***Pak. J. Bot.*, 41(2): 711-730, 2009.**

NONPARAMETRIC METHODS IN COMBINED HETEROSCEDASTIC EXPERIMENTS FOR ASSESSING STABILITY OF WHEAT GENOTYPES IN PAKISTAN

SYED HAIDER SHAH[1], SYED MUNAWAR SHAH[2], M. INAYAT KHAN[3], MAQSOOD AHMED[4], IMDAD HUSSAIN[5] AND K. M. ESKRIDGE[6]

[1]Department of Statistics, University of Balochistan, Quetta, Pakistan
[2]Department of Economics, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan
[3]Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan
[4]Department of Environmental Management and Policy, BUITEMS, Quetta, Pakistan
[5]Department of Business administration, Iqra University, Karachi, Pakistan
[6] Department of Statistics, University of Nebraska, Lincoln, USA

Abstract

Genotype performance in multienvironment trials (METs) are usually analyzed by parametric analysis of variance (ANOVA) and stability models. The results of these models can give misleading inferences when some sensitive assumptions are not satisfied. In this paper, assumptions of combined ANOVA are scrutinized in detail to justify the validity of use of 5 nonparametric stability methods ($S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$, $S_i^{(6)}$ and $YS_i^{(1)}$, $YS_i^{(2)}$) applied to 20 genotypes grown in 40 hetroscedastic and nonnormal environments in Pakistan for the year 2004-05. There is a severe heterogeneity problem in the data because the ratio of the largest estimated mean squares error (MSE) for individual environments randomized complete block design (RCBD) to the smallest MSE is approximately (1.00/0.02=50). Out of 40 environments individual coefficients of determination (R^2), 27 are less than 0.70. This leads to violation of linearity assumption in the model. Standardized residual plots vs. individual environments plots and normal probability plot are indicators of the violation of homogeneity, normality assumptions and absence of outliers. No linear relationship was established between the natural logs of the error variance and the natural log of environments' mean, which again violates coefficient of variation (CV) assumption. Remedial transformations as suggested in literature were not successful to stabilize environments MSEs and could not normalize the data, so as a last resort in this regard nonparametric stability methods seem to justify the analysis of genotype x environment interactions (GEI). The low values of modified rank-sum statistics $YS_i^{(1)}$ and $YS_i^{(2)}$ were positively and significantly associated with mean yield but the other nonparametric methods were not correlated with mean yield. The results of principal component analysis and correlation analysis of nonparametric stability methods indicate that the use of modified rank-sum method would be justifiable for simultaneous selection for high yield and stability. Using modified rank-sum method, the genotypes G7, G3, G15, G5 and G12 were found to be the most stable with yield, whereas G14 and G19 were the least stable genotypes.

Introduction

One of the most challenging issues in plant breeding process to accurately analyze genotype x environment interaction (GEI) is based on data from multienvironment trials. GEI is a universal issue that relating to all living organisms, from humans to plants and bacteria (Kang, 1998). Usually GEI is the nonadditive component of two or more experiments with the same genotypes combined over environments. The process for selecting high yield and stable genotypes usually involves three stages of experimentation: At stage-1, genotypes are tested at a single location; at stage-2, the selected genotypes are tested in a multilocation trials (genotype x location); and finally at stage-3, the most promising genotypes with new set of genotypes are tested for several years under a range of locations (genotype x location x year) (Linn & Binns, 1994).

Corresponding author Email: mrhaidershah@yahoo.com

Genotype performance trials are usually analyzed by various ANOVA models which are based on assumptions that may often not be satisfied. Departure from one or more assumptions can affect both the type-1 error and the sensitivity of F or t tests. Excellent and detailed discussions about assumptions, the consequences their invalidates and remedial steps involved for ANOVA are described by some classic papers (Eisenhart, 1947; Cochran, 1947).

The primary goal of most plant breeders is to identify stable genotypes, whose yield performance remains high across a range of environmental conditions. Stability analysis is only relevant if GEI is present (Hussain *et al*., 2000). Basically there are two broad categories of GEI: crossover and noncrossover (usual) interaction. A crossover interaction (discordance) exists if the ranking of the genotypes is not identical in different environments. If the ranking is identical, crossover interaction is nonexistent (concordance) (de Kroon & Laan 1981; Truberg & Huehn, 2000). Measures of GEI and stability are common tools applied by biometricians who have developed numerous methods to analyze it (Lin *et al*., 1986; Becker & Leon 1988; Flores *et al*., 1998; Mohammadi & Amri, 2008).

Huehn (1996) indicated that there are two major approaches for studying GEI to determine the adaptation characterization of genotypes. The first and most common approach is parametric, which relies heavily on distributional assumptions about genotypic, environmental, and G x E effects. The second approach is the nonparametric approach (rank-based methods), based on weak assumptions. Numerous univariate and multivariate (parametric and nonparametric) methods have been developed by statisticians and applied by plant breeders to analyze GEI at the end of plant breeding programs.

The univariate parametric stability statistics are commonly used by plant breeders to analyze GEI. The essential ideas of stability analysis are described in Lin *et al*., (1986); Hussain *et al*., (2000) and Backer & Leon (1988) are concerned with describing as to how a genotype responds to differing environmental conditions. The parametric stability methods have good properties under statistical assumptions of a normal distribution of independent errors with homogenous variance and no outliers. However, many of these measures may not perform well if any or all of these assumptions are violated, specifically the assumptions of homogeneity of mean square errors (MSEs), nonmixture of normal distributions and data having no outliers (Huehn, 1990). Due to GEI the relative differences among genotypes render performance over environments inconsistent. MSEs are rarely homogeneous in multienvironment or regional yield trials. MSEs are influenced by specific circumstances and tend to be lower in low yielding environments (Bowman & Watson, 1997).

Thus it is advisable to search for alternative approaches such as nonparametric methods that are more robust and are valid in the absence of variance homogeneity, no outliers and normality assumptions. The importance of nonparametric methods in modern statistics has been growing dramatically since their inception in the mid-1930s. Requiring few or no assumptions about the populations from which the data is obtained, nonparametric methods emerged as a very useful methodology among statisticians and researchers performing data analyses. Today, these techniques are being applied to an ever-growing variety of experimental designs in the social, behavioral, biological and physical sciences (Hollander & Wolfe, 1999). These methods are based on the ranks of original observations or residuals of combined ANOVA model.

Huehn (1979) and Nassar & Huehn (1987) proposed four nonparametric indices of stability which are: (1) $S_i^{(1)}$ is the mean of absolute rank differences of a genotype over environments, (2) $S_i^{(2)}$ is the variance among the ranks over the environments, (3) $S_i^{(3)}$ and $S_i^{(6)}$ are the sum of the absolute deviations and sum of squares of ranks for each genotype relative to the mean of ranks respectively.

Kang (1988) proposed the rank sum method which integrates yield and Shukla's stability unbiased variance in one index (Shukla, 1972). In this study, method used for selecting high yielding and stable genotypes was the consideration of yield and Huehn's first two stability statistics ($S_i^{(1)}$ and $S_i^{(2)}$) are used as selection criteria. Based on this method, ranks are assigned for mean yield, with the genotype with highest yield receiving the rank of I. Similarly the ranks are assigned for $S_i^{(1)}$ and $S_i^{(2)}$ with lowest estimated value receiving the rank of I and they are denoted by $YS_i^{(1)}$ and $YS_i^{(2)}$. The two ranks for each genotype are summed and the lowest rank sum is regarded as most stable.

According to Huehn (1990) the nonparametric procedures have a number of advantages over parametric stability methods e.g., they reduce the bias caused by outliers, no assumptions are needed about the distribution of the observed values, they are easy to use and interpret, and additions or deletions of one or few genotypes do not cause much variation in the results. There is theoretical justification for the use of nonparametric methods in the assessment of yield stability analysis of combined heteroscedastic and nonnormal ANOVA. The parametric procedures are not robust specifically when the distribution of data is nonnormal, heteroscedastic and mixture of normal distributions. It is a known fact that the parametric methods have large power values than their nonparametric counterparts when all classical assumptions hold. But the adequate application of parametric methods need fulfillment of some strict statistical assumptions. Interestingly, there has been no report of using nonparametric stability methods applied on multienvironmental crop data analyzing GEI in Pakistan. Most stability analysis methods are based on joint linear regression of genotype yield on an environmental index derived from the average performance of all genotypes in an environment (Finlay & Wilconsin, 1963; Eberhart & Russell, 1966) used by Imtiaz *et al*., (1988); Khan *et al*., (1988); Sial *et al*., (2000) and Javaid *et al*., (2006).

This paper attempts to diagnose as thoroughly as possible the assumptions of parametric combined ANOVA by using plots and statistical tests, with a view to justifying the use of nonparametric stability statistics and the relationship among nonparametric stability methods and to identify the most stable and the least preferred wheat genotypes by using 5 nonparametric stability methods and graphs.

Materials and Methods

Statistical analysis was performed on grain yield data of 20 wheat genotypes collected from 40 heteroscedastic environments in Pakistan for the year 2004-05. The National Uniformity Wheat Yield Trials (NUWYT) was conducted at various agro climatic regions of the country and coordinated by National Agricultural Research Centre, Wheat Program, Islamabad, Pakistan. The descriptive statistics of the environments are given in Table 1.

Table 1. Mean square error (MSE), mean, coefficient of variation, coefficient of determination, superiority measure P_j, maximum, minimum and range, for the wheat data across 40 heteroscedastic environments.

Code	Environments	MSE	Mean	CV (%)	R^2 (%)	P_j	MAX	MIN	Range
E1	Quetta	0.54	3.19	23.13	42.78	0.33	3.88	2.17	1.71
E2	Dhadar Bolan	1.00	3.59	27.81	39.18	0.99	4.84	2.58	2.26
E3	Tandojam	0.09	4.72	6.50	92.35	1.22	6.00	2.81	3.19
E4	Dadu	0.04	1.92	10.37	59.04	0.12	2.38	1.60	0.78
E5	Sanghar	0.14	4.45	8.33	92.73	3.18	6.72	3.18	3.54
E6	Sakrand	0.43	4.00	16.40	60.44	0.50	4.75	2.54	2.21
E7	Larkana	0.08	3.49	8.19	68.46	0.24	4.09	2.96	1.13
E8	Sukkur	0.16	2.67	14.89	63.95	0.26	3.27	1.75	1.52
E9	R.Y. Khan1	0.22	4.54	10.31	69.89	0.84	5.69	2.99	2.70
E10	R.Y. Khan2	0.08	3.94	7.30	78.05	0.53	4.86	2.99	1.87
E11	Bahawalpur	0.14	4.39	8.39	80.78	0.47	5.13	2.83	2.30
E12	Haroonabad	0.16	3.30	11.95	89.09	1.67	4.86	1.24	3.62
E13	Lodhran	0.33	2.98	19.39	39.40	0.55	3.96	1.94	2.02
E14	Multan	0.43	3.33	19.70	45.06	0.59	4.29	2.21	2.08
E15	Vehari	0.12	3.16	11.14	64.18	0.14	3.54	2.00	1.54
E16	Khanewal1	0.09	2.73	11.01	73.86	0.40	3.52	1.73	1.79
E17	Khanewal2	0.10	2.84	11.17	64.14	0.40	3.67	1.90	1.77
E18	Layyah	0.26	3.03	16.90	42.59	0.14	3.46	2.21	1.25
E19	Bhakkar	0.13	4.25	8.44	82.04	1.30	5.73	3.15	2.58
E20	D.G. Khan	0.10	3.83	8.19	82.47	0.41	4.55	2.43	2.12
E21	Muzaffargarh	0.09	4.08	7.38	79.95	0.40	4.83	3.02	1.81
E22	Piplan	0.16	3.03	13.20	54.50	0.34	3.77	2.13	1.64
E23	Jhang	0.13	4.33	8.31	83.93	0.93	5.52	3.17	2.35
E24	Gojra	0.05	3.43	6.53	73.06	0.18	3.95	2.75	1.20
E25	Faisalabad1	0.21	4.57	10.13	69.06	0.42	5.29	3.30	1.99
E26	Faisalabad2	0.07	4.40	6.03	72.67	0.13	4.76	3.39	1.37
E27	Yousaf Wala	0.16	4.76	8.28	79.25	0.68	5.73	2.83	2.9
E28	Kasur1	0.18	4.58	9.34	75.51	0.80	5.69	2.71	2.98
E29	Kasur2	0.20	4.10	11.04	57.90	0.32	4.77	3.15	1.62
E30	Sheikhupura	0.12	4.16	8.30	63.13	0.22	4.71	3.33	1.38
E31	Sargodha	0.11	2.44	13.85	70.91	0.30	3.10	1.75	1.35
E32	Pindi	0.30	2.62	21.04	36.87	0.17	3.13	2.06	1.07
E33	Gujranwala	0.10	3.71	8.350	72.21	0.24	4.27	2.58	1.69
E34	Bhimber	0.02	3.90	3.71	52.81	0.06	4.24	3.65	0.59
E35	Islamabad	0.16	2.30	17.57	60.25	0.21	2.87	1.62	1.25
E36	D.I. Khan	0.34	3.10	18.69	43.87	0.26	3.69	1.83	1.86
E37	Tarnab	0.26	3.75	13.65	65.65	0.51	4.58	2.22	2.36
E38	Pirsabak	0.51	3.61	19.68	53.28	0.79	4.70	2.47	2.23
E39	Mardan	0.33	4.08	14.11	62.84	0.66	5.07	2.99	2.08
E40	Charsada	0.34	5.14	11.39	74.92	1.68	6.78	3.53	3.25

This study was conducted using 20 wheat genotypes, i.e. G1(V-01078), G2(99B4012), G3(Wafaq 2001), G4(RWM-9313), G5(V-00125), G6(DIAMOND), G7(PR-84), G8(TW 0135), G9(V-00055), G10(99B2278), G11(KT-7), G12(V-01180), G13(DN-47), G14(V-9021), G15(CT-00062), G16(7-03), G17(PR-86), G18(V-02192), G19(V-002493) and G20(L. Check). These genotypes were developed by various plant breeders at different research institutes/stations of Pakistan. The most widely grown bread wheat genotype Wafaq 2001 was included as National check alongwith one local check at each environment. Of 20 genotypes used, 10 were from Punjab, 5 from NWFP, 3 from Sindh provinces, one from National Agricultural Research Centre and one local check was a variety already being used in that environment by the farmers. At each environment 20 genotypes were grown in a randomized complete block design (RCBD) with 4 replications. Grain yield was obtained by expressing plot grain yields and converting them on hectare basis (t ha^{-1}). In this paper superiority measure (Linn & Binns, 1988) was used to assess the stability of environment's yield performance and identify those environments having outliers. Superiority measure (P_j) was defined as the distance mean square between the environment's response and the maximum response over genotypes. Bartlett's, Hartley's and Cochran's tests and standardized residuals vs. environment plots, as described by Snedecor & Cochran (1980) Zar, (1996) and Kuehl, (2000), were performed to asses the homogeneity of variances prior to using combined analysis. If the homogeneity tests and plots show that MSEs are severely heterogeneous the validity and accuracy of using the classical F-test of GEI is in question (Yates & Cochran, 1938). Several diagnostic plots are recommended by statisticians for detecting failures in the assumptions (unusual observations, nonnormality, heterogeneity) and other inadequacies in the combined ANOVA. Examination of the assumptions through statistical tests and residual plots should be an automatic part of any ANOVA before using parametric stability methods, ignored by the researchers in most of the genotype performance trails. In the combined experiment analysis, genotypes and environments were considered fixed while blocks assumed random.

A combined two factor linear analysis of variance model was used on the data:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \rho_{K(j)} + \varepsilon_{ijk} \quad i = 1,2, \ldots ,g \ ; \quad j = 1,2, \ldots , e \ ; k = 1,2, \ldots , r$$

where Y_{ijk}; observation on the ith genotype in the jth in the kth block, μ; overall mean, α_i; fixed effect of genotype i, β_j; fixed effect of environment j, $\alpha\beta_{ij}$; fixed interaction effect of genotype i and environment j, $\rho_{k(j)}$; random block within environment effect, ε_{ijk}; experimental error are assumed to be randomly distributed with zero mean and homogeneous variance σ^2. If these assumptions hold true, the model may be analyzed by combined ANOVA. Such an ANOVA is naturally the first step if we want to find out if GEI exists in combined experiment.

Nonparametric stability analyses were performed to identify the stable and unstable genotypes across heteroscedastic environments. Five nonparametric stability statistics were used to assay the combined data. Four of these were proposed by Huehn, (1979) and Nassar & Huehn (1987), which combine mean yield and stability. The data was arranged in a two way table with *g* rows (genotypes) and *e* columns (environments). We denote r_{ij} be the rank of ith genotype in jth environment and r_i as the mean rank across all environments for the ith genotype under the null hypothesis of maximum stability. Huehn stability statistics was computed by using the ranks r_{ij} of the aligned observations ($Y^*_{ij} = Y_{ij} - Y_i - Y$), where, Y_i is the mean yield of the ith genotype, Y_{ij} is the yield of ith genotype in the jth environment and $Y_{..}$ is the grand average across all environments. The correction has been proposed to get independence from genotypic effects, i.e., to reflect the rank of interaction effect plus experimental error (Nassar & Huehn, 1987). The statistics based on aligned yield ranks of genotypes in each environment are expressed as follows:

$$S_i^{(1)} = 2\sum_{j}^{e-1} \sum_{j'=j+1}^{e} \left| r_{ij} - r_{ij'} \right| \Big/ \left[e(e-1) \right]$$

$$S_i^{(2)} = \sum_{j=1}^{e} \left(r_{ij} - \bar{r}_{i.} \right)^2 \Big/ (e-1)$$

$$S_i^{(3)} = \frac{\sum_{j=1}^{e} \left(r_{ij} - \bar{r}_{i.} \right)^2}{\bar{r}_{i.}}$$

$$S_i^{(6)} = \frac{\sum_{j=1}^{e} \left| r_{ij} - \bar{r}_{i.} \right|}{\bar{r}_{i.}}$$

The lowest value for each of these statistics indicates maximum stability for the genotype. According to Huehn (1990), the use of $S_i^{(1)}$ was preferred to $S_i^{(2)}$ for many practical considerations reported to be easy to calculate, interpret and is an efficient for test of significance. These statistics do not require homogeneity of variances assumption and was developed for determining stability of genotypes in nonnormal and heteroscedastic data. In this study, few values were missing for certain genotypes at certain environments. Use of nonparametric stability statistics is theoretically robust for missing values. Missing value was replaced with the average rank of the genotype for the environments (Huehn, 1990).

Nassar & Huehn (1987) and Huehn & Nassar (1989) proposed two tests of significance for $S_i^{(1)}$ and $S_i^{(2)}$ The explicit formulae for means $E(S_i^{(k)})$ and variances $var(S_i^{(k)})$ are respectively:

$$E\left\{S_i^{(1)}\right\} = \frac{g^2-1}{3g} \qquad Var\left\{S_i^{(1)}\right\} = \frac{(g^2-1)\left[(g^2-4)(e+3)+30\right]}{45g^2 e(e-1)}$$

$$E\left\{S_i^{(2)}\right\} = \frac{g^2-1}{12} \qquad Var\left\{S_i^{(2)}\right\} = \frac{(g^2-1)}{36e}\left[\frac{g^2-4}{5} + \frac{g^2-1}{2(e-1)}\right]$$

The approximate statistical tests of significance and properties of $S_i^{(1)}$ and $S_i^{(2)}$ were developed from Nassar & Huehn, (1987) and Huehn & Nassar (1989; 1991) based on the normal distribution under the null hypothesis of no genotype environment interaction effects. The statistic $Z_i^{(K)} = [Si^{(k)} - E(S_i^{(k)})]^2/var(S_i^{(k)})$ and here k = 1, 2; has an approximate χ^2 distribution with 1 degrees of freedom and the statistic $S_i^{(k)} = S_i^{(k)} = \sum_{i=1}^{g} Z_i^{(k)}$, k = 1,2; may be approximated by a χ^2distribution with *g* degrees of freedom.

Kang's (1988) rank-sum-method as modified by Yue *at el.*, (1997) is the fifth nonparametric stability statistic which selects simultaneously a genotype with high and stable yield across environments. The modified rank-sum method in which both yield (in ranks) and first two Huehn (1987) nonparametric stability statistics $S_i^{(1)}$ and $S_i^{(2)}$ (in ranks) are combined, was used to evaluate the characteristics of 20 wheat genotypes. These two rank-sum statistics $YS_i^{(1)}$ and $YS_i^{(2)}$ computed as Rank-sum (yield rank + ranks of the rank-based stability statistic, $S_i^{(1)}$ and $S_i^{(2)}$). This method assumes equal weight for yield and stability statistics for $YS_i^{(1)}$ and $YS_i^{(2)}$. The modified rank-sums provide a compromise statistic when both yield and stability are considered essential in a breeding programme. However, plant breeders may prefer to assign more weight to yield than to stability statistics.

The stability statistics were compared using Spearman's rank correlation coefficient. Spearman's rank correlation coefficient, as calculated from the ranks of 5 nonparametric stability statistics results in measuring the linear relationship between the 5 methods. Principal component analysis (PCA) method was used for stratifying nonparametric stability methods and genotypes. The combined experimental yield data were statistically analyzed using SPSS version 15.0 (2006) for plots and correlation matrix and MS EXCEL (version 2003) for nonparametric stability methods with spreadsheet formulae commands.

Results and Discussion

In this study at each environment, a randomized complete block design (RCBD) with 4 replicates was separately analyzed before pooling the parametric combined experiment. The environments MSEs results were obtained across 40 different environments for yield variable ranged from 0.02 to 1.00. The most noticeable feature of the separate analyses was that the largest mean squares error (MSE) at environment 2 is 1.00 and the smallest one as computed for environment number 34 was 0.02. This is indicator of wide variation. Out of 40 MSE values, 31 range between 0.02 and 0.30, while the remaining are 0.33, 0.33, 0.34, 0.34, 0.43, 0.43, 0.51, 0.54 and 1 (Table 1). The ratio of the largest MSE to the smallest MSE is (1.00/.02=50). There were wide and severe variations among MSEs across heteroscedastic environments, which explicitly pinpoint that the experimental conditions differ among environments and also indicate that when combined for parametric stability analysis the experimental data becomes heteroscedastic and thus we have problem in data. Rogan & Kaselman (1977) have reported that the rate of type 1 error varies as a function of degree of variance heterogeneity and, consequently it should not be assumed that the ANOVA F-test is always robust to variance heterogeneity even when design is balanced. Cochran (1947) and Eisenhart (1947) reported that nonhomogeneous variances had a greater impact on the estimates of treatments effects and their variances than on tests of significance. Improper blocking usually inflates experimental variance (Warren & Mendez, 1982) which may result in acceptance of false null hypothesis or rejection of true null hypothesis. Experimental MSEs (mixture of various levels of variations) across 40 environments were found to be heterogeneous even at $p< 0.001$ using Bartlett's test, Cochran's test and Hartley's test (Table 2). The Z score (standardized residual) method establishes rule-of-thumb limits outside of which an observation is deemed to be an outlier i.e., according to standard normal distribution criterion observations with Z scores greater than 3 in absolute values are considered outliers (improbable). For some highly skewed multienvironment data sets, observations with Z scores greater than 1 in absolute value may be outlier (Rousseeuw & Leroy 1987). Figure 1 explicitly indicates severe heteroscedastic problem across environments along with ample outliers. However, by their nature majority of multienvironmental experimental MSEs were heterogeneous.

Table 2. Bartlett, Hartley and Cochran tests for homogeneous MSEs of 40 individual environments with 4 blocks and 20 genotypes.

Variance homogeneity tests	Untransformed = Yield (t ha^{-1})	Transformed = log Yield (t ha^{-1})
Bartlett	χ^2= 633.24**	χ^2= 812.31**
Hartley	H^a = 50**	H = 58.33**
Cochran	C^b = 0.12**	C = 0.12**

** Significant at p = 0.01

[a] H = (Largest of 40 MSEs) / (Smallest of 40 MSEs)

[b] C = (Largest MSE) / (ΣMSE_j)

Transformation approach was used to decide the appropriate transformation type (power or root) of response variable Y^{λ} to stabilize MSEs and normalize the data (Box & Cox, 1964). In our study using Box-Cox transformation method the maximum likelihood estimate was $\hat{\lambda}$=0.12 (or $\hat{\lambda}$ close to 0) which suggests and that transformed response variable (yield) is either logarithmic or $Y^{0.12}$. Results also suggest that after transformation Bartlett's test, Cochran's test and Hartley's test values further increased (Table 2). But both of these transformations were not successful in sufficiently stabilizing MSEs and bringing the distribution of the observations close enough to normality to meet the robustness properties of the parametric inference procedures. It was therefore necessary that as a last resort we make use of nonparametric stability methods were for analyzing GEI. Bradley (1982) stated that if the response variable consists of mixture of normal distributions or there is major departure from model's assumptions, then such transformation will not be helpful. In this study, the other reason of failure of transformation is that the combined experimental data has no patternoized relationship between environments' mean and variance (MSE) as shown in Fig. 3.

The coefficient of determination (R^2), coefficient of variation (CV), superiority measure (P_j) and range values (Table 1) and plots shown in Figs. 1 & 2 clearly indicate that combined experimental data is problematic (hetroscedastic, nonnormal and ample outliers). Data have been drawn from populations that are heterogeneous, mixture of normal distributions, skewed, or have more observations in their tails than is true of the normal distribution (heavy-tailed distributions). According to Linn & Binns (1988) the modified stability measure (P_j) of the environments E34, E4, E26 E15, E18, E32 and E24 with low (P_j) values indicated high relative stability and these environments had lowest yield performance except E26. Whereas E5, E40, E12, E19, E3, E2 and E23 with high (P_j) values revealed that these environments were problematic and had highest average yields except for E12. All environments had irregular hetroscedasticity irrespective of high or low mean (Table 1).

Agricultural scientists use the coefficient of variation (CV) as a measure of population variability. They use the CV to accept or reject the validity of trials. Use of CV is valid only when the simple regression coefficient *b* value of the regression of the natural log of the environments' MSE on the natural log of the environments' mean equals 2.0 (Bowman & Watson, 1997). In the proposed study of 40 hetroscedastic environment's data sets of wheat crop data revealed no linear relationship i.e., because the estimate of b=0.014 between natural log of MSE and natural log of the mean, which bring the use of the CV for checking validity of crop performance trials into question. Again use of CV may affect the validity of trials due to heterogeneity of MSEs or because of existence of outliers in individual RCBD data across heterogeneous environments.

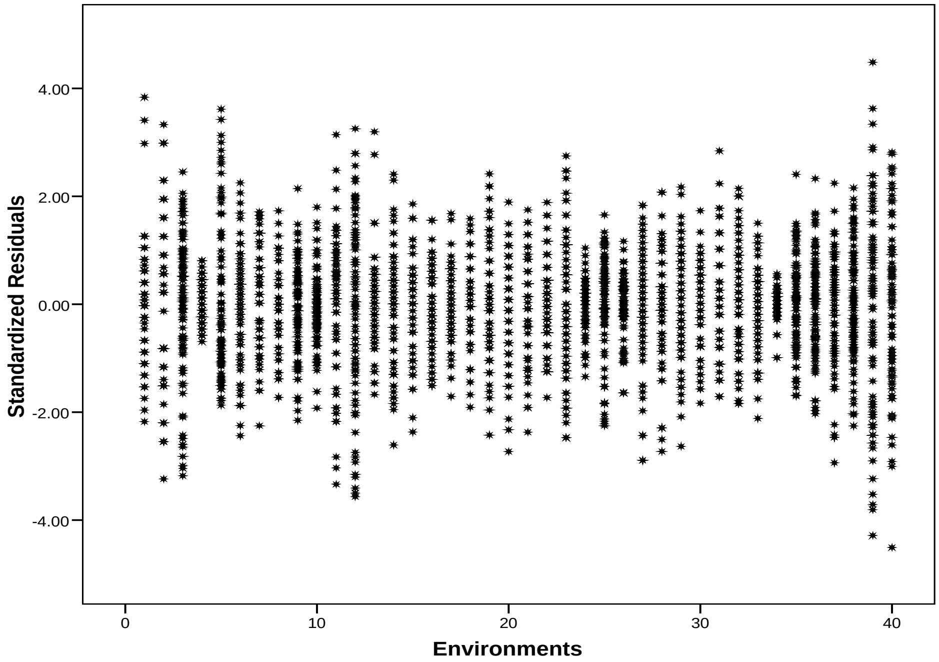

Fig. 1. Standardized residuals against environments depicting heterogeneity among environments with ample outliers.

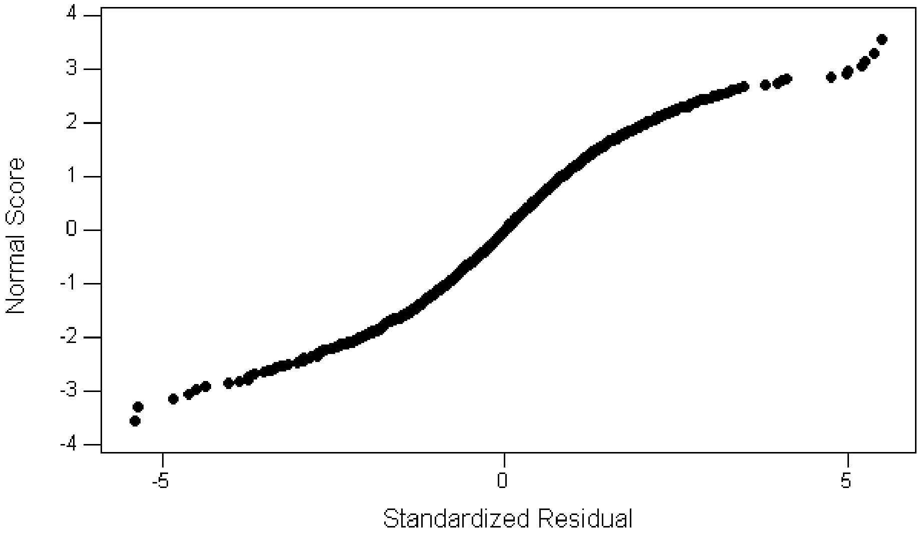

Fig. 2. Normal probability plot showing major violation of the normality assumption and indicating the mixture of normal distributions with outliers (Minitab 13.20).

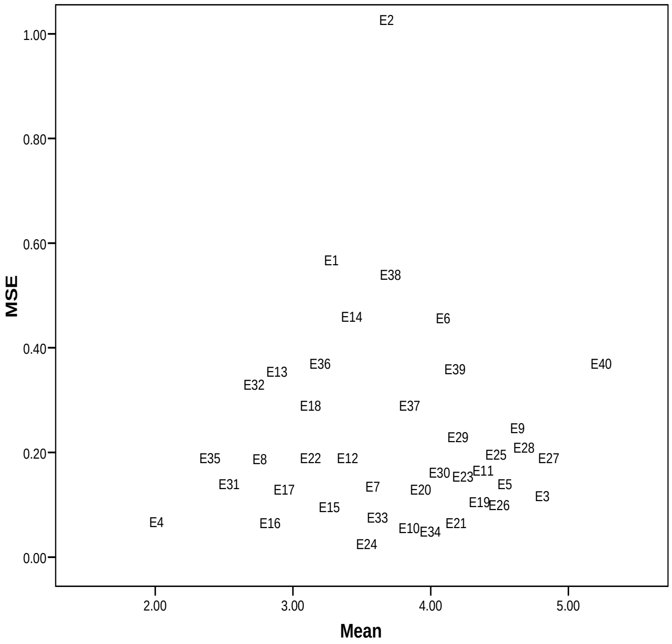

Fig. 3. Plot of environment's MSEs versus Means for 20 genotypes across 40 hetroscedastics environments.

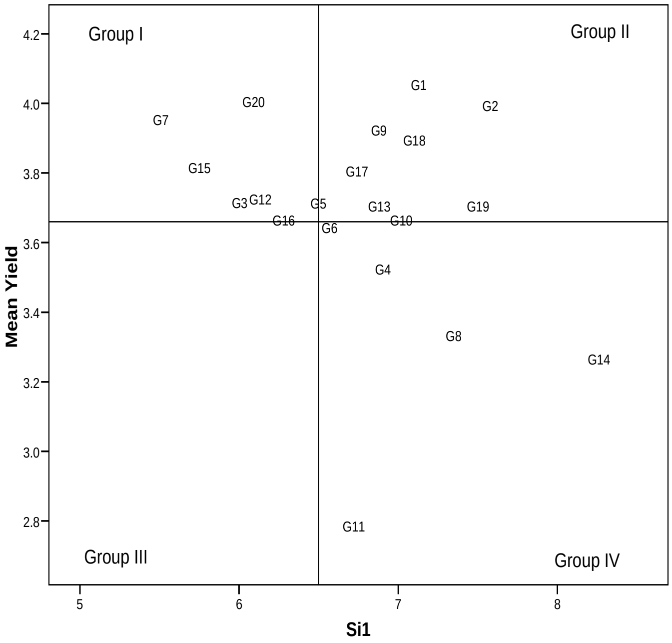

Fig. 4. Classification of 20 wheat genotypes based on the grain yield (t ha^{-1}) and $Si^{(1)(}$ across 40 hetroscedastic environments. The grand mean for grain yield was 3.66 (t ha^{-1}) and the mean of the $Si^{(1)}$ was 6.65.

For more than 50% environments, R^2 are less than 0.70. The results reported in Table 1 are indicator of violation of strong linearity assumption in the model. Cornell & Berger (1987) point out some factors in the data set that lower the value of R^2. One of these factors which is relevant to our study, is that response variable yield contains ample outliers within the range of environments means (Fig. 1). The outlier observation will have larger residual than the other observations. The residuals of outliers made both SSE and SSY large. As a result $R^2 = 1 - SSE/SSY$ approaches zero. Hence, outlier observations within the range of environment means have a reducing effect on R^2 and change strong linear association into moderate linear relationship. In particular, the linearity of response has been convincingly shown not to hold in many international multienvironment trials (Byth *et al*., 1976), including even for IMETs with carefully managed environments (Chapman *et al*., 1996).

It is a common practice among agricultural scientists that they usually ignore some strict and sensitive parametric assumptions e.g., homogeneity, normality and no outlier assumptions prior to undertaking the combined experiments. Any misunderstanding of statistics and statistical methods might lead to misleading and invalid inferences.

In many environments, problems encountered are in connection with the design of field experiments. There are a number of factors such as insect or bird attacks, crop diseases, lack of management, improper blocking and other accidental causes, especially in developing countries like Pakistan, may affect plots of some blocks of the small single RCBD experiments and may contribute to low yield of one or two genotypes (outliers) in comparison to others.

It is an invariable fact that in the METs the relative differences among genotypes across environments are inconsistent due to GEI. The differences in ranks of the 20 genotypes across 40 heteroscedastic environments indicated the existence of GEI (data not shown). The mean grain yield of 20 genotypes across 40 environments varied from 2.74 (t ha^{-1}) (G11) to 4.01 (t ha^{-1}) (G1) is shown in Table 3. The genotypes of G1, G2, G3, G5, G7, G9, G12, G16, G17, G18 and G20 had higher than grand mean grain yield while the rest had equal or lower values than grand mean grain yield. Taking the top 20% high yielding genotypes as a criterion for the assessment of the genotypes, G1, G2, G7 and G20 gave the best yields, with mean yields greater than 3.9 (t ha^{-1}) (Table 3). The environmental mean yield across the genotypes varied from 1.24 (t ha^{-1}) in E12 to 6.78 (t ha^{-1}) in E40 respectively and closely followed by 6.72 (t ha^{-1}) in E5 (Table1). The mean yield and estimates of the 5 nonparametric statistics for the evaluation of genotypes over heteroscedastic environments are presented in Table 3.

The nonparametric statistics $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ values are based on transformed yield values for each genotype. Genotypes with low $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ are considered to have the high stability (Huehn, 1996). The statistical tests of significance $S_i^{(1)}$ and $S_i^{(2)}$ were proposed by Nassar & Huehn (1987). For each ith genotype, $Z_i^{(k)}$ k=1,2 values were calculated based on the $E(S_i^{(k)})$ and $Var(S_i^{(k)})$ under the null hypothesis H_0, that the mean stability for the ith genotype is $E(S_i^{(k)})$ against the alternative that the mean stability deviates from this expectation. If H_0 is rejected, the ith genotype may be stable [i.e., $S_i^{(k)} < E(S_i^{(k)})$]or unstable [i.e., $S_i^{(k)} > E(S_i^{(k)})$] (Huehn & Nassar, 1991). As shown in Table 3 the two overall chi square tests ($\chi^2_{0.05(20)} = 35.32$, $\chi^2_{0.05(20)} = 35.94$) resulted in greater than the tabulated value $\chi^2_{0.05(20)} = 31.41$and there was sufficient evidence for significant differences among the stability values of the 20 genotypes in our study.

Table 3. Mean yield and 5 nonparametric statistics for wheat yield of 20 genotypes in 40 heteroscedastic environments in Pakistan.

Genotype	**Yield ($t\ ha^{-1}$)**	**$S_i^{(1)a}$**	**$Z_i^{(1)b}$**	**$S_i^{(2)a}$**	**$Z_i^{(2)b}$**	**$S_i^{(3)a}$**	**$S_i^{(6)a}$**	**$YS_i^{(1)c}$**	**$YS_i^{(2)c}$**
G1	4.01	7.05	0.66	36.94	0.58	145.17	21.69	17	17
G2	3.95	7.50	3.00	42.13	3.38	170.27	23.28	21	21
G3	3.75	5.93	2.13	26.56	1.91	103.07	17.13	11	11
G4	3.56	6.83	0.13	35.45	0.21	117.15	17.25	27	28
G5	3.67	6.42	0.22	31.37	0.15	108.76	16.53	15	16
G6	3.6	6.49	0.11	31.13	0.19	118.15	18.31	21	20
G7	3.91	5.43	6.16	23.39	4.16	99.17	16.39	5	5
G8	3.29	7.27	1.61	39.12	1.47	137.44	19.66	31	31
G9	3.88	6.80	0.09	35.07	0.14	130.9	19.89	17	18
G10	3.62	6.88	0.21	35.22	0.17	136	19.78	27	26
G11	2.74	6.64	0.00	32.92	0.00	121.4	18.25	25	25
G12	3.76	6.02	1.65	27.18	1.58	94.01	15.17	17	16
G13	3.66	6.80	0.09	34.91	0.12	128.13	20.05	26	25
G14	3.22	8.18	9.70	50.39	12.58	172.78	22.07	36	36
G15	3.77	5.67	3.98	24.18	3.52	85.92	14.49	15	15
G16	3.62	6.20	0.84	28.40	1.01	101.61	16.73	22	22
G17	3.76	6.66	0.00	33.32	0.00	122.6	19.47	25	25
G18	3.85	7.02	0.57	36.64	0.49	143.26	21.75	30	30
G19	3.66	7.42	2.45	41.48	2.90	160.99	22.92	34	34
G20	3.96	6.01	1.72	27.57	1.38	109.15	16.62	20	21
Sum			35.32		35.94				
	$E(S_i^{(1)})$	6.65	$E(S_i^{(2)})$	33.25					
	$Var(S_i^{(1)})$	0.24	$Var(S_i^{(2)})$	23.36					
	$\chi^2_{0.05(20)} = 31.41$		$\chi^2_{0.05(1)} = 3.84$						

a $S_i^{(1)}$ average absolute rank dispersion of a genotype over environments, $S_i^{(2)}$ the variance among the ranks over environments, $S_i^{(3)}$ and $S_i^{(6)}$ are the sum of absolute deviations and sum of squares of ranks for each genotype relative to the mean of ranks respectively. b $Z_i^{(1)}$ and $Z_i^{(2)}$ are chi-square χ^2 test statistics for $S_i^{(1)}$ and $S_i^{(2)}$ c $YS_i^{(1)}$ and $YS_i^{(2)}$ are the modified rank-sum of Kang (1988) Statistics.

Corresponding to Huehn's first two methods $S_i^{(1)}$ and $S_i^{(2)}$, G7 had the smallest changes in ranks and thus regarded as the most stable genotype unlike G14, which was significantly ($p<0.005$) unstable. The next most stable genotype was G15, followed by G3 (Table 3). The nonparametric statistics $S_i^{(1)}$ and $S_i^{(2)}$, are almost in complete linear agreement in ranking with 20 genotypes across environment as shown in Tables 4 & 5, while exact linear relationship was found in Sagherloo *et al*., (2008) and Rose IV *et al*., (2007). By using the $S_i^{(3)}$, the genotypes G15, G12 and G7 with minimum $S_i^{(3)}$ value under heteroscedastic environments were considered to be stable and the genotypes G14, G2 and G19 were unstable. According to $S_j^{(6)}$ the genotypes G15, G12 and G7 with low $S_j^{(6)}$ values were stable genotypes and the genotypes G2, G19 and G14 with high $S_j^{(6)}$ values were unstable.

Rose IV *et al*., (2007) have reported that examination of the formulae for the $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ (rank-based statistics) revealed that small number of genotypes when assayed with these statistics have been found to be extremely sensitive but may give misleading results, whereas in our study a large number of genotypes with similar statistics gave valid results.

With reference to Kang (1988) and Yue *et al.,* (1997) stability statistics $YS_i^{(1)}$ and $YS_i^{(2)}$ rank-sum of the 20 genotypes are shown in Table 4. The genotypes G7, G3 and

G15 with lower values of $YS_i^{(1)}$ and $YS_i^{(2)}$ were considered to be stable while G14, G19 and G8 with higher $YS_i^{(1)}$ and $YS_i^{(2)}$ were considered to be unstable. The most unstable genotype according to $S_i^{(1)}$, $S_i^{(2)}$, $YS_i^{(1)}$ and $YS_i^{(2)}$ was G14, which ranked 19th for mean yield (Table 4). G1 had the highest mean yield and ranked 5th and 6th according to $YS_i^{(1)}$ and $YS_i^{(2)}$. It is pertinent to note that the stable genotypes identified by all 5 rank-based stability methods in this study had grain yields above the grand average. With reference to the 5 distributions free methods applied in our study G7 and G15 were identified as the most stable genotypes, and G14 and G19 as unstable ones. The remaining genotypes were intermediate between these two strata's. As a result of this study, G7 was recommended for national release in Pakistan, as it adequately demonstrated stability across wide variety of environments and also ranked 4th in yield performance.

The rank correlation between mean yield was significantly and positively correlated with Kang's (1988) modified rank-sum statistics ($YS_i^{(1)}$ and $YS_i^{(2)}$) but it was not correlated with $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ (Table 5). The stability statistics $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ were positively and significantly correlated ($p<0.01$), indicating that the four statistics were similar in classifying the genotypes according to their stability under heteroscedastic environmental conditions (Table 5). Consequently, only one of these statistics would be sufficient to select the stable genotype in a plant breeding program. Similar results were obtained by Mohammadi *et al*., (2008) in a study on wheat and (Sabaghnia *et al*., 2006) on lentil.

Among Huehn's four rank-based stability statistics $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ the first two $S_i^{(1)}$ and $S_i^{(2)}$, were very much positively linearly associated among themselves. Overall, all $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ could be highly satisfactory measures for stability, whereas $S_i^{(1)}$ was better than others (Huehn, 1990).

As a simple but alternative graphical assessment of stability of genotypes was also assessed by mean vs. $S_i^{(1)}$ Plot. The 20 genotypes were then classified into the following four groups on the basis of grain mean yield and $S_i^{(1)}$ across environments. This idea was taken from (Francis & Kennenberg, 1978) in which they grouped genotypes on the basis of mean yield and coefficient of variation across environments.

Group 1: High mean yield (above the average of all genotypes) and low $S_i^{(1)}$ (below the expected value of $S_i^{(1)}$)
Group 2: High mean and high $S_i^{(1);}$
Group 3: Low mean and low $S_i^{(1)}$:
Group 4: Low mean and high $S_i^{(1)}$;

According to these configurations, from Figure 4 genotypes falling in group 1 can be considered as stable. Group 1 contains that G7, G15, G20, G3 and G12 are most stable, and well adapted to all environments i.e., those have general adaptability (Fig. 4). Genotypes G1, G2, G9 and G18 fall in group 2 that indicates increasing sensitivity to environmental change and greater specificity of adaptability to high yielding environments. Group 3 reflects poorly adapted genotypes to all environments. Only G16 falls in this group. Group 4, contains G6, G4, G11, G14 and G8. This group reflects greater resistance to environmental fluctuation and increasing adaptability to low yielding environments. Fig. 4 clearly indicates that G11 is an outlier genotype with reference to the lowest mean yield while G14 is also outlier corresponding to low mean yield and the highest value of average rank dispersion $S_i^{(1)}$.

Table 4. Ranks of 20 wheat genotypes based on wheat yield and 5 nonparametric stability statistics.

Genotype	Yield	$S_i^{(1)}$	$S_i^{(2)}$	$S_i^{(3)}$	$S_i^{(6)}$	Rank Sum	
						$YS_i^{(1)}$	$YS_i^{(2)}$
G1	1	16	16	17	16	5	6
G2	3	19	19	19	20	9	9
G3	10	3	3	5	7	2	2
G4	17	13	14	8	8	15	16
G5	11	7	8	6	4	3	4
G6	16	8	7	9	10	9	8
G7	4	1	1	3	3	1	1
G8	18	17	17	15	12	18	18
G9	5	11	12	13	14	5	7
G10	14	14	13	14	13	15	15
G11	20	9	9	10	9	12	12
G12	8	5	4	2	2	5	4
G13	12	12	11	12	15	14	12
G14	19	20	20	20	18	20	20
G15	7	2	2	1	1	3	3
G16	14	6	6	4	6	11	11
G17	8	10	10	11	11	12	12
G18	6	15	15	16	17	17	17
G19	12	18	18	18	19	19	19
G20	2	4	5	7	5	8	9

Table 5. Mean values Y and 5 nonparametric stability statistics for grain yield of twenty wheat genotypes evaluated in 40 environments.

	Y	$S_i^{(1)}$	$S_i^{(2)}$	$S_i^{(3)}$	$S_i^{(6)}$	$YS_i^{(1)}$	$YS_i^{(2)}$
Y	1	0.027	0.013	-0.05	-0.068	0.647**	0.615**
$S_i^{(1)}$	0.027	1	.994**	0.947**	0.907**	0.748**	0.770**
$S_i^{(2)}$	0.013	0.994**	1	0.943**	0.893**	0.730**	0.765**
$S_i^{(3)}$	-0.05	0.947**	0.943**	1	0.965**	0.661**	0.687**
$S_i^{(6)}$	-0.068	0.907**	0.893**	0.965**	1	0.630**	0.640**
$YS_i^{(1)}$	0.647**	0.748**	0.730**	0.661**	0.630**	1	0.990**
$YS_i^{(2)}$	0.615**	0.770**	0.765**	0.687**	0.640**	0.990**	1

** Correlation is significant at the 0.01 level (2-tailed).

Prior to selection, it is quite crucial to be aware of genotypes ranking in each environment and Figure 4 provided mean yield (t ha^{-1}) vs. $S_i^{(1)}$ values accordingly. G7 and G15 are the most stable and well adopted across heteroscedastic environments. G15 has the highest mean rank, while G11 the lowest one. However G7 has larger mean yield than G15.

To understand the relationships among the rank-based statistics, principal component analysis (PCA) was performed on the rank correlation matrix (Table 5). PCA is a multivariate statistical technique which can be used for simplification and dimensionality reduction in a data set by retaining those characteristics that contribute most to its variation. In this regard lower-order principal components are retained and higher order ones are ignored.

The results reported in Table 6 indicate that the loadings of the first two PCAs which explained 93% (74% and 19% by PCA1 and PCA2 respectively) of the variation of original variables. PCA1 is primarily stability and PCA2 is mostly yield. The relationships among the different nonparametric stability statistics are graphically displayed in a biplot of PCA1 versus PCA2 in Fig. 5, where both axes were considered simultaneously. Three groups in Fig. 5 can be defined as;

Group 1: $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$
Group 2: $YS_i^{(1)}$ and $YS_i^{(2)}$
Group 3: Mean yield (Y)

Group 1 that included (Huehn 1979) four stability statistics $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ are shown in Fig. 5. These nonparametric methods were positively linearly correlated with each other and with: $YS_i^{(1)}$ and $YS_i^{(2)}$ ($p<0.01$) but not with mean yield (as reported in Table 4). Huehn (1979) stability statistics $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ provide a measure of stability in the static sense, since a genotype showing a constant performance in all environments does not necessarily respond to improved growing conditions with increased yield. Therefore, stable genotypes according to these methods are adapted for those regions where growing conditions are unfavorable.

Group 2, which contains rank-sum $YS_i^{(1)}$ and $YS_i^{(2)}$ were found to be positively and significantly correlated ($p<0.01$) to each other while moderately positively correlated with mean yield**.** This group consists of statistics that were influenced simultaneously by both mean yield and stability. The statistic rank-sum is related to dynamic stability, while the remaining procedures are associated with static stability.

Yue *at el*., *(*1996) and Yue *at el*., *(*1997) have reported that the rank-sum is related to high yield performance. Therefore rank-sum stability statistics ($YS_i^{(1)}$ and $YS_i^{(2)}$) are related to dynamic concept. Becker & Leon (1988) suggested that a dynamic concept of stability does not require the genotypic response to environmental conditions to be equal for all genotypes. Group 3 contains mean yield separated by the two PCAs axes from the rank-sum methods $YS_i^{(1)}$ and $YS_i^{(2)}$ (Group2) (Fig.4). The stability methods $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$ and $S_i^{(6)}$ were positively linearly and significantly correlated ($P<0.01$), indicating that the four statistics were similar under heteroscedastic environmental conditions (Table 5). As a result, only one of these statistics would be sufficient to select stable genotypes in a breeding program. Scapim *et al*., (2000) found significantly positive correlations between $S_i^{(1)}$ and $S_i^{(2)}$ in maize. Flores *et al*., (1998) also stated high positive rank correlation between $S_i^{(1)}$ and $S_i^{(2)}$ in fababean and peas. Piepho & Lotito (1992) have reported that generally, the results for the large data sets are more constant than for the small data sets and they found strong positive linear relation between $S_i^{(1)}$ and $S_i^{(2)}$ in sugar beet ($p<0.001$).

The results of PC analysis performed for the ranks of genotypes according to 5 rank-based methods are shown in the transpose of Table 5. The biplot explained 81% of the variation of the original variables in study 53% and 28% by PCA1 and PCA2, respectively (Table 7). A biplot based on first two PCAs applied to detect locations of genotypes (Fig. 6). Genotypes that had PCA1 Scores <0 were identified as higher yielding (except G14 near- zero PCA1 score) and those that had PCA1 scores > 0 were identified as lower yielding (G12 and G15). However genotypes G7 and G3 are high yielding and stable but having moderate and identical both PCAs scores (Fig. 6). In the biplot it is noticeable that the genotype 14 (G14) is most unstable and outlier genotype across environments, and G13 with scores for axis 1 and 2 close to zero, showed the smallest interaction.

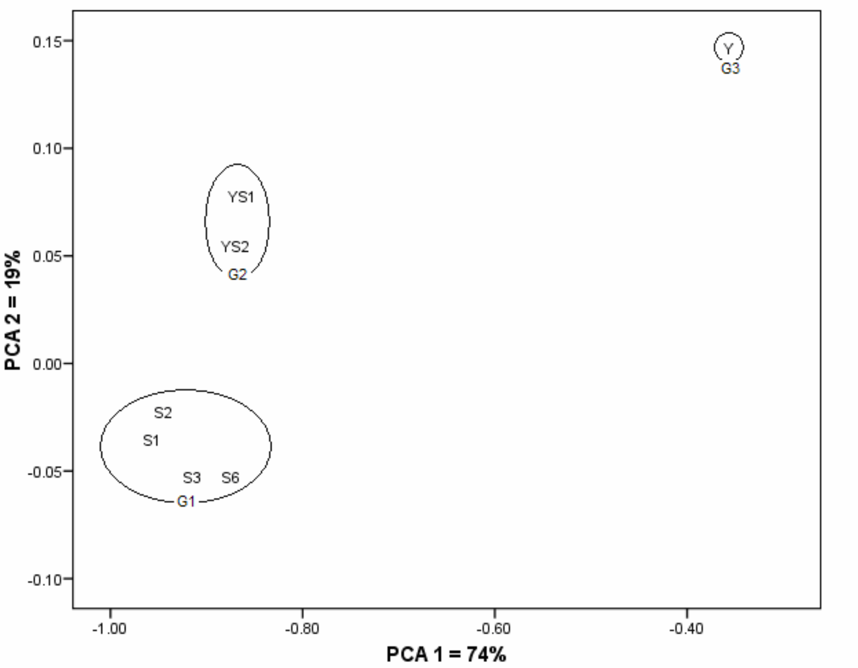

Fig. 5. Biplot of the first two principle component of ranks of stability of yield, estimated by 5 rank-based methods using yield data from 20 genotypes across 40 environments.

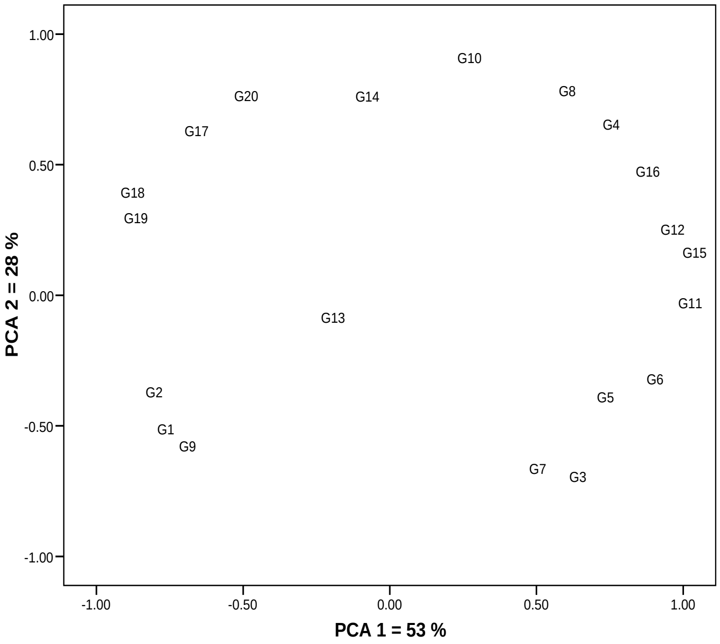

Fig. 6. Biplot of first two principal components for studied genotypes.

Table 6. Loadings of rank derived from different nonparametric stability measures for PCA 1 and PCA 2.

Stability statistics	Principal component	
	PCA1	PCA2
Yield	-0.37	0.14
Si(1)	-0.97	-0.03
Si(2)	-0.96	-0.03
Si(3)	-0.93	-0.06
Si(6)	-0.89	-0.06
YSi(1)	-0.88	0.07
YSi(2)	-0.89	0.06
Variance explained (%)	74	19

Table 7. First two principal components loadings of ranks obtained from 5 rank-based methods used to analyze G X E interaction of wheat genotype yields.

Genotypes	Principal component axis	
	PCA1	PCA2
V-01078 (G1)	-0.81	-0.46
99B4012 (G2)	-0.85	-0.43
Wafaq 2001 (G3)	0.60	-0.75
RWM-9313 (G4)	0.71	0.60
V-00125 (G5)	0.69	-0.45
DIAMOND (G6)	0.86	-0.38
PR-84 (G7)	0.46	-0.72
TW 0135 (G8)	0.56	0.73
V-00055 (G9)	-0.74	-0.64
99B2278 (G10)	0.22	0.85
KT-7 (G11)	0.97	0.02
V-01180 (G12)	0.92	0.19
DN-47 (G13)	-0.24	-0.14
V-9021 (G14)	-0.12	0.70
CT-00062 (G15)	0.99	0.10
7_03 (G16)	0.83	0.41
PR-86 (G17)	-0.71	0.57
V-02192 (G18)	-0.92	0.33
V-002493 (G19)	-0.92	0.34
L. Check (G20)	-0.54	0.70
Eigenvalue	10.58	5.62
Variance Explained (%)	53	28

In this study environmental conditions i.e., MSEs are severely different (heteroscedastic) as presented in Table 1. This is very true of Pakistan because climatic conditions differ from province to province and within the provinces as well. Moreover the quality of management also differs among the cultivars as well as among the experimental stations situated in various provinces of Pakistan. Thus, stable genotypes as detected in this study recommended for those Pakistani regions, where multienvironmental trials growing conditions are widely different.

Conclusions

The most widely used stability parametric methods may be sensitive to violation of normality, homogeneity, no mixture of normal distributions and absence of many outliers' assumptions in combined analysis of variance model. Violation of these assumptions may make point and interval estimation, and testing of hypothesis difficult. There is general agreement among researchers about superiority of parametric methods over nonparametric methods in terms of power; but if one or more of the underlying parametric assumptions are severely violated, the power advantage may not be there. For retaining parametric stability measures, various transformations (logarithms and roots) can be applied to rectify the violation of these assumptions, but none of these could be successful. As a last resort under such situations, most of the researchers prefer the use of appropriate nonparametric stability methods. In the present study 5 most widely applied nonparametric stability measures were applied to 20 genotypes grown across 40 hetroscedastic environments. The focus of the present study was thoroughly testing parametric combined ANOVA model assumptions, rectification and justification of nonparametric stability methods and their application to the data from NUWYT experiments in Pakistan for the year (2004-05). The results of the study revealed severe violation of assumptions that justify the use of nonparametric stability methods to analyze GEI in METs. The results of the study using graphical, tabular and statistical tests explicitly indicate that these METs data are not suitable for parametric stability methods and therefore recommendations based on those methods may be seriously misleading. However, recommendations based on the above suggested non-parametric methods combined with multivariate methods can be used to formulate more plausible recommendations which are based on the evidence from the data and less relying on the assumptions.

Acknowledgements

We are grateful to Wheat Program, National Agricultural Research Centre, Islamabad, Pakistan for providing data, the Higher Education Commission, Pakistan for financial support and the Department of Statistics, University of Nebraska, Lincoln, USA for providing facilities.

References

Becker, H.C. and J. Leon. 1988. Stability analysis in plant breeding. *Plant Breed.,* 101: 1-23.

Bowman, D.T. and C.E. Watson.1997. Measures of validity in cultivar performance trials. *Agron. J.,* 89: 860-866.

Box, G.E.P. and D.R. Cox. 1964. An analysis of transformations*. J. R. Statist. Soc. B.,* 26: 211-246.

Bradley, J.V. 1982. The insidious L-Shaped distance. *Bull. Psychonomic Soc.,* 20: 85-88.

Byth, D.E., R.L. Eisemann and I.H. DeLacy. 1976. Two-way pattern analysis of a large data set to evaluate genotype adaptation. *Heredity,* 37: 215-230.

Chapman, S.C., G.O. Edmeades and J. Crossa. 1996. Pattern analysis of gains from selection for drought tolerance in tropical maize populations. In: Plant adaptation and crop improvement. (Eds.): M. Cooper, G.L. Hammer. CAB International, UK, 513-527.

Cochran, W.G. 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics,* 4: 22-38.

Cornell, J.A. and R.D. Berger. 1987. Factors that influence the value of the coefficient of determination in simple linear and nonlinear regression models. *Phytopathology,* 77: 63-70.

de Kroon, J. and P. van der Laan. 1981. Distribution-free test procedures in two-way layouts: a concept of rank-interaction. *Stat Neeri,* 35:189-213.

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

Eisenhart, C. 1947. The assumptions underlying the analysis of variance. *Biometrics,* 3: 1-21.

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

Flores, F., M.T. Moreno and J.I. Cubero. 1998. A comparison of univariate and multivariate method to analyze environments. *Field Crops Res.,* 56: 271-286

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

Hollander, M. and D.A. Wolfe. 1999. *Nonparametric Statistical Methods.* 2nd. ed J. Wiley, New York.

Huehn, M and R. Nassar. 1989. On tests of significance for nonparametric measures of phenotypic stability. *Biometrics,* 45: 997-1000.

Huehn, M. 1990. Nonparametric measures of phenotypic stability: part 1. Theory, *Euphytica,* 47: 189-194.

Huehn, M. 1996. Nonparametric analysis of genotype x environment interactions by ranks. In: *Genotype by environment interaction.* (Eds.): M.S. kang and H.G. Gauch. CRC Press, Boca Raton, FL, USA, pp. 213-228.

Huehn, M. and R. Nassar. 1991. Phenotypic stability of genotypes over environment: On tests of significance for two nonparametric measures. *Biometrics*, 47: 1196-1197.

Huehn, V.M. 1979. Beitrage zur erfassung der phanotypischen stabilitat. *EDV Med. Biol.,* 10: 112-117.

Hussein, M.A., A. Bjornstad and A.H. Aastveit. 2000. SASG X ESTAB: A SAS program for computing genotype X environment stability statistics. *Agron. J.,* 92: 454-459.

Javed, H.I., M.A. Masood, S.R. Chughtai, H.N. Malik, M. Hussain and A. Saleem. 2006. Performance of maize genotypes on the basis of stability Analysis in Pakistan. *Asian J. Plant Sci.,* 5: 207-210.

Kang, M.S. 1988. A rank-sum method for selecting high yielding stable corn genotypes. *Cereal Res. Comm.,* 16: 113-115.

Kang, M.S. 1998. Using genotype-by environment interaction for crop cultivar development. *Adv. Agron.,* 62: 199-253.

Khan, I.A., B.A. Malik and M. Bashir. 1988. Investigation of genotype x environment interaction for seed yield in chickpea (*Cicer arietinum* L.). *Pak. J. Bot.,* 20: 201-204.

Kuehl, R.O. 2000. *Design of Experiments Statistical Principles of Research Design and Analysis.* 2nd. (Ed.). Duxbury.

Lin, C.S. and M.R. Binns. 1994. Concepts & Methods for Analyzing Regional Trial Data for Cultivar & Location selection. In: *Jules J. Plant Breed. Review,* 12: 271-297.

Lin, C.S. and M.R. Binns.1988. A superiority measure of cultivar performance for cultivar x location data. *Can. J. Plant Sci.,* 68: 193-198.

Lin, C.S., M.R. Binns and L.P. Lefkovitch. 1986. Stability analysis: where do we stand? *Crop Sci.,* 26: 894-900.

Mohammadi, R. and A. Amri. 2008. Comparison of parametric and nonparametric methods for selecting stable and adapted durum wheat genotypes in variable environments. *Euphytica,* 159: 419-432.

Nassar, R. and M. Huehn. 1987. Studies on estimation of phenotypic stability: tests of significance for nonparametric measures of phenotypic stability. *Biometrics*, 43: 45-53.

Piepho, H.P. and S. Lotito. 1992. Rank correlation among parametric and nonparametric measures of phenotypic stability. *Euphytica,* 64: 221-225.

Rogan, J.C. and H.J. Keselman. 1977. Is the ANOVA F- Test to Variance Heterogeneity when sample sizes are equal?: An Investigation *Via* Coefficient of Variation. *Amr. Edu. Res*. J., 14: 493-498.

Rose, IV LW, M.K. Das and C.M. Taliaferro. 2007. A comparison of dry matter yield stability assessment methods for small numbers of genotypes of bermudagrass. *Euphytica,*(published on line Nov 28 2007 DOI 10.1007/s10681-007-9620-2).

Rousseeuw, P.J. and A.M. Leroy. 1987. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons, Inc.

Sabaghnia, N., H. Dehghani and S.H. Sabaghpour. 2006. Nonparametric methods for interpreting genotype x environment interaction of lentil genotypes. *Crop Sci.,* 46: 1100-1106.

Scapim, C.A., V.R. Oliveira, A.L. Braccinil, C.D. Cruz, C.A.B. Andrade and M.C.G. Vidigal. 2000. Yield stability in maize (*Zea mays* L.) and correlations among the parameters of the Eberhart and Russell, Lin and Binns and Huehn models. *Genet. Mol. Biol.,* 23: 387-393.

Segherloo, A.E., S.H. Sabaghpour, H. Dehghani and M. Kamrani. 2008. Nonparametric measures of phenotypic stability in chickpea genotypes (*Cicer arietinum* L.). *Euphytica,* 162: 221-229.

Shukla, G.K. 1972. Some aspects of partitioning genotype-environmental components of variability. *Heredity*, 28: 237-245.

Sial, M.A., M.A. Arain and M. Ahmad 2000. Genotype x environment interaction on bread wheat grown over multiple sites and years in Pakistan. *Pak. J. Bot.*, 32: 85-91.

Snedecor, G.W. and W.G. Cochran. 1980. *Statistical Methods.* 7th (ed.) Iowa State University Press, USA.

Truberg, B. and M. Heuhn. 2000. Contribution to the analysis of genotype by environment interactions: comparison of different parametric and nonparametric tests for interactions with emphasis on crossover interactions. *Agron. Crop Sci.,* 185: 267-274.

Warren, J.A. and I. Mendez. 1982. Methods for estimating background variation in field experiments. *Agron. J.,* 74: 1004-1009.

Yates, F. and W.G. Cochran. 1938. The analysis of groups of experiments. *J. Agric. Sci.,* 28: 556-580.

Yue, G., W.F. Heer, J.L. Moyer, L.D. Maddux, E.L. Sorensen and G.H. Liang. 1996. Comparison of nonparametric and parametric stability estimates to evaluate alfalfa cultivars. *J. Genet & Breed.,* 50: 67-74.

Yue, G.L., K.L. Roozeboom, W.T. Schapaugh and G.H. Liang. 1997. Evaluation of soybean cultivars using parametric and nonparametric stability estimates. *Plant Breed.,* 116: 271-275.

Zar, J.H. 1996. *Biostatistical Analysis.* 3rd. (ed.). Prentice Hall, Englewood Cliffs, New Jersey.

(Received for publication 24 July 2008)