Non-parametric approach to the study of phenotypic stability

D.F. Ferreira[1], S.B. Fernandes[2], A.T. Bruzi[3] and M.A.P. Ramalho[2]

[1]Departamento de Ciências Exatas, Universidade Federal de Lavras, Lavras, MG, Brasil
[2]Departamento de Biologia, Universidade Federal de Lavras, Lavras, MG, Brasil
[3]Departamento de Agricultura, Universidade Federal de Lavras, Lavras, MG, Brasil

Corresponding author: S.B. Fernandes
E-mail: samuelfernandes@agronomo.eng.br

ABSTRACT. The aim of this study was to undertake the theoretical derivations of non-parametric methods, which use linear regressions based on rank order, for stability analyses. These methods were extension different parametric methods used for stability analyses and the result was compared with a standard non-parametric method. Intensive computational methods (e.g., bootstrap and permutation) were applied, and data from the plant-breeding program of the Biology Department of UFLA (Minas Gerais, Brazil) were used to illustrate and compare the tests. The non-parametric stability methods were effective for the evaluation of phenotypic stability. In the presence of variance heterogeneity, the non-parametric methods exhibited greater power of discrimination when determining the phenotypic stability of genotypes.

Key words: Bootstrap; Permutation; Genotype x environment interaction

INTRODUCTION

The evaluation of cultivars in diverse environments is a practice that is routinely adopted by breeders when evaluating potential cultivars, and this practice is attributed to the occurrence of genotype x environment interactions. The risks associated with the recommendation and adoption of cultivars are reduced with the evaluation and identification of more stable and adapted cultivars, which have greater homeostatic capacity in the face of unpredictable environmental variations (Lin and Binns, 1988). Although there is no unanimous concept for phenotypic stability, various methods and models have been used to study and explain genotype x environment interactions. Among the methods used to determine stability, those that use single-segmented and bi-segmented linear regression stand out. In addition, other methods are being used, which may be considered complementary to the existing regression methods (e.g., Lin and Binns, 1991).

One widely used method proposes the calculation of a recommendation risk index for each genotype as a parameter used to evaluate phenotypic stability (Annicchiarico, 1992; Annicchiarico et al., 1995). Multivariate methods also stand out in studies of phenotypic stability, including the widely used additive multiplicative model interaction method (Gauch and Zobel, 1988; Crossa, 1990; Yau, 1995). Other univariate techniques used for the evaluation of stability include the ecovalence method (Wricke, 1964) and the Plaisted and Paterson (1959) method.

Non-parametric techniques are widely used in experimentation as a whole, especially when the basic presuppositions of normality and heteroscedasticity are not verified. There are constant problems associated with experiments conducted during the final phase of breeding, especially heteroscedasticity problems in experiments conducted at different locations and during different years (Piepho, 1997). Thus, appropriate non-parametric techniques could be used to study phenotypic stability in these situations.

In light of this, the aim of this study was to obtain theoretical derivations of non-parametric methods, which use linear regressions based on rank order, for stability analyses. Moreover, these methods were extensions of the methods of Plaisted and Paterson (1959), Wricke (1964), Lin and Binns (1988, 1991), Annicchiarico (1992), and Annicchiarico et al. (1995), and the methods can also be compared to those of Huhn (1990).

MATERIAL AND METHODS

The mean response of the i-th genotype in the j-th environment was initially represented by Y_{ij}, where $i = 1, 2, ..., p$ and $j = 1, 2, ..., q$. R_{ij} was defined as the rank of the i-th genotype in the j-th environment based on the increasing order of the Y_{ij} values, and the rank of the lowest and highest values (Y_{ij}) were equal to 1 and pq, respectively. When the same values were observed, the arithmetic mean of the ranks that would be attributed if there were no tie was taken. The resulting mean rank was used to represent the tied observations. In the same way, R_{ij} values are obtained if Y_{ij} was replaced by the effect of the interaction, and they constituted a second alternative for obtaining ranks. A third alternative in which the genotypes were ordered in each environment was used, and the values of 1 and p were attributed to the lowest and highest phenotypic values, respectively. Tied cases were treated in a way similar to the methods described above.

The first method adopted was simple linear regression based on rank order, and the statistical model was as follows:

$$R_{ij} = \beta_{0i} + \beta_{1i}X_j + \delta_{ij} \qquad \text{(Equation 1)}$$

where $X_j = \cdot_j - \cdot_\cdot$ is the environmental index, i = 1, 2, ..., p and j = 1, 2, ..., q. In this case, the environmental index was the deviation from the overall mean value of the ranks for each reported environment. The least square method for linear models was used to estimate the parameters and to obtain the sums of squares and products.

The main hypothesis tested was H_O: b_{li} = 1, and alternatives using intensive computation were evaluated. Under the null hypothesis, if the linear coefficient was equal to 1, it was possible to generate a null distribution. Moreover, the ranks of the genotypes in the environments were permuted, and the estimates of each permutation were obtained. A pivotal statistic, similar to the Student *t*-statistic, was idealized using:

$$t = \frac{\hat{\beta}_{li} - 1}{S(\hat{\beta}_{li})} \qquad \text{(Equation 2)}$$

where S ($\hat{\beta}_{li}$) was the square root of the element corresponding to b_{li} in the matrix $(X'X)^{-1}$ of the least square method multiplied by the regression deviation variance of ANOVA obtained from the ranks; and $\hat{\beta}_{li}$ was the estimator of b_{li}. In each permutation undertaken, respective estimates were substituted in this expression, and the observed value was recorded. The process was repeated thousands of times until the null distribution was obtained. If the value obtained in the original sample belonged to the two tailed α of this distribution, the null hypothesis was rejected at this nominal significance level (100α%). As the total number of permutations was very high, only a portion (sample) of the permutations was performed. In general, a total of 10,000 permutations were used.

A viable alternative used to test this hypothesis is the resampling of the ranks with replacement, which is known as bootstrapping. The two tests were conducted, and the basic difference between them was characterized by the use or lack of use of replacements when sampling the ranks. It should be emphasized that when attributing ranks to genotypes within a single environment, the third alternative is not adequate for the proposed studies. This is because the environmental effect is artificially eliminated, since the mean values of environmental ranks are equal.

The non-parametric generalization of the methodology of Lin and Binns (1991) was then derived such that each genotype was compared to the maximum performance in each environment-based rank. The overall stability index of each genotype was broken down (as previously described) into genetic and interaction components. Thus, the following equation was obtained:

$$Pi = \frac{\sum_{j-1}^{q}\left(R_{\ddot{g}} - M_j\right)^2}{2q} \qquad \text{(Equation 3)}$$

where P_i represents the stability index of genotype i; R_{ij} represents the rank of the i-th genotype in the j-th environment; and M_j is the value of the maximum responses among all of the genotypes in environment j.

This expression was further broken down into:

$$P_i = \left[q(\bar{R}_{i\cdot} - \bar{M})^2 + \sum_{j-1}^{q}\left(R_{ij} - \bar{R}_{i\cdot} - M_j + \bar{M}\right)^2\right]/2q \qquad \text{(Equation 4)}$$

where $\bar{R}_{i\cdot} = \sum_{j-i}^{q} R_{\ddot{g}}/q$ is the mean rank value of genotype i; and $\bar{M} = \sum_{j-i}^{q} M_j/q$ is the mean value of maximum response ranks. This breakdown of P_i predicted its division in estimates of the genetic effect (first component) and the effect of the contribution of genotype i to the interaction (second

component of the expression). Therefore, the cultivars that contributed most to the interaction were inferred. The unbalanced state in this case may be easily resolved, without any loss or additional difficulty, by simply considering q_j in the place of q in the previous expressions. The F test carried out for P_i was replaced by a bootstrap alternative and randomization (permutation).

The non-parametric generalization of the risk index of each genotype was proposed by Annicchiarico (1992) and Annicchiarico et al. (1995). Initially, each genotype had its value standardized in each environment using its transformation into ranks. The mean value of the ranks ($\overline{R}_{\bullet j}$) in the j-th environment was calculated, and the values of the i-th genotype in the j-th environment were taken as a percentage of the environmental mean value using the following equation:

$$V_{ij} = 100R_{ij} / \overline{R}_{\cdot j} \qquad \text{(Equation 5)}$$

Once these standardized values were obtained, the mean value ($\overline{V}_{i\cdot}$) and variance of the ranks ($s_{V_i}^2$) of the i-th genotype over the q environments were estimated. From each genotype, the non-parametric risk index (I_i) was obtained in the following manner:

$$I_i = \overline{V}_{i\cdot} - Z_{(i-\alpha)} S_{Vi} \qquad \text{(Equation 6)}$$

where is the quantile 1-α of the standard normal distribution; and is the standard deviation of the i-th genotype in relation to the ranks taken as a percentage of the mean value of the environments (over q environments).

The Wricke (1964) method was also generalized for non-parametric cases. Having obtained one of the transformations into ranks, the non-parametric ecovalence was estimated as follows:

$$W_i = \sum_{j=1}^{q} \left(R_{ij} - R_{i\cdot} - R_{\cdot j} + R_{\cdot\cdot}\right)^2 \qquad \text{(Equation 7)}$$

The stability parameter estimates derived using the methodologies proposed by Nassar and Huhn (1987), Huhn (1990), Huhn and Nassar (1989, 1991), and Nassar et al. (1994) were also obtained for comparison purposes.

Initially, the null hypothesis assumed that all of the genotypes showed similar stability, and it was tested using two statistics (one for each parameter) based on the following equation:

$$K_l = \sum_{i=1}^{P} Z_{li} = \sum_{i=1}^{P} \left[S_{li} - E(S_{lj})\right]^2 / V(S_{lj}) \qquad \text{(Equation 8)}$$

In this equation, l = 1, 2, which asymptotically follows a chi-square distribution with d.f. ($\nu = p$) under the null hypothesis. Each parcel (Z_{li}) of the sum was itself used if the null hypothesis was rejected by the overall test during the identification of genotype or genotypes that contributed in a differentiated manner to the interaction (i.e., those that were considered unstable). Each (Z_{li}) had chi-square asymptotic distribution with $\nu = 1$ d.f.

In the last stage, an illustrative example of the analysis results was presented using all

of the previously described techniques, which were applied using data from the common bean-breeding program of the Department of Biology of the Universidade Federal de Lavras (UFLA), Minas Gerais, Brazil. Ten genotypes were evaluated in 20 environments, and incomplete block trials with three replicates were used (Bruzi et al., 2007). All of the methods used to study phenotypic stability were implemented in the R program.

RESULTS

The data were initially subjected to joint ANOVA, and the d.f. were fitted using the Cochran (1954) procedure, because heterogeneity of the individual variances was found. The sources associated with genotype variation, environment type, and genotype x environment interaction exhibited significant results ($P < 0.05$).

The results of the simple linear regression model analyses for all 10 genotypes studied are shown in Table 1 (including eight lines, the F_2 generation of the multiple hybrids, and the mixture of lines), and the original data were used for these analyses. Highly discrepant results were found between permutation (randomization) and bootstrap analysis and the original t-test. Using the t-test, the Iapar 81 and Talismã genotypes presented regression coefficients that were significantly ($P < 0.01$) higher and lower than 1, respectively. Nevertheless, in the permutation (randomization) and bootstrap tests, no P value was less than 5%. The Scott and Knott (1974) test was applied to the mean values of the genotypes, and it verified that the best-performing cultivar was OP-S-16 (Table 1). Using the same test, the multiple hybrid (MH) and the mixture of line groups were allocated to the second yield group (group b).

Table 1. Simple linear regression analyses with the yield data.

Genotypes	β_0	β_1	R^2	t_c ($\beta_1 = 1$)	P value (permutation)	P value (bootstrap)	P value (t)
Pérola	1756.3[ba]	1.0588	95.71	1.1138	0.2794	0.2779	0.2556
Iapar 81	1672.6[d]	1.1409	91.19	1.6873	0.1050	0.1063	0.0067
Esal 693	1603.8[d]	0.9196	95.00	-1.6165	0.1231	0.1223	0.1211
MA-I-25	1826.3[b]	1.0260	91.67	0.3579	0.7306	0.7179	0.6138
Talismã	1704.6[c]	0.8622	84.85	-1.6042	0.1225	0.1291	0.0081
Carioca-MG	1703.2[c]	1.0743	93.07	1.0761	0.2925	0.2968	0.1511
OP-S-16	1947.1[a]	0.9293	89.71	-0.9524	0.3537	0.3541	0.1726
Magnífico	1557.4[d]	1.0881	89.11	0.9830	0.3359	0.3409	0.0891
MH	1839.5[b]	1.0002	94.35	0.0047	0.9958	0.9950	0.9957
MP	1758.6[b]	0.9000	90.54	-1.4591	0.1624	0.1552	0.0539

[a]Mean values of genotypes that have the same letter belong to the same group as defined by the Scott and Knott (1974) test with a nominal significance level of 5%. Tests were performed via bootstrap, permutation, and the conventional t-test with 10,000 Monte Carlo simulations.

Estimates of the stability parameters for the situation in which the set of pq values were transformed into ranks (1 to pq) are shown in Table 2. Moreover, the estimated regression coefficients, the coefficient of determination, and the P values associated with hypothesis testing of angular coefficients equal to 1 for all of the p genotypes are also shown. When the estimates of the angular coefficients obtained from the original data and the data transformed into ranks were

compared, the results suggest that only four genotypes maintained their classification relative to the ordering of the coefficients (Pérola, MA-I-25, Carioca-MG, and MP). An interesting result regarding the MH sample was found, in which the sample had an estimate greater than 1 (1.0002) in the original data, but the results of the rank analysis indicated an estimate of less than 1 (0.94). However, in both cases, the null hypothesis that included a parameter equal to 1 was not rejected (P > 0.05).

Table 2. Analysis of simple linear regression with data transformed into ranks.

Genotypes	β_0	β_1	R^2	t_c ($\beta_1 = 1$)	P value (permutation)	P value (bootstrap)
Pérola	100.30	1.0443	92.01	0.6118	0.5249	0.5264
Iapar 81	92.80	1.1465	89.44	1.5788	0.1116	0.1130
Esal 693	91.95	0.9983	91.73	-0.0238	0.9804	0.9807
MA-I-25	106.10	1.0291	85.73	0.2941	0.7565	0.7641
Talismã	100.15	0.9523	81.79	-0.4501	0.6351	0.6374
Carioca-MG	96.55	1.0915	89.67	1.0491	0.2747	0.2899
OP-S-16	117.20	0.7488	88.36	-3.9216	0.0006	0.0004
Magnífico	88.30	1.1673	89.82	1.8065	0.0694	0.0686
MH	108.00	0.9400	91.62	-0.8947	0.3513	0.3551
MP	103.65	0.8815	85.92	-1.4080	0.1526	0.1536

Tests were conducted via bootstrap and permutation with 10,000 Monte Carlo simulations.

The OP-S-16 line exhibited the greatest mean value of the ranks (117.20%) in relation to the regression coefficient. This value was significantly less than one (P < 0.01), and it exhibited a value of 0.7488, which may be considered different from zero. The other cultivars, including the Iapar 81 and Talismã lines, exhibited regression coefficients equal to 1 (P > 0.05). The Magnífico cultivar, with the lowest mean value of the ranks (88.30%), exhibited the greatest regression coefficient (1.17), but the observed P value was greater than 0.05 (0.0686).

The estimates of the stability coefficients of Lin and Binns (1991) are shown in Table 3. These estimates were obtained from the original data, but the tests were performed using the distribution obtained via permutation or bootstrap. Under these conditions, the OP-S-16 line again exhibited the best performance. Among the cultivars that most contributed to the interaction, Magnífico stands out, and it differed significantly from the maximum (P < 0.01).

Considering the stability parameters of Lin and Binns (1991), using data transformed into ranks (Table 4), the lowest contributions to the interaction continued to be the OP-S-16 and Esal 693 lines. However, in those cases, the lines of greatest heterogeneity or heterozygosity stood out amongst those that contributed least to the interaction. Similarly, lower P_i's estimates obtained using the original data were observed for the OP-S-16 line and the MH population. Although there was agreement among the estimates obtained from the original data, there were some inversions in the overall classification. For instance, Magnífico, Iapar 81, and Carioca-MG were among the populations that most contributed to the interaction. In spite of exhibiting a high value (10.56%), the Talismã cultivar may be classified in an intermediate group of lines that contributed to the interaction. In relation to P_i, the same three lines (Magnífico, Iapar 81, and Carioca-MG) exhibited the greatest estimates.

Table 3. Estimates of the stability coefficients using the Lin and Binns (1991) method on the original data (yield).

Genotypes	P_i	Genetic effect	Deviation of the interaction	Contribution interaction (%)	F_i	P value (permutation)	P value (bootstrap)
Pérola	123,929	84,252	39,677	8.02	1.56	0.9257	0.9239
Iapar 81	183,946	122,105	61,840	12.51	2.32	0.2855	0.2878
Esal 693	182,946	158,461	24,485	4.95	2.30	0.3033	0.3078
MA-I-25	96,102	57,942	38,160	7.72	1.21	0.9942	0.9936
Talismã	170,654	106,772	63,882	12.92	2.15	0.4504	0.4428
Carioca-MG	168,678	107,427	61,251	12.39	2.12	0.4689	0.4714
OP-S-16	49,687	24,132	25,556	5.17	0.63	1.0000	1.0000
Magnífico	274,784	185,645	89,139	18.03	3.46	0.0012	0.0006
MH	94,952	53,539	41,413	8.38	1.20	0.9943	0.9943
MP	132,358	83,292	49,065	9.92	1.67	0.8733	0.8698

Tests were performed via bootstrap and permutation with 10,000 Monte Carlo simulations.

Table 4. Estimates of the stability coefficients using the Lin and Binns (1988, 1991) method on data transformed into ranks.

Genotypes	P_i	Genetic effect	Deviation of the interaction	Contribution interaction (%)	F_i	P value (permutation)	P value (bootstrap)
Pérola	759.2	479.0	280.3	8.94	1.58	0.3935	0.3940
Iapar 81	1198.5	739.2	459.3	14.64	2.50	0.0031	0.0036
Esal 693	952.6	772.2	180.4	5.75	1.99	0.0850	0.0824
MA-I-25	633.9	316.3	317.6	10.13	1.32	0.6897	0.6919
Talismã	814.8	483.6	331.2	10.56	1.70	0.2727	0.2626
Carioca-MG	1034.9	602.0	432.9	13.80	2.16	0.0324	0.0326
OP-S-16	195.5	98.7	96.8	3.09	0.41	0.9999	1.0000
Magnífico	1492.2	922.4	569.9	18.17	3.11	0.0000	0.0001
MH	502.5	270.3	232.2	7.40	1.05	0.9234	0.9235
MP	616.9	380.9	236.0	7.52	1.29	0.7234	0.7310

Tests were performed via bootstrap and permutation with 10,000 Monte Carlo simulations.

The non-parametric risk index (based on ranks) and the parametric risk index are shown in Table 5, and the mean values of genotypic variance of the ranks are expressed as a percentage of the environmental mean. Large changes were observed regarding the classification of genotypic risks when the original index (IO_i) and the index based on ranks (IP_i) were compared.

Table 6 shows the ecovalence values obtained from the original data, including data transformed into ranks (from 1 to pq), transformed into ranks per environment, and modified using the Huhn (1990) method before being transformed into ranks per environment. The three methods were highly divergent, and the results indicated the stable and less stable genotypes. Transformations into rankings using the Huhn (1990) method showed greater agreement with the original ecovalence estimates (data not shown).

Table 5. Estimates of the non-parametric risk index (IP$_i$), original index (IO$_i$), mean, and standard deviation of genotypes based on data transformed initially into ranks and subsequently into percentages of the environmental mean value.

Genotypes	$\overline{R}_{i\bullet}$ (%)	S^2_i	IP$_i$ (non-parametric)	IO$_i$ (original)
Pérola	102.72	43.8571	73.14	90.15
Iapar 81	88.18	59.6568	47.94	76.35
Esal 693	68.18	40.8159	40.65	82.98
MA-I-25	115.45	49.8016	81.86	90.89
Talismã	92.72	59.2544	52.76	84.88
Carioca-MG	94.54	50.2624	60.64	81.86
OP-S-16	135.45	48.8136	102.53	103.70
Magnífico	70.90	43.7080	41.42	66.36
MH	128.18	44.3306	98.28	97.96
MP	103.63	46.4859	72.28	91.12

Normal asymptotic approximation and $\alpha = 0.25$ were used.

Table 6. Estimates of non-parametric ecovalence based on data transformed into ranks, which were based on the values of all genotypes in all environments, the values of the genotypes in each environment and the values of the genotypes per environment using the Huhn (1990) transformation.

Genotypes	W_{1i} (ranks from 1 to pq)	W_{2i} (original ranks/ environments)	W_{3i} (transformed data ranks/ environments)
Pérola	5522.26	110.55	129.75
Iapar 81	10,086.06	204.55	213.00
Esal 693	5127.41	95.75	96.95
MA-I-25	10,116.46	142.55	158.95
Talismã	11,660.41	201.80	223.20
Carioca-MG	8311.81	145.20	170.20
OP-S-16	7823.06	136.95	177.00
Magnífico	10,419.26	109.80	198.95
MH	4820.86	112.95	120.55
MP	8073.41	124.20	144.55

The parameters associated with the Huhn (1990) method were used to analyze the 10 genotypes that were evaluated over 20 environments (Table 7). The results indicated that the P values of Z_{1i} and Z_{2i} were significant for the genotypes Talismã at 0.0592 and 0.0434, for the Esal 693 at 0.0535 and 0.0691, and for Iapar 81 at 0.0871 and 0.0873, respectively.

In Table 8, Spearman correlations are shown between the main estimates of parameters that were used to evaluate phenotypic stability. The regression coefficient estimator based on the original data only significantly correlated (P < 0.01) with the regression of ranks estimator. In contrast, the regression coefficient estimator based on ranks correlated significantly (P < 0.05) with the original P$_i$ values as well as those based on ranks and those with the risk indices, but it did not correlate with the ecovalences and Huhn (1990) parameters. The P$_i$ values correlated significantly with most of the other parameters, and their partitioning relative to the interaction contribution exhibited significantly higher correlations with the ecovalences.

Table 7. Estimates of stability parameters based on the Huhn (1990) method for each genotype, including S_{1i}, S_{2i}, Z_{1i} and corresponding P values, and Z_{2i} and corresponding P values for the hypothesis tests associated with the equality of the genotype stability effect.

Genotypes	S_{1i}	S_{2i}	Z_{1i}	P value (Z_{1i})	Z_{2i}	P value (Z_{2i})
Pérola	3.0157	6.8289	0.6234	0.4297	0.6735	0.4118
Iapar 81	3.9157	11.2105	2.9266	0.0871	2.9233	0.0873
Esal 693	2.6052	5.1026	3.7251	0.0535	3.3039	0.0691
MA-I-25	3.3842	8.3657	0.0547	0.8150	0.0044	0.9466
Talismã	3.9789	11.7473	3.5577	0.0592	4.0796	0.0434
Carioca-MG	3.4947	8.9578	0.2926	0.5885	0.1671	0.6826
OP-S-16	3.5578	9.3157	0.5133	0.4737	0.3788	0.5382
Magnífico	3.7315	10.4710	1.4375	0.2305	1.6453	0.1995
MH	2.9105	6.3447	1.1707	0.2792	1.2107	0.2711
MP	3.2157	7.6078	0.0547	0.8150	0.1375	0.7107
Mean/total	3.3000	8.2500	14.3566	0.1573	14.5244	0.1504
Variance	0.1296	2.9982	-	-	-	-

Table 8. Spearman correlations between estimates of the different methods used to evaluate phenotypic stability.

	β_{id}	β_{ip}	P_{id}	CI_d	P_{ip}	CI_p	IP_i	IO_i	W_{1i}	W_{2i}	W_{3i}	Z_{1i}	Z_{2i}
β_{id}	1.00	0.83	0.32	0.32	0.58	0.64	-0.16	-0.58	0.15	0.10	0.21	-0.04	-0.09
β_{ip}		1.00	0.70	0.50	0.89	0.84	-0.59	-0.89	0.42	0.04	0.28	0.22	0.19
P_{id}			1.00	0.60	0.93	0.68	-0.96	-0.93	0.48	-0.02	0.37	0.59	0.59
CI_d				1.00	0.59	0.84	-0.41	-0.59	0.68	0.37	0.70	0.18	0.30
P_{ip}					1.00	0.81	-0.87	-1.00	0.49	0.05	0.38	0.47	0.45
CI_p						1.00	-0.50	-0.81	0.76	0.38	0.64	0.10	0.15
IP_i							1.00	-0.87	-0.33	0.14	-0.19	-0.65	-0.62
IO_i								1.00	-0.49	-0.05	-0.38	-0.47	-0.45
W_{1i}									1.00	0.53	0.84	0.01	0.09
W_{2i}										1.00	0.68	-0.15	-0.07
W_{3i}											1.00	0.15	0.25
Z_{1i}												1.00	0.98
Z_{2i}													1.00

Modulus correlations that are higher than 0.63 and 0.76 are significantly different from zero at 5 and 1%, respectively.

DISCUSSION

The presence of significant effects of genotypes and environments were confirmed by the F-test. The significance of the genotype x environment interactions showed that there was differential behavior of genotypes relative to the environments. Thus, the study of phenotypic stability for these populations was justified. The occurrence of variance heterogeneity is an important aspect that should be highlighted. As widely reported in specialized statistical literature,

heterogeneity affects the performance of tests, so inference may be compromised. Therefore, the application of non-parametric methods to these data set was justified.

Using regression analyses with the original data, the Iapar 81 and Talismã genotypes were considered responsive and non-responsive, respectively (Table 1). Nevertheless, this inference may potentially be attributed to type I error, because there was heterogeneity of the residual variances among the 20 environments. It is important to mention that the fits were considered adequate in light of the high coefficient of determination values ($R^2 > 89\%$). The results of the intensive computational tests (permutation and bootstrap) indicated that the regression coefficients for all of the genotypes were considered equal to 1 (i.e., they exhibit broad stability). Furthermore, it should be highlighted that the application of intensive computational tests to the original data may have been subjected to the effects of different scales, and this would explain the results obtained in which all of the intercepts were considered equal to 1 ($P > 0.05$). The results of the permutation and bootstrap tests were markedly similar, and they could hence primarily be considered equivalent.

Based on the Scott and Knott (1974) test, the OP-S-16 line was considered adapted, but it was not responsive to environmental conditions (Table 1). Therefore, it is fitting to emphasize that this is a line obtained in the UFLA breeding program. Moreover, its performance was expected to be high, because it was derived from two well-adapted lines (Ouro Negro and Pérola), which are relatively resistant to angular leaf spot and to anthracnose.

Some changes occurred when data transformed into ranks were used (Table 2). With the exception of the Magnífico line, the fit of the model worsened, and this might be verified by the reduction of R^2. Moreover, the aforementioned line may be considered responsive. Based on the regression coefficient presented in Table 2, the OP-S-16 line cannot be considered biologically stable. Biological stability tends to occur in cultivars with the greatest adaptation (yield). The other cultivars, including the Iapar 81 and Talismã lines have broad stability. Furthermore, the bootstrap and permutation tests were again found to be equivalent.

The estimates of deviation in relation to the maximum performance (Pi) based on the Lin and Binns (1991) method (Table 3) showed that the genotypes that contributed the least to the interaction were Esal 693 and OP-S-16, which is in agreement with the results of tests involving regression coefficients and the original data (Table 1). Moreover, using the Lin and Binns (1988, 1991) method, Talismã and Magnífico contributed the most. Therefore, the OP-S-16, MH, MP, and MA-I-25 genotypes exhibited the lowest estimates, and were the populations nearest to the maximum (Table 3). As expected, this behavior was observed in a group of well-adapted lines and population mixtures (only one line was differentiated from the maximum). The fact that the mixture and multiple hybrids exhibited similar behaviors at the maximum and relatively low contributions to the interaction was likely due to a phenomenon known as genetic homeostasis. This theory suggests that a genotype (line, population, hybrid, or mixture) composed of a large quantity of heterozygous loci and mixtures of genotypes is more stable under adverse environmental conditions because of buffering. The bootstrap and permutation tests were also equivalent for the Lin and Binns (1988, 1991) method with the data transformed into ranks.

Considering that the P_i parameter used the ranks (Table 4), the OP-S-16 and Esal 693 lines continued contributing little to the interaction. However, in these cases, the populations with greater heterogeneity or heterozygosity also stood out among those that contributed less to the interaction, which reinforced the genetic homeostasis hypothesis. Estimates of the P values obtained using intensive computational tests led to the conclusion that the Magnífico ($P < 0.01$), Iapar 81 ($P < 0.01$), and Carioca-MG ($P < 0.05$) cultivars deviated significantly from the maximum, and the lines were considered unstable. The non-parametric method differed considerably from the parametric

method, and it allowed greater genotypic differentiation to be obtained in regard to stability.

The changes that occurred during the classification of the genotypes considering genotypic risk (Table 5), original index (IO_i), and the index based on ranks (IP_i), indicated that the effect of scales (heterogeneous variances) might lead to big changes in the inference of the risks associated with the adoption of a cultivar. This fact emphasizes the importance of the non-parametric methods developed or presented in this study. The OP-S-16 line was the highest yielding, with a mean rank value equal to 135.45%, and the variances were very similar (ranging from 40.8 to 59.3).

Although some changes occurred, the two indices highlighted that the genotype of least risk was the OP-S-16 line, because in the worst cases with 75% reliability, its mean yield was 2.53 or 3.70% greater than the general mean value of the environment. The variances were very similar, ranging from 40.8 to 59.3, which illustrated the homogeneity of the genotypic oscillations over the environments at the percentage scale that were obtained from the transformation of the original data into ranks. Therefore, using this method, the Esal 693, MH, Pérola, and MP populations may be highlighted as those that contributed least to the interaction, and the Magnífico, Iapar 81, and Talismã population contributed the most.

The hypothesis that there were no differences between the phenotypic stabilities of the genotypes was not rejected based on the nominal significance value of 5%, and this was shown by the estimated statistics and respective P values, with the exception of the Talismã line (Table 7). The Talismã line showed the greatest statistical test values, with $S_{1,5} = 3.98$, $S_{2,5} = 11.75$, $Z_{1,5} = 3.56$ (P = 0.0592), and $Z_{2,5} = 4.08$ (P = 0.0434). Moreover, Talismã was the population that contributed the most to the interaction, followed closely by the Esal 693 and Iapar 81 lines. The populations that contributed the least, based on the lowest Z statistics values, Si values, or values closest to their means, were MA-I-25, MP, Carioca-MG, and OP-S-16.

It should be emphasized that some methods may be considered complementary, and may have different purposes in relation to genotypic classification with regard to stability and adaptability. Thus, the use of some of these methodologies may be recommended in an associated way for the improved identification of genotypes and their classification with regard to stability. Furthermore, the fact that low correlations were obtained should also be emphasized. A possible reason may stem from the fact that the group of genotypes studied included well-adapted and stable lines. Moreover, they were also recommended cultivars in use, and they exhibited resistance to diseases. Thus, the stability parameters were quite similar among the genotypes, which resulted in low variability of the parameter estimates, and as a potential consequence, agreements were not greatly expected among the methods studied. The risk indices correlated with various other estimated parameters, and their correlations were negative; therefore, the higher the values, the lower the risk (Table 8). The opposite occurred with many of the other parameters, in which the higher their values, the less stable the genotypes. The ecovalences only correlated significantly among themselves and with the parameters related to the contribution of each genotype to the interaction (partitioning of the P_i values).

The parameters of Huhn (1990) only showed significant correlations among themselves and with the non-parametric risk indices (Table 8). Moreover, the correlation between the Huhn (1990) parameter estimates was very high (0.98), and it was considered the strongest among all the estimates. The non-parametric methods of stability proved to be effective in evaluating phenotypic stability, and intensive computational statistics allowed the inferences to be made. Moreover, the association with non-parametric methods resulted in efficient methods used to draw valid conclusions with respect to the phenotypic stability of the genotypes. In the presence of variance heterogeneity, the non-parametric methods showed greater discrimination power when

determining the phenotypic stability of the genotypes, and the computational statistical methods proved to be attractive.

In conclusion, the non-parametric approach proved to be effective, and it exhibited greater power to discriminate phenotypic stability in the presence of variance heterogeneity. The use of intensive computational statistics (bootstrap and permutation tests) is a good alternative for studies of adaptability and stability in the presence of heteroscedasticity, because there is no difference between the two methods. The correlation with the Huhn (1990) parameters was only significant when associated with the non-parametric risk index. This result indicated that most methods used may not contain the same information, but are considered complementary.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Annicchiarico P (1992). Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. *J. Genet. Breed.* 46: 269-278.

Annicchiarico P, Bertolini M and Mazzinelli G (1995). Analysis of genotype-environment interactions for maize hybrids in Italy. *J. Genet. Breed.* 49: 61-68.

Cochran WG (1954). Some methods for strengthening the common c^2 tests. *Biometrics* 10: 417-451. http://dx.doi.org/10.2307/3001616

Bruzi AT, Ramalho MAP and Abreu AFB (2007). Desempenho de famílias do cruzamento entre linhagens de feijões andinos e mesoamericanos em produtividade e resistência a *Phaeoisariopsis griseola. Cienc. Agrotec.* 31: 650-655. http://dx.doi.org/10.1590/S1413-70542007000300008

Crossa J (1990). Statistical analysis of multilocation trials. *Adv. Agron.* 44: 55-85. http://dx.doi.org/10.1016/S0065-2113(08)60818-4

Gauch HG, Jr. and Zobel RW (1988). Predictive and postdictive success of statistical analyses of yield trials. *Theor. Appl. Genet.* 76: 1-10. http://dx.doi.org/10.1007/BF00288824

Huhn M (1990). Nonparametric measures of phenotypic stability. Part 1: theory. *Euphytica* 47: 89-194.

Huhn M and Nassar R (1989). On tests of significance for nonparametric measures of phenotypic stability. *Biometrics* 45: 997-1000. http://dx.doi.org/10.2307/2531698

Huhn M and Nassar R (1991). Phenotypic stability of genotypes over environments: on tests of significance for two nonparametric measures. *Biometrics* 47: 1196-1197.

Lin CS and Binns MR (1988). A superiority measure of cultivar performance for cultivar x location data. *Can. J. Plant Sci.* 68: 193-198. http://dx.doi.org/10.4141/cjps88-018

Lin CS and Binns MR (1991). Assessment of a method for cultivar selection based on regional trial data. *Theor. Appl. Genet.* 82: 379-388. http://dx.doi.org/10.1007/BF02190626

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. http://dx.doi.org/10.2307/2531947

Nassar R, Leon J and Huhn M (1994). Tests of significance for combined measures of plant-stability and performance. *Biometrical J.* 36: 109-123. http://dx.doi.org/10.1002/bimj.4710360115

Piepho HP (1997). Distribution-free tests for one-way homoscedasticity in a two-way classification. *Biometrics* 53: 340-351. http://dx.doi.org/10.2307/2533119

Plaisted RL and Paterson LC (1959). A technique for evaluating the ability of selections to yield consistently in different locations or seasons. *Am. Potato J.* 36: 381-385. http://dx.doi.org/10.1007/BF02852735

Scott A and Knott M (1974). Cluster-analysis method for grouping means in analysis of variance. *Biometrics* 30: 507-512. http://dx.doi.org/10.2307/2529204

Wricke G (1964). Zur berechnung der okovalenz bei sommerweizen und hafer. *Z. Pflanzenzuchtung-. J. Plant Breed* 52: 127-138.
Yau SK (1995). Regression and AMMI analysis of genotype x environment interactions: an empirical comparison. *Agron. J.* 87: 121-126. http://dx.doi.org/10.2134/agronj1995.00021962008700010021x