Ciência Rural, Santa Maria, v.53:5, e20220181, 2023 http://doi.org/10.1590/0103-8478cr20220181

ISSNe 1678-4596
CROP PRODUCTION

Does soybean sample size impact Tukey’s test for non-additivity?

Rafael Rodrigues de Souza[1] Marcos Toebe[2*] Anderson Chuquel Mello[1]
Karina Chertok Bittencourt[2] Iris Cristina Datsch Toebe[3]

[1]Departamento de Fitotecnia, Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.
[2]Departamento de Ciências Agronômicas e Ambientais, Universidade Federal de Santa Maria (UFSM), 98400-000, Frederico Westphalen, RS, Brasil. E-mail: m.toebe@gmail.com. *Corresponding author.
[3]Programa de Pós-graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

ABSTRACT: *This study analyzed the interference of sample size on Tukey’s test for non-additivity and found the sample size to optimize the test for soybean grain yield. Six experiments were conducted in a completely randomized block design with either 20 or 30 cultivars and three repetitions of each treatment. Grain yield was determined per plant, totaling 9,000 sampled plants. Next, sample scenarios up to 100 plants were simulated, estimating F statistic for a degree of freedom of the error in each scenario. After that, the optimal sample size was defined via power models and maximum curvature point. Results showed the number of sampled plants per experimental unit influences the estimates of Tukey’s test for non-additivity. Also, the sampling of 14 to 19 plants per experimental unit allows for maintaining the accuracy of the test.*
Key words: *analysis of variance, experimental planning, Glycine max, mathematical assumptions.*

O tamanho de amostra em soja impacta o teste de não aditividade de Tukey?

RESUMO: *Os objetivos deste estudo foram analisar a interferência do tamanho amostral no teste de não aditividade de Tukey e encontrar o tamanho de amostra para otimizar o teste para a produtividade de grãos em soja. Seis experimentos em delineamento de blocos ao acaso foram conduzidos com 20 ou 30 cultivares de soja em três repetições de cada tratamento. A produtividade de grãos foi definida por planta, totalizando 9.000 plantas amostradas. A seguir, foram simulados cenários amostrais de até 100 plantas, estimando a estatística F para um grau de liberdade do erro em cada cenário. Após, foi definido o tamanho amostral ótimo via modelos de potência e pontos de máxima curvatura. Os resultados mostram que o número de plantas amostradas por unidade experimental influencia as estimativas do teste de não aditividade de Tukey. Além disso, a amostragem de 14 a 19 plantas por unidade experimental possibilita manter a acurácia do teste.*
Palavras-chave: *análise de variância, Glycine max, planejamento experimental, pressuposições matemáticas.*

Early research on the analysis of variance, including studies seen as classic references in the field, bring numerous discussions regarding alternative ways of verifying whether mathematical assumptions are met or violated (BARTLETT, 1947; COCHRAN, 1947; EISENHART, 1947), which is still the subject of more recent reports (WELHAM et al., 2015; BUTLER, 2021). TUKEY (1949) highlighted an assumption that is often forgotten in scientific research, such as the additivity of the mathematical model, proposing a methodology to measure it, named “Tukey’s test for non-additivity”. This tool has the purpose of separating a degree of freedom from the experimental error, which consisted of the interaction between the main factors of the analysis of variance, with subsequent application of the F test to identify the presence or absence of additivity in the mathematical model (TUKEY, 1949; BUTLER, 2021). The premise of this methodology is that dilation of a degree of freedom of the error is only possible when rows and columns, that is, the main effects, are not additive (TUKEY, 1949). Therefore, the assertiveness of the analysis of variance tends to be compromised, requiring either transformation or the use of non-Gaussian methodologies (BUTLER, 2021).

Exceptionally, this method is the one that gained greater visibility in identifying the additivity of analysis-of-variance models (ŠIMEČEK & ŠIMEČKOVA, 2013). However, the factors that possibly interfere with its estimates have not been deeply investigated yet. Soybean studies that use the

Received 03.29.22 Approved 06.09.22 Returned by the author 07.19.22
CR-2022-0181.R1
Editors: Leandro Souza da Silva Alessandro Dal’Col Lúcio

analysis of variance (SOUZA et al., 2021; SODRÉ FILHO et al., 2022) use different sample sizes per experimental unit, with variations from 5 to 20 sampled plants. Nevertheless, SOUZA et al. (2022) showed a certain variation of the *F* statistic as a function of sample size for soybean, and such statistic is used in the methodology by TUKEY (1949). Thus, the number of sampled plants per experimental unit could be a factor affecting the bias of estimates from Tukey's test for non-additivity. On this basis, this study analyzed the inference of sample size in Tukey's test for non-additivity and found the sample size to optimize the test for soybean grain yield.

During the 2017/2018 growing season, six experiments with soybean were carried out. Three of them were performed on-farm in the municipality of Erval Seco (27º31'60" S latitude, 53º28'11" W longitude, and 517 m altitude), which presents a soil classified as Dystrophic Red Latosol. The other three experiments were conducted in the experimental area of the Federal University of Pampa – Itaqui Campus (29º09'21" S latitude, 56º33'02" W longitude, and 74 m altitude), located in the municipality of Itaqui, with a soil of the Haplic Plinthosol type. Both locations are in the state of Rio Grande do Sul, Brazil. Sowings in Erval Seco were performed on 10/24/2017 (E1), 11/15/2017 (E2), and 12/05/2017 (E3), and in Itaqui, on 11/02/2017 (E4), 11/30/2017 (E5), and 12/21/2017 (E6). The climate in these locations is characterized as humid subtropical, with no dry season defined (WREGE et al., 2012).

Each experiment was conducted in a randomized complete block design, with three repetitions of each treatment (genotype), and thus three blocks per experiment, which contained one repetition of all treatments. The experimental units consisted of 5 rows 3.0 m long, spaced 0.45 m apart, and a population of 30 plants per m^2 was considered. From a useful area of 2.70 m^2, in each experimental unit, 20 plants were collected after 95% of the plot had reached physiological maturity. Thus, a total of 9,000 plants were measured individually. In each harvested plant, grain yield was determined by weighing grains, with a posterior correction to 13% humidity. Thirty commercial cultivars were assessed in E1, E2, and E3, and 20 cultivars in E4, E5, and E6. All cultivars are of the indeterminate-growth type with a relative maturity group ranging from ≥ 5.0 to ≤ 6.9. The cultivars used in experiments E4, E5, and E6 were '50I52 RSF IPRO', '54I52 RSF IPRO', '5855 RSF IPRO', '58I60 RSF', '5958 RSF IPRO', '59I60 RSF IPRO', '61I59 RSF IPRO', '63I64 RSF IPRO', '6563 RSF IPRO', '68I70 RSF IPRO', '6968 RSF', '7166 RSF IPRO', 'Don Mario 5.9 I', 'NA 5909 RG', 'NS 5959 IPRO', 'NS 6535 IPRO', 'M 5730 IPRO', 'M 5838 IPRO', 'M 5947 IPRO', and 'M 6410 IPRO'. As for experiments E1, E2 e E3, besides the 20 cultivars above, cultivars '53I54 RSF IPRO', '95R51', '95Y52', '96Y90', 'AS 3570IPRO', 'AS 3590IPRO', 'BMX Potência RR', 'BRS6203 RR', 'M5892 IPRO', and 'TMG7062 IPRO' were added.

The statistical analyses were performed through specific routines built in the R environment (R DEVELOPMENT CORE TEAM, 2022). First, data were stratified per experimental unit in all experiments. Next, 31 sampling scenarios were planned (*n* = 1, 2, …, 20, 25, …, 50, 60, ..., 100 plants per experimental unit), so that, in each scenario, resamplings with reposition (bootstrap) were simulated 10,000 times (EFRON, 1979). This procedure was adopted using the *sample*() function. After that, for each resampling, a multiple linear regression was applied using the *lm*() function, considering grain yield per plant as a dependent variable and the effect of genotypes and blocks as independent variables. Each model was then squared, obtaining the square of the effect of the genotype × block interaction ($\lambda_{Gi\beta r}$). Such an operation makes it possible to isolate a degree of freedom of the experimental error, as highlighted by TUKEY (1949). For the verification of the analysis-of-variance model additivity, the $\lambda_{Gi\beta r}$ parameter was added to it and the analysis was performed using the *aov*() function. The following mathematical model was used: $Y_{ir} = m + G_i + \beta_r + \varepsilon_{ir} + \lambda_{Gi\beta r}$, where Y_{ir} is the value observed in the response variable in plot *ir*, *m* is the overall mean, G_i is the fixed effect of level *i* of the genotype factor, being *i* = 1, 2, ..., 30 for E1, E2 and E3 and *i* = 1, 2, ..., 20 for E4, E5 and E6, β_r is the random effect of level *r* (*r* = 1, 2, 3) of the block, ε_{ir} is the effect of the experimental error, and $\lambda_{Gi\beta r}$ previously described. Afterwards, the *F* statistic value of the $\lambda_{Gi\beta r}$ parameter with one degree of freedom was extracted. This statistic was calculated 1,860,000 times (31 sample sizes per experimental unit × 10,000 resamplings × 6 reference experiments).

The values extracted in each sampling scenario were subjected to descriptive analysis, calculating minimum, 2.5 percentiles, mean, 97.5 percentiles, and maximum values. The ninety-five percent confidence interval width ($\text{CI}_{95\%}$) was obtained as the difference between the 97.5 and 2.5 percentiles. Then, $\text{CI}_{95\%}$ estimates and the number of plants per experimental unit (planned sampling scenarios) were fitted through the *nls*() function with the following power model: $CI_{95\%} = a \times n^{\beta} + \varepsilon$, where α is the coefficient of interception, *n* is the sample size, β

is the exponential rate of decay, and ε is the error of random effect. In order to verify the fitting quality of the power model, the following quality indicators were used: coefficient of determination (R^2), root mean square error (RMSE), and Willmott's agreement index (d). Finally, such a model was considered in each experiment to apply four maximum-curvature-point methods (general, perpendicular distances, linear plateau response, and spline), described by SILVA & LIMA (2017), using the *maxcurv*() function from the soilphysics package, which were used to estimate the optimal sample size for Tukey's test for non-additivity.

The number of sampled plants per experimental unit interfered directly with the estimates of Tukey's methodology for non-additivity (Figure 1). This result brings insights into a poorly documented response of this test, that is, to sample size, which shows the expansion of a degree of freedom of the experimental error, as proposed by TUKEY (1949), is influenced by the number of samples used. Thus, a higher tendency of overestimating *F* results is observed in small sampling scenarios, such as when ≤ 3 plants are sampled. This estimate bias remains until the sampling of ≤ 8 plants, gradually reducing $CI_{95\%}$ and, consequently, providing more reliable estimates. SOUZA et al. (2022) also observed an exponential decreasing response when analyzing the response of $CI_{95\%}$ to the *F* test applied on the effect of genotypes, and other studies have also shown similar results for different statistics, such as in TOEBE et al. (2018) and BITTENCOURT et al.

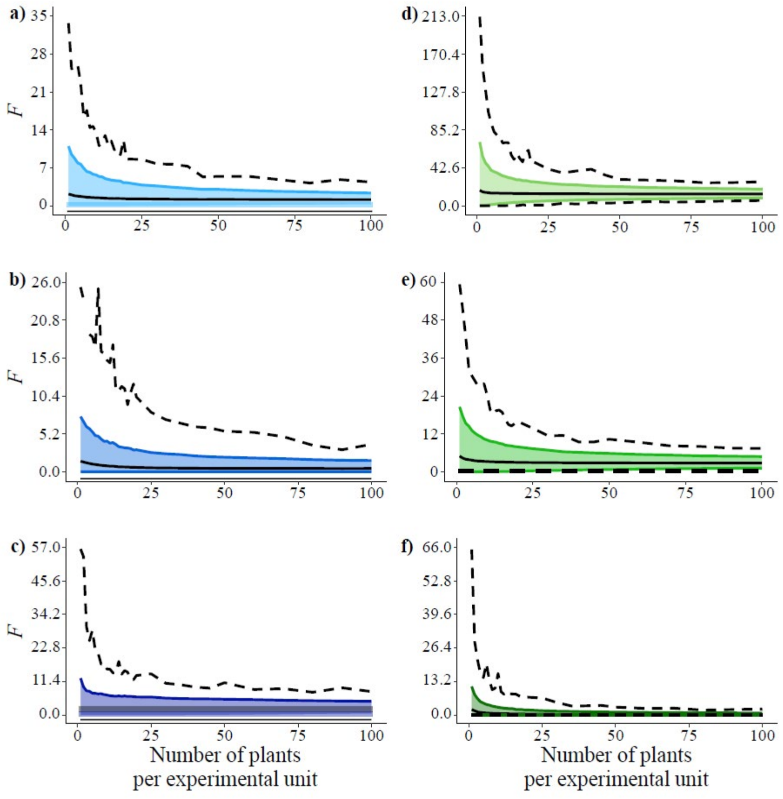

Figure 1 – Minimum, 2.5 percentiles, mean, 97.5 percentiles, and maximum values for the first sowing date (E1 – a), second sowing date (E2 – b), and third sowing date (E3 – c) in Erval Seco, and first sowing date (E4 – d), second sowing date (E5 – e), and third sowing date (E6 – f) in Itaqui for *F* statistic of a degree of freedom of the experimental error, according to the methodology described by Tukey (1949) in soybean.

(2022). In addition, the mean property of the *F* test for the model additivity is not constant, so smaller sample sizes show slightly higher *F* values that stabilize once the sample size is increased. Hence, this is an indicator that the precision of Tukey's test for non-additivity is improved in scenarios of greater samples, and thus, the test sensitivity to sample size.

In this sense, sample size determination was performed reliably, once the six parametrized power models showed a satisfactory fit (Table 1),

Table 1 – Coefficient of determination (R^2), root mean square error (RMSE), and d index of the power models, and maximum curvature points and sample sizes for Tukey's test for non-additivity.

Trial[†]	Power model	R^2	RMSE	d index
E1	$CI_{95\%} = 12.0224 \times n^{-0.3433}$	0.98	0.33	0.99
E2	$CI_{95\%} = 8.5780 \times n^{-0.3392}$	0.96	0.31	0.99
E3	$CI_{95\%} = 11.2604 \times n^{-0.2088}$	0.95	0.35	0.99
E4	$CI_{95\%} = 73.6162 \times n^{-0.4034}$	0.99	0.92	0.99
E5	$CI_{95\%} = 22.1560 \times n^{-0.3540}$	0.99	0.51	0.99
E6	$CI_{95\%} = 11.5640 \times n^{-0.5749}$	0.99	0.12	0.99
Trial	Maximum curvature method	Maximum Curvature	Maximum $CI_{95\%}$	Sample size
E1	Geral method	2.55	8.72	3
E1	Spline method	12.17	5.10	13
E1	Perpendicular distance method	16.39	4.60	17
E1	Linear plateau response method	26.84	3.89	27
E2	Geral method	1.97	6.82	2
E2	Spline method	12.23	3.67	13
E2	Perpendicular distance method	16.45	3.32	17
E2	Linear plateau response method	27.07	2.80	28
E3	Geral method	1.69	10.09	2
E3	Spline method	14.33	6.46	15
E3	Perpendicular distance method	18.25	6.14	19
E3	Linear plateau response method	35.04	5.36	36
E4	Geral method	10.13	28.93	11
E4	Spline method	11.28	27.70	12
E4	Perpendicular distance method	15.62	24.29	16
E4	Linear plateau response method	23.56	20.58	24
E5	Geral method	4.10	13.48	5
E5	Spline method	11.99	9.19	12
E5	Perpendicular distance method	16.25	8.26	17
E5	Linear plateau response method	26.23	6.97	27
E6	Geral method	3.14	5.99	4
E6	Spline method	9.06	3.26	10
E6	Perpendicular distance method	13.64	2.58	14
E6	Linear plateau response method	15.97	2.35	16

[†] E1: first sowing date (October 24, 2017), E2: second sowing date (November 15, 2017), and E3: third sowing date (December 05, 2017) in Erval Seco–RS; E4: first sowing date (November 02, 2017), E5: second sowing date (November 30, 2017), and E6: third sowing date (December 21, 2017) in Itaqui–RS.

based on quality indicators. R^2 values were ≥ 0.95 and d ≥ 0.99, and RMSE did not exceed 0.92, where the least efficient model was the one parametrized for E4, although it still reached a high precision (WILLMOTT et al., 2012). The four methods for defining the maximum curvature points of each model obtained quite different sample sizes, with recommendations varying from ≥ 2 to ≤ 36 plants per experimental unit. As observed, such an oscillation in sample dimensioning depends on the technique used. BITTENCOURT et al. (2022), when using the same four maximum-curvature-point methods, also noted different results. The same authors verified smaller sample sizes were obtained through the general and spline methods than through the perpendicular distances and linear plateau response, which also occurred in this study (Figure 2). The general method, for instance, presented sample size values varying from 2 to 5 plants per experimental unit between experiments, while the linear plateau response method suggested 16 to 36 plants. Based on the $CI_{95\%}$, the values recommended by the general method (≤ 5 plants) may lead to biased results, once $CI_{95\%}$ has not reached stabilization yet with those sample sizes. The same situation is valid for the spline method (≤ 15 plants), especially as shown in figures 2a, 2b, and 2d.

The perpendicular distances and linear plateau response methods show greater sample size results, which are closer to $CI_{95\%}$ stabilization point, and thus, more efficient in the dimensioning

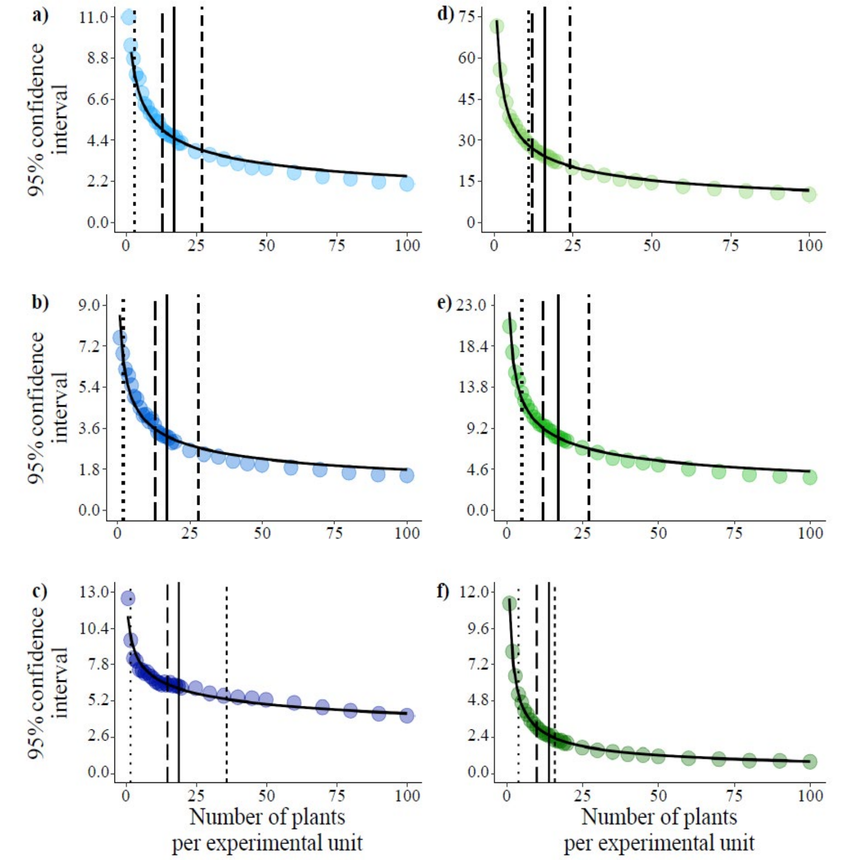

Figure 2 – Sample size determination via power model and maximum curvature points for Tukey's test for non-additivity for the first sowing date (E1 – a), second sowing date (E2 – b), and third sowing date (E3 – c) in Erval Seco, and first sowing date (E4 – d), second sowing date (E5 – e), and third sowing date (E6 – f) in Itaqui.

of sample size in this case. Importantly, although the recommendations obtained through the linear plateau response method reached 36 plants per experimental unit, the number of plants defined through the perpendicular distances' method is ≤ 19. Thus, the linear plateau response not only resulted in a considerably larger number of plants than the perpendicular distances' but also very little precision is gained if compared to the latter. BITTENCOURT et al. (2022) observed the same situation when defining sample size for the overall experimental mean in cauliflower seedlings, and SOUZA et al. (2022) used the perpendicular distances' method to estimate sample size for precision statistics in soybean. Such studies reinforce the use of the perpendicular distances' method, and for this, we recommend the sampling of ≥ 14 to ≤ 19 plants per experimental unit in order to optimize the estimates of Tukey's test for non-additivity, which will enable the accurate verification of the additivity assumption in analysis-of-variance models for soybean grain yield. However, the recommendations here made should not be followed without performing preliminary studies in experiments carried out in extremely different conditions than the ones here described, and should merely serve as a starting point for researchers that measure different traits in soybean.

ACKNOWLEDGEMENTS

We thank the scholarship students and volunteers for their help in experimental conduction and data acquisition; Eluizio Ferrari Manfio and Érico Luis for making the area available for the experiment in Erval Seco – RS. To the Universidade Federal de Santa Maria (UFSM), Universidade Federal do Pampa (UNIPAMPA), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Programa de educação tutorial (PET), FAPERGS / Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Process number 16/2551-0000257-6 ARD/PPP) and to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil - Finance code 001 for scholarships and financial support.

DECLARATION OF CONFLICT OF INTERESTS

We have no conflict of interest to declare.

AUTHORS' CONTRIBUTIONS

Conceptualization: RRS. Data acquisition: RRS and ACM Design of methodology and data analysis: RRS and MT. Supervision and coordination: MT and ICDT. RRS and KCB prepared the draft of the manuscript. All authors critically revised the manuscript and approved of the final version.

REFERENCES

BARTLETT, M.S. The use of transformations. **Biometrics**, v.3, p.39–57, 1947. Available from: <https://doi.org/10.2307/3001536>. Accessed: Jan. 22, 2022. doi: 10.2307/3001536.

BITTENCOURT, K. C. et al. What is the best way to define sample size for cauliflower seedlings? **Ciência Rural**, v.52, e20210747, 2022. doi: 10.1590/0103-8478cr20210747.

BUTLER, R.C. Popularity leads to bad habits: Alternatives to "the statistics" routine of significance, "alphabet soup" and dynamite plots. **Annals of Applied Biology**, v.180, p.1–14, 2021. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/aab.12734>. Accessed: Feb. 25, 2022. doi: 10.1111/aab.12734.

COCHRAN, W.G. Some consequences when the assumptions for the analysis of variance are not satisfied. **Biometrics**, v.3, p.22–38, 1947. Available from: <https://doi.org/10.2307/3001535>. Accessed: Jan. 22, 2022. doi: 10.2307/3001535.

EFRON, B. Bootstrap methods: another look at the jackknife. **Annals of Statistic**, v.7, p.1–26, 1979. Available from: <https://doi.org/10.1214/aos/1176344552>. Accessed: Feb. 01, 2022. doi: 10.1214/aos/1176344552.

EISENHART, C. The assumptions underlying the analysis of variance. **Biometrics**, v.3, p.1–21, 1947. Available from: <https://doi.org/10.2307/3001534>. Accessed: Jan. 22, 2022. doi: 10.2307/3001534.

R DEVELOPMENT CORE TEAM. **R**: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2022.

SILVA, A.R. da; LIMA, R.P. Determination of maximum curvature point with the R package soilphysics. **International Journal of Current Research**, v.9, p.45241–45245, 2017. Available from: <https://www.journalcra.com/sites/default/files/issue-pdf/20162.pdf>. Accessed: Jan. 28, 2022.

ŠIMEČEK, P.; ŠIMEČKOVA, M. Modification of Tukey's additivity test. **Journal of Statistical Planning and Inference**, v.143, p.197–201, 2013. Available from: <https://doi.org/10.1016/j.jspi.2012.07.002>. Accessed: Mar. 03, 2022. doi: 10.1016/j.jspi.2012.07.002.

SODRÉ FILHO, J. et al. Intercropping sorghum and grasses during off-season in Brazilian Cerrado. **Scientia Agricola**, v.79, e20200284, 2022. Available from: <https://www.scielo.br/j/sa/a/hZdytDZ7FtrCsZhVYSndhdQ/?lang=en>. Accessed: Feb. 02, 2022. doi: 10.1590/1678-992X-2020-0284.

SOUZA, R.R. de. et al. Soybean yield variability per plant in subtropical climate: sample size definition and prediction models for precision statistics. **European Journal of Agronomy**, v.136, 126489, 2022. Available from: <https://doi.org/10.1016/j.eja.2022.126489>. Accessed: Mar. 15, 2022. doi: 10.1016/j.eja.2022.126489.

SOUZA, R.R. de. et al. Soybean grain yield in highland and lowland cultivation systems: A genotype by environment interaction approach. **Annals of Applied Biology**, v.179, p.302–318, 2021. Available from: <https://onlinelibrary.wiley.

com/doi/abs/10.1111/aab.12709>. Accessed: Feb. 01, 2022. doi: 10.1111/aab.12709.

TOEBE, M. et al. Sample size for estimating mean and coefficient of variation in species of crotalarias. **Anais da Academia Brasileira de Ciências**, v.90, p.1705–1715, 2018. Available from: <https://doi.org/10.1590/0001-3765201820170813>. Accessed: Jan. 28, 2022. doi: 10.1590/0001-3765201820170813.

TUKEY, J.W. One degree of freedom for non-additivity. **Biometrics**, v.5, p.232–242, 1949. Available from: <https://doi.org/10.2307/3001938>. Accessed: Jan. 22, 2022. doi: 10.2307/3001938.

WELHAM, S.J. et al. **Statistical methods in biology**: **Design and analysis of experiments and regression**. Boca Raton: CRC Press, 2015. 608p.

WILLMOTT, C.J. et al. A refined index of model performance. **International Journal of Climatology**, v.32, p.2088–2094, 2012. Available from: <https://doi.org/10.1002/joc.2419>. Accessed: Fev. 27, 2022. doi: 10.1002/joc.2419.

WREGE, M.S. et al. **Climatic Atlas of the South Region of Brazil**: **States of Paraná, Santa Catarina and Rio Grande do Sul**, Brasília: EMBRAPA, 2012. 334p.