of very large number of genotypes in contrast to not very number of environments ( $K = 50$  and  $N = 5$ ) normal populations were generated with ( $\sqrt{\beta_1} = 0$  and  $\beta_2 = 3$ ) of  $v_{ij}$  effects and symmetric distribution of  $\sigma_i^2$ 's. From these populations, ranking abilities of various measures of stability were obtained and are as under

$W_i$	MINQUE	MLE	$L_i$	$FP_i$	$B_i$	$W_{i(A\text{MMI})}$	$S_i^{(1)}$	$S_i^{(2)}$	$P_i^{(1)}$	$P_i^{(2)}$	$R_i$	$P_i^{(3)}$
0.720	0.720	0.709	0.705	0.474	0.602	0.720	0.541	0.560	0.562	0.579	0.647	0.652

If we compare the ranking abilities of different stability measures among themselves in the above situation it is seen that it is not necessary that the difference between various stability measures tends to fade away with larger number of genotypes and smaller number of environments. This is true at least for the stability measures considered in this study. The difference in ranking ability even in non-parametric measures is up to 0.111 (between  $S_i^{(1)}$  and  $P_i^{(3)}$ ) whereas the overall difference is up to 0.246 (between  $FP_i$  and  $W_i / \text{MINQUE} / W_{i(A\text{MMI})}$ ). The case where  $K = 30$ ,  $N = 5$  in Table 2 is also of this kind. If we see the corresponding differences in this case these are 0.123 and 0.239 respectively; not much change in difference with increase in number of genotypes from  $K = 30$  to  $K = 50$ . Rather the influence of number of environments is much higher to fade away the differences in ranking abilities of various stability measures, which is evident from Table 2.

The ranking abilities of various measures of stability given a normal distribution of  $v_{ij}$  effects and non-symmetric a priori distribution of  $\sigma_i^2$ 's are presented in Table 3. Ranking abilities of all the stability measures with all the combinations of  $K$  and  $N$  are improved as compared to that of Table 2. Here also  $W_i / \text{MINQUE} / W_{i(A\text{MMI})}$  tops in all the combinations of  $K$  and  $N$ , and the ranking ability of MLE approaches  $W_i / \text{MINQUE} / W_{i(A\text{MMI})}$  with the increase in the number of genotypes and environments; for (30,15) and (30,20) combinations of  $K$  and  $N$ , the ranking ability of MLE is same as that of  $W_i / \text{MINQUE} / W_{i(A\text{MMI})}$ . Shifting to non-symmetric a priori distribution for  $\sigma_i^2$ 's led to a rise in the ranking abilities only and there was no change in the conclusions drawn from Table 2.

Regarding the ranking ability, it is obvious from Table 2 and Table 3 that, with increase in the number of genotype and/or environments there is considerable improvement in the ranking ability of the stability measures and the choice of best performing stability measure seems less critical rather than the choice of the number of environments and genotypes. The poor ranking abilities of the stability measures with smaller number of environments and genotypes may be partly due to statistical errors in the stability estimates, which may be best reduced by including in the analysis as many genotypes and environments as possible. Another factor that influences the ranking ability is the distribution of stability variance  $\sigma_i^2$ . One may observe a sharp rise in the ranking ability of all the stability measures from Table 2 to Table 3. This reveals the fact that wider the distance between  $\sigma_i^2$ 's ( $i = 1, 2, \dots, K$ ) larger will be the ranking ability.

**Table 3.** Ranking abilities of various measures of stability given the normal distribution ( $\sqrt{\beta_1} = 0, \beta_2 = 3$ ) of  $v_{ij}$  effects and non-symmetric a priori distribution of  $\sigma_i^2$

K	N	$W_i$	MINQUE	MLE	$L_i$	$FP_i$	$B_i$	$W_{i(AMM)}$	$S_i^{(1)}$	$S_i^{(2)}$	$P_i^{(1)}$	$P_i^{(2)}$	$R_i$	$P_i^{(3)}$
10	5	0.866	0.866	0.828	0.858	0.701	0.820	0.866	0.614	0.627	0.721	0.733	0.809	0.815
10	10	0.928	0.928	0.915	0.921	0.733	0.869	0.928	0.787	0.809	0.859	0.872	0.877	0.882
10	15	0.950	0.950	0.944	0.945	0.758	0.885	0.950	0.848	0.862	0.906	0.913	0.910	0.915
10	20	0.962	0.962	0.959	0.959	0.767	0.897	0.962	0.885	0.895	0.931	0.935	0.929	0.933
20	5	0.873	0.873	0.850	0.866	0.681	0.811	0.873	0.665	0.684	0.720	0.736	0.823	0.827
20	10	0.930	0.930	0.924	0.923	0.714	0.845	0.930	0.829	0.847	0.862	0.876	0.887	0.890
20	15	0.948	0.948	0.947	0.944	0.732	0.861	0.948	0.880	0.892	0.905	0.913	0.913	0.917
20	20	0.961	0.961	0.960	0.957	0.751	0.875	0.961	0.910	0.919	0.928	0.935	0.932	0.934
30	5	0.874	0.874	0.863	0.866	0.679	0.801	0.874	0.692	0.711	0.728	0.745	0.823	0.826
30	10	0.928	0.928	0.924	0.922	0.706	0.834	0.928	0.836	0.853	0.857	0.873	0.886	0.889
30	15	0.951	0.951	0.951	0.947	0.728	0.857	0.951	0.893	0.903	0.908	0.917	0.918	0.920
30	20	0.961	0.961	0.961	0.957	0.738	0.866	0.961	0.918	0.925	0.930	0.936	0.932	0.934

Now, we may attempt to summarize the results obtained from various non-normal distributions in Table 4. A complete presentation of results and discussion for these distributions is given in the thesis of Raju [16]. Table 4 shows the stability measures with best ranking ability, given varying parent distribution of  $v_{ij}$  effects; symmetric a priori distributions of  $\sigma_i^2$ 's. With Johnson's  $S_b$  distribution for interaction, ranking ability of MLE is found to be superior. However one may prefer  $W_i / \text{MINQUE} / W_{i(\text{AMMI})}$  if the number of genotypes or environments is small. With Johnson's  $S_u$  distribution for interaction,  $L_i$  is found to be superior in terms of ranking ability. Most of the small samples drawn from location contaminated normal distribution with large  $(\mu_1 - \mu_2)$  and very large proportion ( $W$  approaching unity) behave like the samples drawn from a normal distribution with mean associated to  $W$ . In such cases  $W_i / \text{MINQUE} / W_{i(\text{AMMI})}$  is obviously expected to have the best ranking ability. However when the number of environments increases ( $N \geq 10$ ), the departure from normality, the characteristic of parent population, becomes apparent in which case the rank based measure  $P_i^{(3)}$  is found to have the best ranking ability exhibiting some kind of robustness to the distribution of interaction effects departing from normality. If  $W$  is not too large, one may prefer  $P_i^{(3)}$  for smaller number of environments and  $P_i^{(2)}$  for larger number of environments. With scale contaminated normal distribution for interaction, the stability measure  $P_i^{(3)}$  may be recommended. With uniform and normal contaminated with uniform distributions, which slightly depart from normality,  $W_i/\text{MINQUE}/W_{i(\text{AMMI})}$  is found to be superior as expected.

To conclude, given a normal distribution of interaction effects, it is best to estimate stability by the MINQUE of  $\sigma_i^2$  (or equivalently by  $W_i / W_{i(\text{AMMI})}$ ). The situation does not change dramatically under mild departures from normality. However, in view of the observations made in this study it is recommended that the  $v_{ij}$ 's realized from any real data should be tested for their normality before applying any stability measure (provided the number of environments is sufficiently large). There is enough literature available for testing normality (eg. Shapiro-Wilk's test). If the departure from normality is significant (extreme and longer tailed distributions) plant breeders may choose the robust measures like  $L_i$ ,  $P_i^{(2)}$  and  $P_i^{(3)}$  accordingly.



**Table 4.** Stability measures with best ranking ability for different distributions of  $v_{ij}$  effects; different  $K$  and  $N$ ; symmetric a priori distribution of  $\sigma_i^2$ 's ( $d = 4$ )

Verteilung number	Distribution	K = 10			K = 20			K = 30				
		N = 5	N = 10	N = 15	N = 5	N = 10	N = 15	N = 5	N = 10	N = 15	N = 20	
(1)	Normal	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA
(2)	$S_b$	WMA	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE	MLE
(3)	$S_b$	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA
(4)	$S_u$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$
(5)	$S_u$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$
(6)	$S_u$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$	$L_1$
(7)	LCN	WMA	$P_1^{(3)}$	$P_1^{(3)}$	WMA	$P_1^{(3)}$	$P_1^{(3)}$	WMA	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$
(8)	LCN	$P_1^{(3)}$	$P_1^{(2)}$	$P_1^{(2)}$	$P_1^{(3)}$	$P_1^{(2)}$	$P_1^{(2)}$	$P_1^{(3)}$	$P_1^{(2)}$	$P_1^{(2)}$	$P_1^{(2)}$	$P_1^{(2)}$
(9)	SCN	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$
(10)	SCN	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$	$P_1^{(3)}$
(11)	NCU	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA
(12)	Uniform	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA	WMA

Where  $WMA = W_i / \text{MINQUE} / W_{i(A,MMI)}$

## REFERENCES

- [1] Becker, H.C. and Leon, J. (1988). Stability analysis in plant breeding. *Plant Breeding*, **101**, 1-23.
- [2] Brindley, R.D. and Bradley, R.A. (1985). Some new results on Grubbs' estimates. *J. Amer. Statist. Assoc.*, **80**, 711-714.
- [3] Cohen, A.C. (1967). Estimation in mixtures of two normal distributions. *Technometrics*, **9**, 15-28.
- [4] Finlay, K. and Wilkinson, G.N. (1963). The analysis of adaptation in a plant-breeding programme. *Austr. J. Agril. Res.*, **14**, 742-754.
- [5] Huhn, M. (1990). Non parametric measures of phenotypic stability. Part 1: Theory. *Euphytica*, **47**, 189-194.
- [6] Jaech, J.L. (1985). *Statistical Analysis of Measurement Errors*. Wiley, New York.
- [7] Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149-176.
- [8] Kendall, M.G. and Stuart, A. (1987). *Advanced Theory of Statistics*. 1: Distribution Theory. Griffin, London.
- [9] Lin, C.S., Binns, M.R. and Levkovitch, L.P. (1986). Stability analysis: Where do we stand? *Crop Science*, **26**, 894-900.
- [10] Nassar, R. and Huhn, M. (1987). Studies on estimation of phenotypic stability: Tests of significance for nonparametric measures of phenotypic stability. *Biometrics*, **43**, 45-53.
- [11] Pearson, E.S. and Hartley, H.O. (eds) (1976). *Biometrika Tables for Statisticians*. II. Biometrika Trust, London.
- [12] Piepho, H.P. (1992). Vergleichende Untersuchungen der statistischen Eigenschaften verschiedener stabilitätsmasse mit Anwendungen auf Hafer, Winterraps, Ackerbohnen sowie Futter- und zuckerriiben. Doctoral thesis, Kiel.
- [13] Piepho, H.P. (1993). Use of the maximum likelihood method in the analysis of phenotypic stability. *Biom. J.*, **35**, 815-822.
- [14] Piepho, H.P. (1994). Ranking ability of various measures of phenotypic stability. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie*, **25**, 172-180.
- [15] Piepho, H.P. and Lotito, S. (1992). Rank correlation among parametric and non parametric measures of phenotypic stability. *Euphytica*, **64**, 221-225.
- [16] Raju, B.M.K. (2002). On some statistical aspects of assessing sensitivity of crop varieties. Unpublished Ph.D. Thesis, P.G. School, IARI, New Delhi.
- [17] Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.*, **65**, 161-172.
- [18] Shukla, G.K. (1972). Some statistical aspects of partitioning genotype environmental componenets of variability. *Heredity*, **29**, 237-245.
- [19] Wricke, G. (1965). Die Erfassung der Wechselwirkungen zwischen genotyp und umwelt bei quantitativen Eggenenschaften. *Zeitschrift für pflanzenzuchtung*, **53**, 266-343.