

Violations of normality and data loss compromise accurate soybean genotypes selection

Violações da normalidade e perda de dados comprometem a seleção precisa de genótipos da soja

M. V. Loro^{1*}; I. R. Carvalho²; M. M. Moraes¹; G. M. Froehlich¹; V. L. Bueno¹;
A. Guareschi¹; D. A. Bromberger¹

¹Departamento de Fitotecnia, Universidade Federal de Santa Maria, 97105-900, Santa Maria-RS, Brasil

²Universidade Regional do Noroeste do Estado do Rio Grande do Sul, 98700-000, Ijuí-RS, Brasil

*muriloloro@gmail.com

(Recebido em 18 de agosto de 2025; aceito em 23 de junho de 2026)

In agronomic experiments, violations of normality of errors and imbalances are frequent, compromising the accuracy of inferences. Evaluating adaptability and stability methodologies regarding their robustness under these conditions is essential to ensure reliability in genotype selection and to assist breeding programs in environments with non-ideal data. This study aimed to evaluate the performance of biometric models in identifying promising genotypes using data from experiments that violated the normality assumption in errors and exhibited experimental unbalance. Grain yield data from 18 soybean genotypes evaluated in six Rio Grande do Sul environments were used. Five experimental scenarios were generated to examine the normality of errors and experimental unbalance at levels of 10%, 20%, and 30%. Adaptability and stability methods were applied, including those of Wricke, Annicchiarico, Lin and Binns, Eberhart and Russell, AMMI, and GGE biplot. The genotypes were ranked and compared to identify robust parameters using Spearman correlations, principal component analysis and Euclidean distance. The Annicchiarico and Lin and Binns methodologies show greater agreement in classifying genotypes under conditions of normality violation and a 10% unbalance. The parameters of the GGE and AMMI models exhibit lower agreement in classifying genotypes in data with violations of normality and unbalances of 20% and 30%.

Keywords: genetic improvement, genotype $\times$ environment interaction, biometric models.

Em experimentos agrônômicos, violações da normalidade dos erros e desbalanceamentos são frequentes, comprometendo a precisão das inferências. Avaliar metodologias de adaptabilidade e estabilidade quanto à sua robustez nessas condições é essencial para garantir confiabilidade na seleção de genótipos e auxiliar programas de melhoramento em ambientes com dados não ideais. Este estudo teve como objetivo avaliar o desempenho de modelos biométricos na identificação de genótipos promissores, utilizando dados de experimentos que violaram a suposição de normalidade nos erros e apresentaram desbalanceamento experimental. Foram utilizados dados de produtividade de grãos de 18 genótipos de soja avaliados em seis ambientes do Rio Grande do Sul. Cinco cenários experimentais foram gerados para examinar a normalidade dos erros e o desbalanceamento experimental nos níveis de 10%, 20% e 30%. Foram aplicados métodos de adaptabilidade e estabilidade, incluindo os de Wricke, Annicchiarico, Lin e Binns, Eberhart e Russell, AMMI e GGE biplot. Os genótipos foram ranqueados e comparados para identificar parâmetros robustos, utilizando correlações de Spearman, análise de componentes principais e distância euclidiana. As metodologias de Annicchiarico e Lin and Binns demonstram maior concordância na classificação de genótipos em condições de violação da normalidade e desbalanceamento de 10%. Os parâmetros dos modelos GGE e AMMI apresentam menor concordância na classificação de genótipos em dados com violações da normalidade e desbalanceamentos de 20% e 30%.

Palavras-chave: melhoramento genético, interação genótipo $\times$ ambiente, modelos biométricos.

1. INTRODUCTION

The evaluation of superior genotypes is conducted through a series of experiments that involve networks of environments [1]. In these cases, the phenotypic responses of the genotypes (G) are studied as a function of the interaction with these environments (E). The phenotypic expression of a genotype refers to the response of gene expression under the influence of

environmental factors, including soil and climate [2]. When evaluating genotypes in tests, the $G \times E$ interaction is significant; biometric models are used to estimate stability and adaptability parameters, thereby highlighting promising genotypes for the evaluated conditions [3].

The number of genotypes and environments is generally significant, requiring substantial labor and time to conduct the experiments. Since the tests are performed in multiple environments, there are challenges in installation, execution, and evaluation. However, it is important to maintain the homogeneity and standardization of the procedures involved in conducting experiments to minimize systematic errors arising from both controllable and non-controllable sources of variation that may affect statistical inferences.

Certain assumptions must be met to analyze the effects of the factors that constitute the mathematical model linked to an experiment evaluating genotypes across multiple environments. The normality of errors is one of the primary premises that must be met to enable the application of variance analysis and validate the results of the hypothesis [4]. Sometimes, there are situations where this assumption is not met, and if the analyses are still carried out, certain complications may affect the results, generating irrelevant information [4, 5].

However, in field conditions, losses of experimental units are common, which compromises the orthogonality of the data and results in an increase in the residual variances associated with the mean square error, as these units become uninformative. Value for Cultivation and Use (VCU) trials, in turn, tend to be unbalanced due to the replacement of genotypes over the years [6–8]. Strategies have been adopted to ensure the inferences' reliability to overcome the unbalance, such as using GGE biplot graphs, environmental matrices, and similarity networks between environments [7].

However, studies investigating the impacts of unbalance without imputing missing data and the violation of the assumption of normality of errors on the adaptability and stability parameters used in genotype selection are still scarce. Considering that biometric models exhibit varying sensitivity to these factors, understanding this sensitivity can support more accurate decisions in genetic improvement programs, thereby favoring the selection of more appropriate methodologies and enhancing selection accuracy. In this context, the present study aimed to evaluate the performance of biometric models in identifying promising genotypes, using data from experiments that violated the assumption of normality in errors and exhibited experimental unbalance.

2. MATERIAL AND METHODS

2.1 Experimental design

Grain yield (GY, kg ha^{-1}) data from 18 soybean genotypes were evaluated across six environments in the state of Rio Grande do Sul - RS during the 2020/2021 season: Santa Rosa - RS, Tenente Portela - RS, Campos Borges - RS, Sarandi - RS, Arroio Grande in Várzea - RS, and Arroio Grande in Coxilha - RS. The 18 genotypes were evaluated in a randomized block experimental design with three replications, totaling 324 experimental units in all environments. Sowing was carried out in all environments in the first half of November 2020.

Sowing was carried out with the distribution of 12 viable seeds per linear meter, using base fertilization with 250 kg ha^{-1} of formulated fertilizer 02-23-23 (N-P-K). Base fertilization was standardized in all environments, according to the management protocol adopted for the experimental trial network. The experiments were conducted under rainfed conditions, without supplemental irrigation.

The experimental units consisted of five seeding rows spaced 45 cm apart and five meters long, totaling an area of 11.5 m^2 per plot. Pest and disease control was conducted uniformly in all experiments to minimize biotic stress on the plants. Harvests were carried out in the second half of March 2021, when the plants reached full physiological maturity (phenological stage R7). Grain yield (kg ha^{-1}) was estimated from the manual harvesting of all plants in the two central rows of each experimental unit, with the values subsequently adjusted to a standard moisture content of 13%.

2.2 Statistical procedures

The normality of the errors was diagnosed using the Shapiro-Wilk test, defining the null hypothesis H_0 : the residues follow a normal distribution, and the alternative hypothesis H_1 : the residues do not follow a normal distribution. The significance level was set at 5%. Given the lack of adjustment of the residues to the normal distribution, the data were transformed using the Box-Cox family of transformations. This transformation consisted of the following equation:

$$X' = \frac{X^\lambda - 1}{\lambda}$$

Where, X' is the transformed variable, and λ is the parameter to be estimated from the data, thus the variable X is transformed into a new scale better fitting to the normal distribution, $X \sim N(0, 1)$. λ was defined by the amplitude of $0.1 \leq \lambda \leq 0.9$, and the best fit occurred with a value of $\lambda = 0.5$. Thus, two scenarios were formed: data with normality of errors and balanced (I) and data violating the assumption of normality of errors and balanced (II).

Subsequently, with the data with the normality of errors and balanced (I), random simulations of experimental unbalance were performed, generating three scenarios: data with standard errors and unbalances at 10% (III), 20% (IV), and 30% (V). To generate the scenarios of 10, 20, and 30% unbalance, normal distribution data were used, and thus, 32, 64, and 97 experimental units were randomly removed using the sample function of the R Software [9]. Although the unbalance was performed, all genotypes were evaluated in all environments. Thus, five scenarios were obtained: I: data with the normality of errors and balance; II: data with violation of the assumption of normality in the mistakes and balance; III: data with the normality of errors and unbalance by 10%; IV: data with the normality of errors and unbalanced by 20%; V: data with the normality of errors and unbalanced by 30%.

Since significant genotype $\times$ environment interaction was observed in the five scenarios (I, II, III, IV, and V), the genotype $\times$ environment interaction was studied through adaptability and stability analyses for each scenario. For this purpose, methodologies based on the analysis of variance proposed by Wricke (1965) [10] and Annicchiarico (1992) [11], based on linear regression: Eberhart and Russel (1966) [12]; based on nonparametric statistics: Lin and Binns method (1988) modified by Carneiro [13]; AMMI-Biplot (Additive Main Effects and Multiplicative Interaction) based on the estimation of additive main effects and multiplicative interaction [14]; and GGE biplot [15] were used.

The parameters obtained from the biometric models were estimated for each scenario (I, II, III, IV, and V), considering that some methods provide information on adaptability and stability, whereas others are specifically focused on genotype stability. The genotypes were then ranked according to the specific criteria of each method: (1) Wricke (1965): the ranking was performed by assigning the best classification to the genotype with the lowest covalence value (W_i), indicating greater stability. (2) Annicchiarico: the genotypes were classified based on the confidence index, with the best position being assigned to those with the highest values, considering three environmental conditions: general ($W_{i \text{ Gen}}$), favorable ($W_{i \text{ Fav}}$) and unfavorable ($W_{i \text{ Desf}}$). (3) Lin e Binns: the classification followed a logic similar to that of the Annicchiarico method, with the difference that the genotypes with the lowest values of the P_i parameter received the best positions, also considering the environments in general ($P_{i \text{ Gen}}$), favorable ($P_{i \text{ Fav}}$) and unfavorable ($P_{i \text{ Desf}}$). (4) Eberhart e Russell (1966): the genotypes were ranked for each of the three parameters: (i) Slope coefficient (b_1): best classification assigned to the genotype with the lowest b_1 ; (ii) variance of deviations (S^2): genotypes with the lowest absolute value of S^2 received the best ranks; and (iii) coefficient of determination (R^2): better classifications were attributed to genotypes with the highest R^2 values. (5 and 6) AMMI and GGE: the stability of the genotypes was evaluated based on the scores of the principal components (PC_1 and PC_2), with the best classifications being attributed to those with the lowest absolute values of the scores. Thus, the genotypes with the lowest scores in PC_1 ($PC_{1 \text{ AMMI}}$ and $PC_{1 \text{ GGE}}$) received the first classifications; the same was done for PC_2 ($PC_{2 \text{ AMMI}}$ and $PC_{2 \text{ GGE}}$).

Based on the ranking of the genotypes, a matrix of dimensions 7×254 was constructed, consisting of the following columns: Parameter, genotype, scenario I, II, III, IV, and V. The lines represent the ranks of the genotypes in each scenario for each adaptability and stability parameter. Based on this matrix, the Spearman correlation coefficients were individually calculated between scenarios I and II, III, IV, and V for each parameter. The resulting correlation matrix was used to generate boxplots to visualize the distribution of the correlation coefficients by parameter.

Next, a line called “Ideal” was added to the correlation matrix, consisting of values equal to 1 (one) in all scenarios, as a reference. This new matrix was subjected to principal component analysis (PCA), aiming to identify patterns of similarity between the parameters. In the biplot generated, the parameters closest to the “Ideal” line indicate greater agreement with scenario I and less influence of non-normality or unbalance effects. Subsequently, a column “I” was added to the correlation matrix, containing values equal to 1 (one) for all parameters and reference purposes. The Euclidean distances between the scenarios (I to V) were calculated using this matrix. The scenarios with the smallest distance about scenario I indicate less variation in the genotype ranks, that is, greater stability of the ordering about the base scenario. All analyses were performed using R software version 4.1.3 [9] using the MASS packages [16] to perform the Box-Cox transformation and metan [17] to estimate the adaptability and stability parameters.

3. RESULTS AND DISCUSSION

The comparison of the adaptability and stability parameters between the scenarios based on the genotype classifications revealed that the biometric models present different results in the classification of genotypes according to the scenarios (Table 1). For example, the five best genotypes based on the Wricke (1965) [10] method in scenario I (data with the normality of errors and balance) were G13, followed by G11, G6, G7, and G3; based on scenario II (data with violation of the assumption of normality of errors and balanced) were G13, followed by G11, G17, G7, and G3; based on scenario III (data with normality of errors and unbalanced by 10%) were G13, G6, G7, G3 and G4. By analyzing only, the genotype with the greatest stability, it is possible to show that in all scenarios, except scenario V (data with normality of errors and unbalanced by 30%), the genotype G13 was the most stable. This indicates that although data violates the assumption of normality of errors and unbalances up to 20%, there is a tendency to identify the genotype with greater stability.

Thus, it is observed that scenario II shows greater agreement with scenario I in the classification of genotypes in the first five positions. It is possible to tolerate a reasonable deviation from normality with little effect on the analysis of variance [4]. However, reflections on the adaptability and stability parameters were not found, but it is suggested that the parameter estimates are similar to those of normal data. The violation of the assumption of normality of errors does not lead to significant errors in estimating the mean. However, this is sensitive in the presence of outliers in the database [4]. Therefore, in some models, the bias between the classification of scenarios I and II can be explained by the over- or underestimation of the mean of the genotypes in the environments.

The most significant disagreements in the classification of genotypes were associated with the second principal component of AMMI and GGE (Table 1). For example, for the PCA_2 of AMMI, the genotypes classified as stable in scenario I were G18, G7, G10, and G3, while in the other scenarios, only scenario II corresponded with the same genotype (G10) in the fourth position. This reveals that the PCA_2 of AMMI is sensitive and tends to present biased explanations of the data when using experiments with unbalanced data and with a violation of the assumption of normality of errors. The same is observed for the PCA_2 of GGE, in which low correspondence of the ranking of genotypes is observed. Schmidt et al. (2019) [6] revealed that experimental unbalance overestimates heritability estimates in the broad sense. Port et al. (2024) [8] reported that experimental unbalance of up to 10% is acceptable in maize trials, as it

maintains the accuracy of broad-sense heritability estimates. However, unbalances exceeding this threshold tend to overestimate such estimates.

Table 1: Ranking of genotypes in scenarios of data with normality of errors and balanced (I), data with violation of the assumption of normality in errors and balanced (II); data with normality of errors and unbalanced by 10% (III); data with normality of errors and unbalanced by 20% (IV) and data with normality of errors and unbalanced by 30% (V) for each adaptability and stability parameter.

Scenarios	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18
W_i (Wricke)																		
I	16	13	5	7	10	3	4	14	11	9	2	15	1	12	8	17	6	18
II	16	13	5	7	14	6	4	9	10	11	2	15	1	12	8	17	3	18
III	16	13	4	5	8	2	3	15	11	9	10	14	1	12	7	17	6	18
IV	15	13	7	4	12	3	2	18	8	9	11	14	1	6	5	17	10	16
V	15	12	14	2	10	4	3	18	8	9	7	13	11	5	1	17	6	16
b₁ (Eberhart e Russel)																		
I	18	10	7	6	15	14	8	17	11	3	5	1	4	2	13	9	12	16
II	16	9	8	4	17	14	7	12	11	3	5	1	6	2	15	13	10	18
III	18	12	7	6	9	14	11	17	10	3	4	1	5	2	13	8	15	16
IV	18	16	9	6	8	14	11	15	10	3	4	1	5	2	12	7	13	17
V	18	15	13	8	7	17	11	14	9	3	4	1	5	2	10	6	12	16
S² (Eberhart e Russel)																		
I	16	14	6	8	12	1	2	15	13	11	7	4	10	3	9	18	5	17
II	16	15	7	6	14	2	4	12	13	11	9	5	10	3	8	18	1	17
III	15	14	8	6	11	1	5	16	13	10	12	4	7	2	9	18	3	17
IV	15	14	10	4	13	2	3	18	11	8	9	7	1	5	6	17	12	16
V	15	13	14	3	12	1	5	17	10	7	6	9	11	2	4	18	8	16
R² (Eberhart e Russel)																		
I	10	13	9	12	4	2	7	8	11	15	3	18	1	16	6	17	5	14
II	11	12	9	13	6	3	8	7	10	15	2	18	1	16	5	17	4	14
III	6	11	8	12	7	1	4	10	9	14	15	18	2	16	5	17	3	13
IV	5	8	9	10	11	1	2	16	7	15	14	18	4	13	3	17	6	12
V	7	5	9	6	12	1	2	14	8	16	13	17	15	11	3	18	4	10
W_i Gen (Annicchiarico)																		
I	17	11	7	16	9	5	13	18	15	10	6	14	3	2	1	8	12	4
II	17	11	7	16	10	5	13	18	15	9	6	14	3	2	1	8	12	4
III	17	12	6	16	7	5	15	18	14	9	11	13	1	3	2	8	10	4
IV	17	7	6	16	9	5	14	18	15	11	12	13	3	2	1	10	8	4
V	17	10	3	16	9	5	15	18	14	11	12	13	6	2	1	8	7	4
W_i Fav (Annicchiarico)																		
I	11	14	9	16	7	5	13	17	15	12	8	18	2	3	1	10	6	4
II	11	12	9	17	8	5	14	16	15	13	7	18	3	4	1	10	6	2
III	12	11	8	16	7	4	13	17	15	10	14	18	2	5	1	9	6	3
IV	11	8	7	16	9	3	10	18	15	12	13	17	5	4	1	14	6	2
V	12	7	4	17	10	2	9	18	15	11	13	16	8	5	1	14	6	3
W_i Desf (Annicchiarico)																		
I	17	10	6	15	12	9	13	18	14	4	3	5	2	1	8	7	16	11
II	17	11	6	15	12	9	13	18	14	3	4	7	2	1	8	5	16	10
III	17	12	4	14	9	10	15	18	13	5	3	7	2	1	8	6	16	11
IV	17	12	10	16	8	9	14	18	13	4	3	6	2	1	7	5	15	11
V	17	12	8	14	4	11	15	18	13	7	5	9	2	1	6	3	16	10
P_i Gen (Lin e Binns)																		
I	14	11	7	17	9	6	13	18	15	12	8	16	4	2	1	5	10	3
II	13	11	7	17	8	6	14	18	15	12	9	16	5	4	1	3	10	2
III	14	12	6	17	8	7	15	18	13	10	11	16	2	4	1	5	9	3
IV	15	9	5	17	10	6	14	18	16	12	11	13	2	3	1	7	8	4
V	14	9	2	17	10	5	15	18	16	11	12	13	8	4	1	7	6	3

P _i Fav (Lin e Binns)																		
I	10	12	8	16	6	4	14	17	15	13	11	18	5	3	1	7	9	2
II	9	12	8	17	7	4	15	16	14	13	11	18	5	6	2	3	10	1
III	10	11	6	17	8	4	14	16	15	12	13	18	3	7	1	5	9	2
IV	10	8	6	17	11	4	14	18	15	13	12	16	3	5	1	9	7	2
V	8	7	1	17	11	4	12	18	15	13	14	16	10	5	3	9	6	2
P _i Desf (Lin e Binns)																		
I	17	9	7	16	12	11	14	18	15	8	6	3	2	1	5	4	13	10
II	17	9	6	16	13	11	14	18	15	8	7	5	2	1	4	3	12	10
III	17	12	6	15	9	11	16	18	14	8	5	7	2	1	4	3	13	10
IV	17	12	9	15	8	11	16	18	14	6	5	7	2	1	3	4	13	10
V	17	11	8	15	5	12	16	18	14	7	6	9	3	1	4	2	13	10
PC ₁ AMMI																		
I	16	12	9	2	7	6	8	13	10	14	4	15	3	11	5	17	1	18
II	16	12	11	2	8	6	9	7	10	15	4	14	3	13	5	17	1	18
III	16	13	9	3	11	4	6	15	8	14	7	12	1	10	5	17	2	18
IV	15	13	10	3	14	7	5	18	8	12	6	11	2	9	4	16	1	17
V	16	12	15	3	13	5	6	18	9	10	7	11	2	8	4	14	1	17
PC ₂ AMMI																		
I	11	12	4	8	17	13	2	5	10	3	6	18	7	16	15	14	9	1
II	8	15	4	16	17	11	1	10	12	5	3	18	6	13	14	7	2	9
III	15	1	7	12	3	10	14	17	5	6	16	4	11	9	2	18	13	8
IV	16	10	2	12	5	8	9	14	6	11	17	1	13	3	7	18	15	4
V	8	11	14	12	3	7	10	13	1	9	16	4	17	2	5	18	15	6
PC ₁ GGE																		
I	4	6	10	14	5	2	12	16	13	9	1	15	7	8	11	17	3	18
II	11	8	9	12	6	3	15	10	13	14	5	16	4	1	7	17	2	18
III	7	5	8	15	4	1	14	16	10	6	2	11	12	9	13	17	3	18
IV	17	14	5	13	12	9	3	18	1	10	8	4	7	16	15	6	2	11
V	17	13	11	10	14	6	1	18	3	9	7	8	4	15	12	5	2	16
PC ₂ GGE																		
I	17	12	4	14	6	8	1	18	3	11	5	7	9	16	15	10	2	13
II	16	12	7	15	13	11	6	18	1	9	3	5	10	14	17	8	2	4
III	17	10	6	9	13	5	2	18	4	12	7	8	3	15	11	14	1	16
IV	2	6	12	13	8	3	14	11	16	7	4	15	10	5	9	17	1	18
V	2	5	15	14	8	1	13	16	11	6	4	12	9	7	10	18	3	17

W_i: Wricke ecovalence parameter; b_i: Eberhart and Russell slope coefficient; S²: Eberhart and Russell regression deviation; R²: Eberhart and Russell coefficient of determination; W_i Gen: Annicchiarico general ecovalence parameter; W_i Fav: Annicchiarico favorable ecovalence parameter; W_i Desf: Annicchiarico unfavorable ecovalence parameter; P_i Gen: Lin and Binns general ecovalence parameter; P_i Fav: Lin and Binns favorable ecovalence parameter; P_i Desf: Lin and Binns unfavorable ecovalence parameter; PC₁ AMMI: AMMI first principal component scores; PC₂ AMMI: AMMI second principal component scores; PC₁ GGE: GGE first principal component scores; PC₂ GGE: GGE second principal component scores.

When comparing hypothesis tests of balanced and unbalanced data in the analysis of variance, Landsheer et al. (2015) [18] revealed that the rejection rate of H₀ (null hypothesis) of the main effects was higher in unbalanced data. This indicates overestimating the mean of some treatments and underestimating others, evidencing a greater discrepancy in the mean of the treatments. In this case, with unbalanced data, it can be inferred that in the presence of unbalance, the number of repetitions is reduced, and this promotes effects on the estimation of the mean of the genotypes and environments. These effects of over or underestimation are accentuated as there is a greater level of unbalance.

When comparing the ranked positions of the genotypes in scenario I with the other scenarios, regarding the stability and adaptability parameters estimates, the Spearman correlation coefficient varied from -0.42 to 1.00 (Figure 1). This indicates changes in the concordance of the classification of the genotypes. Spearman correlation coefficients equal to -1 and +1 reflect the perfect negative and positive relationship between two variables.

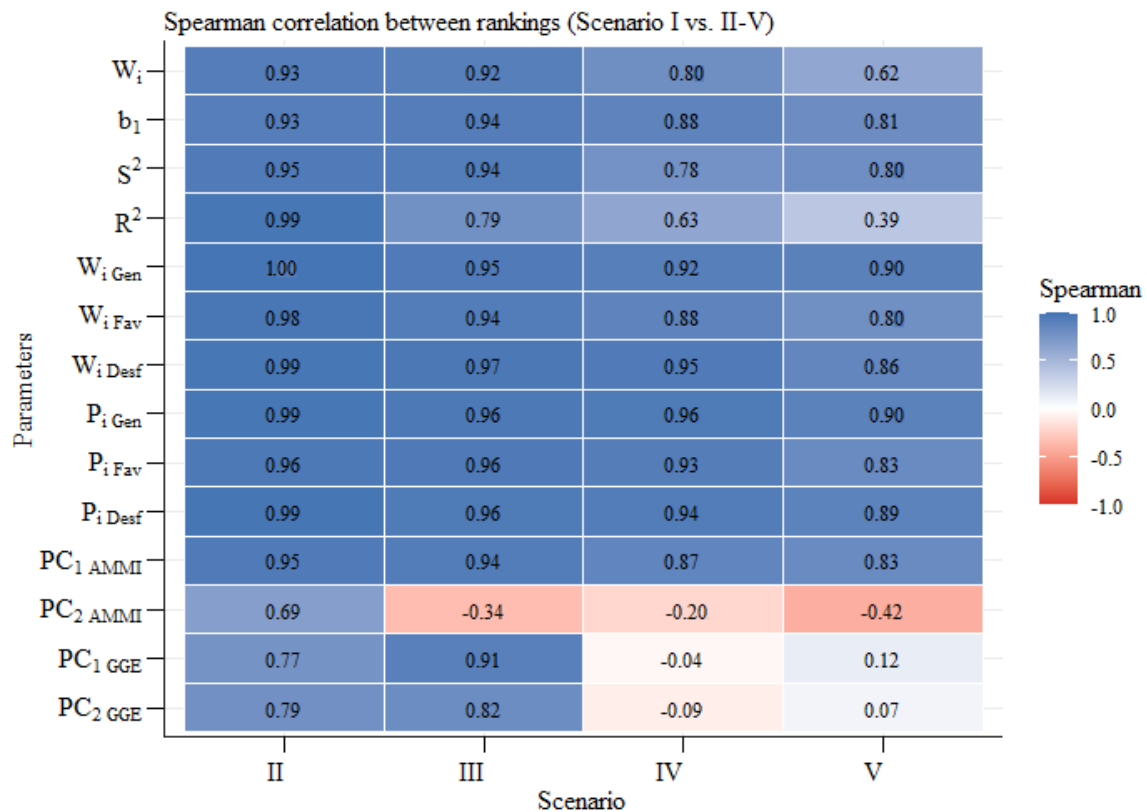


Figure 1: Spearman correlation coefficients between the rankings of genotypes in the scenario with balanced error normality (I) and the scenarios of data with violation of the assumption of normality in errors and balanced (II); data with error normality and unbalanced at 10% (III); at 20% (IV) and at 30% (V) for each adaptability and stability parameter. W_i : Wricke ecovalence parameter; b_1 : Eberhart and Russell slope coefficient; S^2 : Eberhart and Russell regression deviation; R^2 : Eberhart and Russell coefficient of determination; $W_i \text{ Gen}$: Annicchiarico general ecovalence parameter; $W_i \text{ Fav}$: Annicchiarico favorable ecovalence parameter; $W_i \text{ Desf}$: Annicchiarico unfavorable ecovalence parameter; $P_i \text{ Gen}$: Lin and Binns general ecovalence parameter; $P_i \text{ Fav}$: Lin and Binns favorable ecovalence parameter; $P_i \text{ Desf}$: Lin and Binns unfavorable ecovalence parameter; $PC_1 \text{ AMMI}$: AMMI first principal component scores; $PC_2 \text{ AMMI}$: AMMI second principal component scores; $PC_1 \text{ GGE}$: GGE first principal component scores; $PC_2 \text{ GGE}$: GGE second principal component scores.

A gradual reduction in Spearman's correlation coefficients was observed from scenarios II to V in all parameters evaluated. This result indicates that using data that violated the assumption of normality of errors but had a balanced structure resulted in greater agreement in the classification of genotypes compared to the analysis with unbalanced data and without violation of the normality of errors. The parameters b_1 , $W_i \text{ Gen}$, $W_i \text{ Fav}$, $W_i \text{ Desf}$, $P_i \text{ Gen}$, $P_i \text{ Fav}$, $P_i \text{ Desf}$ e $PC_1 \text{ AMMI}$ presented Spearman's correlation coefficients greater than 0.80 in the four scenarios evaluated (II, III, IV, and V), indicating high consistency in the classification of genotypes about the reference scenario (scenario I). $PC_2 \text{ AMMI}$, $PC_1 \text{ GGE}$, and $PC_2 \text{ GGE}$ presented the lowest Spearman's correlation coefficients, with negative values. This result indicates a greater sensitivity of these parameters to variations in the data, especially in conditions that violate the assumption of normality of errors and the presence of unbalance. The $PC_1 \text{ GGE}$ and $PC_2 \text{ GGE}$ parameters exhibited the most significant variation in coefficients, with greater disagreement in the ranking of genotypes observed when applied to data with 20 and 30% unbalance (scenarios IV and V) (Figure 2).

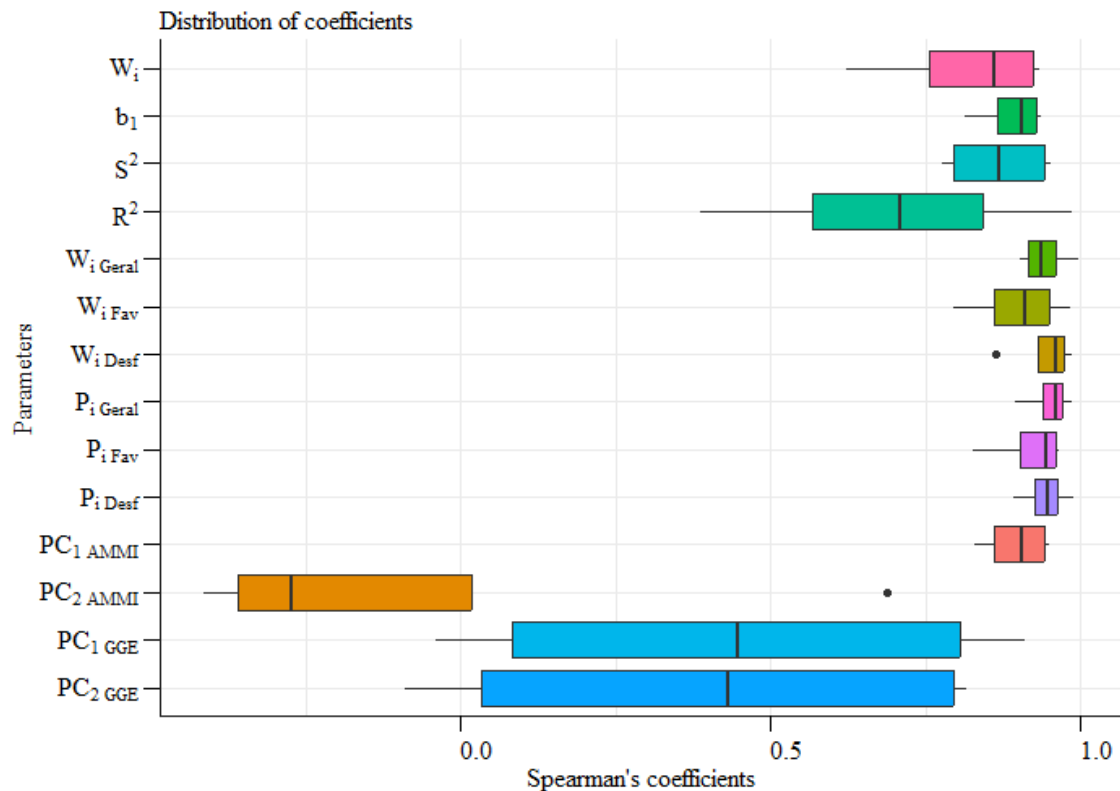


Figure 2: Distribution of Spearman correlation coefficients between genotype rankings regardless of the scenario for each adaptability and stability parameter. W_i : Wricke ecovalence parameter; b_1 : Eberhart and Russell slope coefficient; S^2 : Eberhart and Russell regression deviation; R^2 : Eberhart and Russell coefficient of determination; W_i Gen: Annicchiarico general ecovalence parameter; W_i Fav: Annicchiarico favorable ecovalence parameter; W_i Desf: Annicchiarico unfavorable ecovalence parameter; P_i Gen: Lin and Binns general ecovalence parameter; P_i Fav: Lin and Binns favorable ecovalence parameter; P_i Desf: Lin and Binns unfavorable ecovalence parameter; PC_1 AMMI: AMMI first principal component scores; PC_2 AMMI: AMMI second principal component scores; PC_1 GGE: GGE first principal component scores; PC_2 GGE: GGE second principal component scores.

In the biplot graph, the two principal components explained 98.00% of the total data variability. From the biplot graph, it was confirmed that the parameters b_1 , W_i Gen, W_i Fav, W_i Desf, P_i Gen, P_i Fav, P_i Desf, and PC_1 AMMI present greater agreement with the scenario I (ideal) in the ranking of genotypes since it is visually possible to observe the smallest distances between the points of these parameters about the "Ideal" point (triangle) (Figure 3). Contrary to the parameters, PC_2 AMMI, PC_1 GGE, and PC_2 GGE presented the most incredible distances about the "Ideal" point, revealing greater disagreement with scenario I in the ranking of genotypes. Thus, it can be understood that in situations where these scenarios are evaluated, using the second principal component of AMMI to identify adapted and stable genotypes tends to provide biased information. Similarly, estimates of adaptability and stability parameters using the GGE method should be avoided when there is more than a 10% unbalance. Inferences about GGE method parameters are not recommended when there is more than 30% of estimated data [19]. However, in this study, without estimating missing values, it is observed that caution is needed when using GGE with an unbalance greater than 10%. This result indicates the most significant discrepancy in the indication of genotypes; possibly, the most promising genotypes were classified as the most unstable and with the least adaptability.

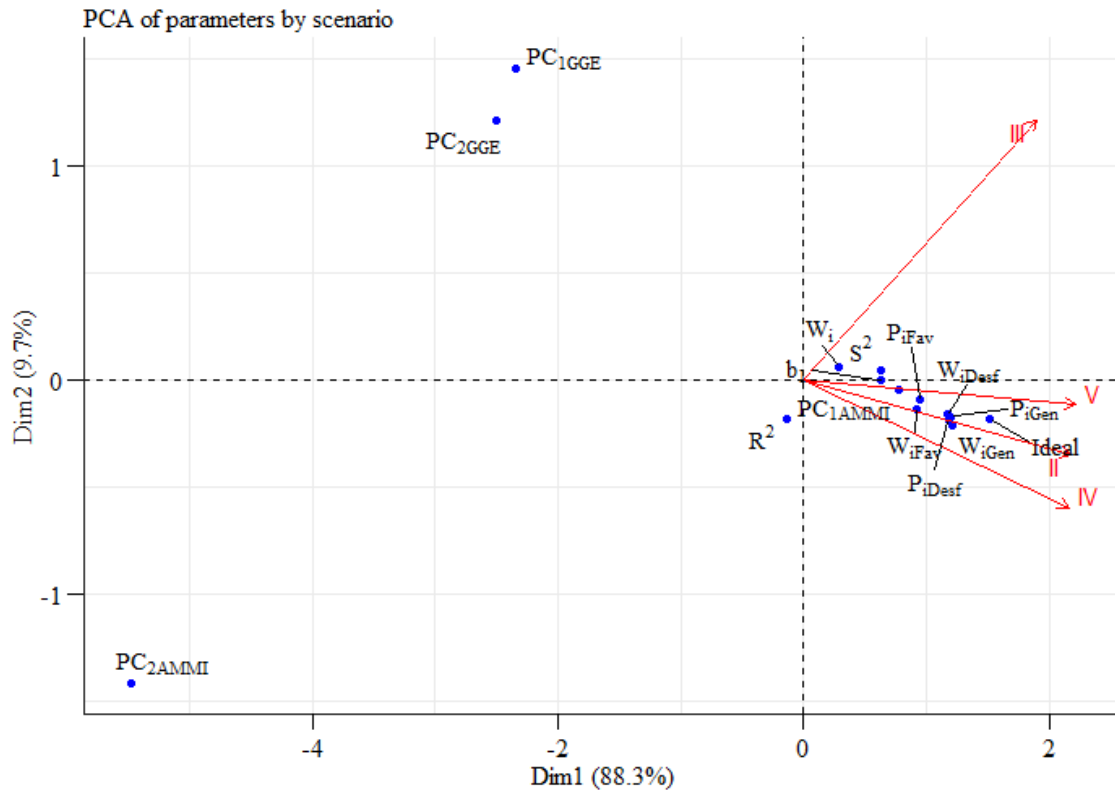


Figure 3: Biplot of principal component analysis for Spearman correlation coefficients between genotype rankings in the scenario with balanced error normality (I) and the scenarios of data with violation of the error normality assumption and balanced (II); data with error normality and unbalanced at 10% (III); at 20% (IV) and at 30% (V) for each adaptability and stability parameter. W_i : Wricke ecovalence parameter; b_1 : Eberhart and Russell slope coefficient; S^2 : Eberhart and Russell regression deviation; R^2 : Eberhart and Russell coefficient of determination; W_i Gen: Annicchiarico general ecovalence parameter; W_i Fav: Annicchiarico favorable ecovalence parameter; W_i Desf: Annicchiarico unfavorable ecovalence parameter; P_i Gen: Lin and Binns general ecovalence parameter; P_i Fav: Lin and Binns favorable ecovalence parameter; P_i Desf: Lin and Binns unfavorable ecovalence parameter; PC_{1AMMI} : AMMI first principal component scores; PC_{2AMMI} : AMMI second principal component scores; PC_{1GGE} : GGE first principal component scores; PC_{2GGE} : GGE second principal component scores.

From the Euclidean distance, it was observed that scenario II presented greater agreement with scenario I regardless of the parameter evaluated (Figure 4). This result indicates that data violating the normality assumption of errors maintain greater similarity with the reference scenario than data with unbalance. In addition, the presence of unbalance $\geq 20\%$ (scenarios IV and V), even with normality met, promotes greater disagreement in the ranking of genotypes. This may be associated with biased mean estimates in scenarios IV and V. This is expected since, in these scenarios, the mean is estimated by fewer observations, causing greater variation. Generally, high variability in mean estimates is not precise and should be viewed cautiously.

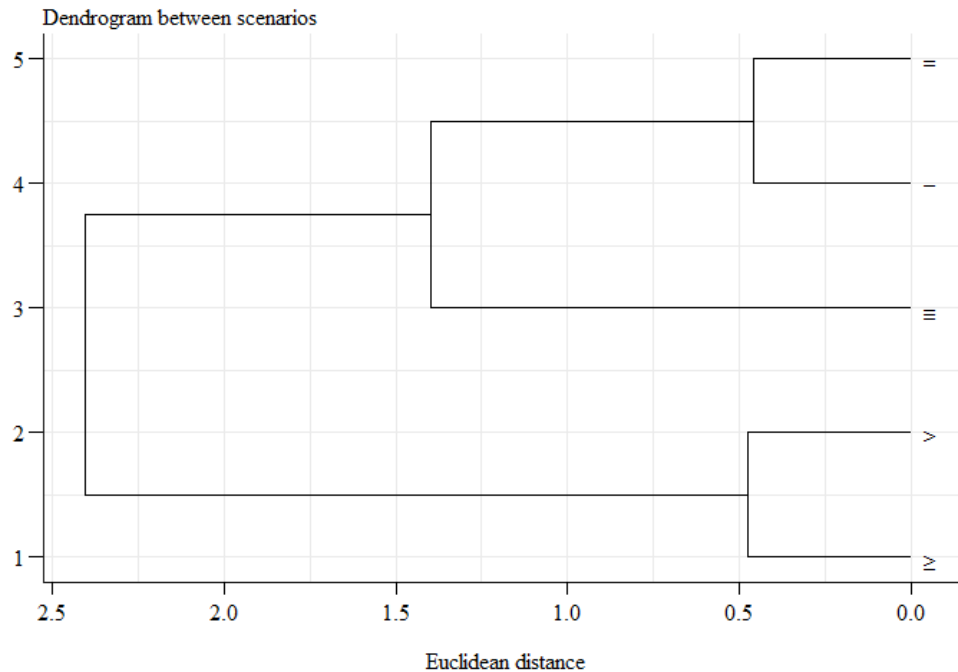


Figure 4: Dendrogram of Euclidean distance measures between scenarios with balanced error normality (I) and scenarios with data that violated the assumption of normality in errors and balanced (II); data with error normality and unbalanced by 10% (III); by 20% (IV) and by 30% (V) based on the adaptability and stability parameters.

The results shown in the study indicate that the estimates of the scores of the main components of the AMMI and GGE methodologies are biased by the scenarios evaluated. Thus, it can be understood that the selection of genotypes by these two methods mentioned, when using data from scenarios II, III, IV, and V in estimating the scores of PC_1 and PC_2 , tends to be biased. This can cause problems in determining the most promising genotypes for environments in general or specific. A thorough evaluation of the analysis should be performed, especially for VCU tests, since the breeder usually indicates only one or a few cultivars, so it is necessary to analyze the changes caused by the unbalance of the data.

4. CONCLUSION

The Annicchiarico and Lin and Binns methodologies present the highest correspondence in the classification of genotypes with data with violation of the assumption of normality of errors and balanced and data with the normality of errors and unbalanced by 10%, about data with the normality of errors and unbalanced by 20% and data with the normality of errors and unbalanced by 30%.

The GGE and AMMI models present high sensitivity in the estimates of the scores of the first two principal components and lower correspondence in the classification of genotypes when the assumption of normality of errors and experimental unbalance of 20 and 30% are violated.

5. REFERENCES

1. Cruz CD, Carneiro PCS, Regazzi AJ. Modelos biométricos aplicados ao melhoramento genético. Viçosa (MG): Editora UFV; 2014.
2. Taiz L, Zeiger E, Moller IM. Fisiologia e desenvolvimento vegetal. Porto Alegre (RS): Artmed Editora; 2017.

3. Carvalho IR, Szareski VJ, Mambrin RB, et al. Biometric models and maize genetic breeding: A review. *Aust J Crop Sci.* 2018;12(11):1796-805. doi: 10.21475/ajcs.18.12.11.p792
4. Pino FA. A questão da não normalidade: Uma revisão. *Rev Econom Agríc.* 2014;61(2):17-33.
5. Schielzeth H, Dingemanse NJ, Nakagawa S, et al. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods Ecol Evol.* 2020;11(9):1141-52. doi: 10.1111/2041-210X.13434
6. Schmidt P, Hartung J, Rath J, et al. Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. *Crop Sci.* 2019;59(2):525-36. doi: 10.2135/cropsci2018.06.0376
7. Rodrigues FC, Silva FCS, Carneiro PCS, et al. Environmental stratification in trials of unbalanced multiyear soybean (*Glycine max* (L.) Merrill) via the integration of GGE Biplot graphs and networks of environmental similarity. *Euphytica.* 2022;218(6):71. doi: 10.21203/rs.3.rs-1027523/v1
8. Port ED, Carvalho IR, Loro MV, et al. Experimental unbalance: reflections on genetic parameters and corn selection gain. *Rev DELOS.* 2024;17(58):e1643. doi: 10.55905/rdelosv17.n58-015
9. R Core Team. R: A language and environment for statistical computing. Vienna (AT): R Foundation for Statistical Computing; 2025.
10. Wricke G. Zur berechnung der okovalenz bei sommerweizen und hafer. 1965;52:127-38.
11. Annicchiarico P. Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. *J Genetics & Breeding.* 1992;46:269-78.
12. Eberhart SA, Russell WA. Stability parameters for comparing varieties 1. *Crop Sci.* 1966;6:36-40.
13. Lin CS, Binns MR. A superiority measure of cultivar performance for cultivar $\times$ location data. *Can J Plant Sci.* 1988;68(1):193-8.
14. Gauch HG. Model selection and validation for yield trials with interaction. *Biometrics.* 1988;44(3):705.
15. Yan W, Hunt LA, Sheng Q, et al. Cultivar evaluation and mega-environment investigation based on the GGE Biplot. *Crop Sci.* 2000;40(3):597-605. doi: 10.2135/cropsci2000.403597x
16. Venables WN, Ripley BD. Modern applied statistics with S. New York (US): Springer; 2002.
17. Olivoto T, Lúcio AD. metan: An R package for multi-environment trial analysis. Jarman S, organizador. *Methods Ecol Evol.* 2020;11(6):783-9. doi: 10.1111/2041-210x.13384
18. Landsheer JA, Van Den Wittenboer G. Unbalanced 2 x 2 Factorial designs and the interaction effect: A troublesome combination. *PLoS ONE.* 2015;10(3):e0121412. doi: 10.1371/journal.pone.0121412
19. Woyann LG, Benin G, Storck L, et al. Estimation of missing values affects important aspects of GGE biplot analysis. *Crop Sci.* 2017;57(1):40-52. doi: 10.2135/cropsci2016.02.0100