Ciência Rural, Santa Maria, v.53:6, e20220110, 2023 http://doi.org/10.1590/0103-8478cr20220110

ISSNe 1678-4596
CROP PRODUCTION

Optimizing Bartlett test: a grain yield analysis in soybean

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 response of the Bartlett test as a function of sample size and to define the optimal sample size for the test with soybean grain yield data. Six experiments were conducted in a randomized block design with 20 or 30 cultivars and three repetitions. Grain yield was determined per plant, totaling 9,000 sampled plants. Next, sample scenarios of 1, 2, ..., 100 plants were simulated and the optimal sample size was defined via maximum curvature points. The increase in sampled plants per experimental unit favors Bartlett test's precision. Also, the sampling of 17 to 20 plants per experimental unit is enough to maintain the accuracy of the test.
Key words: analysis of variance, experimental planning, Glycine max, mathematical assumptions.

Otimizando o teste de Bartlett: uma análise da produtividade de grãos em soja

RESUMO: Os objetivos deste estudo foram analisar a resposta do teste de Bartlett em função do tamanho de amostra e definir o tamanho amostral ótimo para o teste com dados de produtividade de grãos de soja. Foram conduzidos seis experimentos em delineamento de blocos ao acaso com 20 ou 30 cultivares e três repetições. A produtividade de grãos foi definida por planta, totalizando 9.000 plantas amostradas. Logo, foram simulados cenários amostrais de 1, 2, ..., 100 plantas e definido o tamanho amostral ótimo via pontos de máxima curvatura. O aumento de plantas amostradas por unidade experimental favorece a precisão do teste de Bartlett. Além disso, a amostragem de 17 a 20 plantas por unidade experimental é suficiente para manter a acurácia do teste.
Palavras-chave: análise de variância, Glycine max, planejamento experimental, pressuposições matemáticas.

Gaussian inferences are subject to mathematical assumptions that, if violated, may reduce the reliability of results (WELHAM et al., 2015; BUTLER, 2021). The analysis of variance, in particular, which is used for summarizing scientific data, is subject to four assumptions, such as the additivity of the model, error independence, error normality, and homogeneity of variances (BUTLER, 2021). The two latter are normally the hardest ones to meet and; although BLANCA et al. (2017) pointed out that the analysis of variance is robust to normality deviations, such robustness does not include cases with heterogeneous variances (WELHAM et al., 2015). This is because FISHER (1925), when developing such analysis, considered the variances of each treatment to be similar or at least close. If the variation surrounding the mean of each treatment is similar, a grouped error can be calculated (BUTLER, 2021); otherwise, this inference loses reliability.

Many statistical tests can be used in order to evaluate the presence of variance homoscedasticity, being Bartlett test one of the most common (BARTLETT, 1937). However, in cases where variance homoscedasticity is violated, the accuracy of the test used to assess the homogeneity of variances is an important factor to verify. Bartlett test itself is susceptible to normality deviations (BARTLETT, 1937; WELHAM et al., 2015); however, this may not be the only factor that interferes with its estimates. Little is known about the quantitative response of this test as a function of sample size, being samplings often empirically performed for soybean yield traits,

Received 02.27.22 Approved 07.03.22 Returned by the author 08.09.22
CR-2022-0110.R1
Editors: Leandro Souza da Silva Alessandro Dal'Col Lucio

as in SOUZA et al. (2021) and SODRÉ FILHO et al. (2022), who evaluated 20 and 5 plants per experimental unit, respectively. Therefore, in order to optimize the accuracy of the test and identify how sample size interferes with Bartlett's estimates, this study analyzed the response of the Bartlett test as a function of sample size and defined the optimal sample size for soybean grain yield data.

Six experiments with soybean were carried out during the 2017/2018 growing season. Three of them were performed on a farm in the municipality of Erval Seco (27º31'60"S latitude, 53º28'11"W longitude, and 517 m altitude), which were sown on 10/24/2017 (E1), 11/15/2017 (E2), and 12/05/2017 (E3), and the other three experiments were performed 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, which were sown on 11/02/2017 (E4), 11/30/2017 (E5), and 12/21/2017 (E6). Both locations are in the state of Rio Grande do Sul, Brazil, and the climate in both is characterized as humid subtropical, with no dry season defined (WREGE et al., 2012), and soils classified as Dystrophic Red Latosol and Haplic Plinthosol (SANTOS et al., 2018) in Erval Seco and Itaqui, respectively.

In the experiments, a randomized block design was used, with three repetitions. A population of 30 plants per m^2 was set, and each experimental unit consisted of 5 rows 3.0 m long, spaced 0.45 m away, considering as a useful area 2.70 m^2. Within the useful area, 20 plants were collected per experimental unit, after 95% of the plot had reached the stage of physiological maturity, thus 9,000 plants were evaluated in total. In each harvested plant, grain yield was determined through grain weighing, with a posterior correction to 13% moisture. Thirty commercial cultivars were assessed in E1, E2, and E3, and 20 cultivars in E4, E5, and E6. 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. All cultivars are indeterminate growth types with a relative maturity group ranging from ≥5.0 to ≤6.9. All cultural practices were performed following standard recommendations for the crop (SALVADORI et al., 2016).

For the data analysis, specific routines constructed in R software were used (R DEVELOPMENT CORE TEAM, 2022). Initially, the database was subdivided per experimental unit for all experiments (E1, E2, E3, E4, E5, and E6). Next, 31 sampling scenarios of n = 1, 2, …, 20, 25, …, 50, 60, ..., 100) plants per experimental unit were simulated with reposition and 10,000 resamplings (EFRON, 1979) for each experiment, using *sample*() function. Once the values of each experimental unit in the re-samplings per sampling scenario were obtained, the analysis of variance was performed with *aov*() function, according to the following mathematical model: $Y_{ir} = m + G_i + \beta_r + \varepsilon_{ir}$, 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 and ε_{ir} is the effect of the experimental error. The estimates of the error ($\hat{\varepsilon}_{ir}$) obtained by $\hat{\varepsilon}_{ir} = Y_{ir} - (\hat{m} + \hat{G}_i + \hat{\beta}_r)$ were extracted and the Bartlett test was applied at 5% error probability using *bartlett.test*() function. Bartlett's statistic (K^2) was obtained 1,860,000 times (31 sample sizes per experimental unit × 10,000 re-samplings × 6 reference experiments).

Finally, each planned scenario was subject to a descriptive analysis calculating minimum, 2.5 percentiles, mean, 97.5 percentiles, and maximum values. The ninety five percent confidence interval width ($CI_{95\%}$) was obtained as the difference between the 97.5 and 2.5 percentiles. Then, $CI_{95\%}$ estimates were fitted through *nls*() function with the following power model: $CI_{95\%} = \alpha \times n^{\beta} + \varepsilon$, 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. Subsequently, four maximum curvature point methods were used (general, perpendicular distances, linear plateau response, and spline) as described by SILVA & LIMA (2017), using the *maxcurv*() function from the soilphysics package, considering the point reached as a sample size that is representative enough.

As expected, sample size directly interferes with Bartlett test's estimates (Figure 1) when analyzing soybean grain yield per plant. By observing the mean properties of the six trials, an exponential decreasing response is identified, which is also true for the $CI_{95\%}$. This type of response has already been described for other statistics when analyzing $CI_{95\%}$ (TOEBE et al., 2018; PIÑERA-CHAVEZ et al., 2020). Such indicators showed that increasing sample size guarantees a higher precision to the test's estimates (TOEBE et al., 2018).

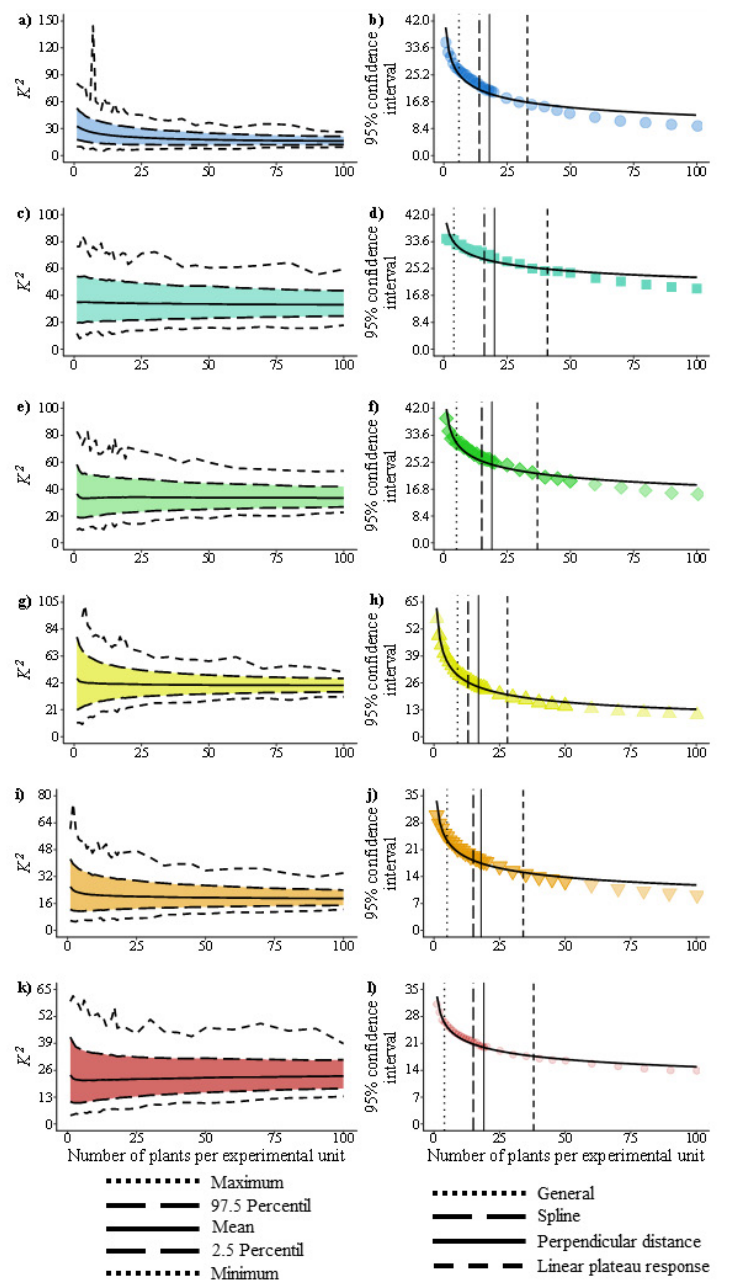

Figure 1 - Descriptive statistics (minimum, 2.5 percentiles, mean, 97.5 percentiles, and maximum values) for trials E1 (a), E2 (c), E3 (e), E4 (g), E5 (i), and E6 (k), and power models for the comparison of four methods for determining the maximum curvature point (general, spline, perpendicular distance, and linear plateau response methods) to estimate Bartlett's K^2 reliably in E1 (b), E2 (d), E3 (f), E4 (h), E5 (j), and E6 (l).

Bartlett test's sensitivity to sample size is identified in small sampling scenarios, as in a number of ≤ 5 plants per experimental unit. In those cases, there is a higher tendency to overestimate the values of the test. However, as observed in figures 1a, 1c, 1e, 1g, 1i, and 1k, an underestimation bias is also possible.

Moreover, four methods to estimate sample size were applied, and compared by the previous fitting of power models (Table 1 and figure 1). The power models showed a satisfactory performance in the six trials, when analyzed using fitting indicators as the coefficient of determination (R^2), root mean square error (RMSE), and Willmott's agreement index (d). This allows to make inferences *a posteriori*, such as the use of maximum curvature points, to be efficiently made (SILVA & LIMA, 2017). Nevertheless, contrasting sample size values were identified, ranging from ≥ 4 to ≤ 41 plants per experimental unit. Perceptibly, such a large variation occurs due to the implemented method since only slight differences can be seen when comparing sample sizes obtained through the same method between trials. An example of this is, when comparing the optimal sample size for the Bartlett test between trials, obtained using the general method, the number of plants only fluctuates from ≥ 4 to ≤ 9 plants per experimental unit. Equally, with the linear plateau response method, variation is little, ranging from ≥ 28 to ≤ 41 plants per experimental unit. The same is observed for the perpendicular distance and spline methods.

Based on the $CI_{95\%}$, small sample sizes, as the ones obtained through the general method (≤ 9 plants) may lead to biased estimates; and although the slightly greater sizes suggested by the spline method (≤ 15 plants) might reduce the bias of the test, such values are still far from optimizing it, that is, $CI_{95\%}$ is still decreasing, meaning the curve has not stabilized yet at

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 Bartlett's test.

Trials	Power model	$R^{2(£)}$	RMSE	d index
E1	$CI_{95\%} = 40.0805 \times n^{-0.2509}$	0.92	1.89	0.98
E2	$CI_{95\%} = 39.5148 \times n^{-0.1229}$	0.84	1.81	0.95
E3	$CI_{95\%} = 41.8149 \times n^{-0.1817}$	0.93	1.52	0.98
E4	$CI_{95\%} = 62.2628 \times n^{-0.3365}$	0.98	1.84	0.99
E5	$CI_{95\%} = 33.7467 \times n^{-0.2281}$	0.94	1.38	0.98
E6	$CI_{95\%} = 33.2686 \times n^{-0.1742}$	0.98	0.69	0.99
Trials	Maximum curvature methods	Maximum Curvature	Maximum $CI_{95\%}$	Sample size
E1	Geral method	5.38	26.28	6
E1	Spline method	13.62	20.81	14
E1	Perpendicular distance method	17.64	19.51	18
E1	Linear plateau response method	32.35	16.75	33
E2	Geral method	3.22	34.23	4
E2	Spline method	15.84	28.14	16
E2	Perpendicular distance method	19.54	27.42	20
E2	Linear plateau response method	40.78	25.05	41
E3	Geral method	4.56	31.73	5
E3	Spline method	14.80	25.63	15
E3	Perpendicular distance method	18.65	24.57	19
E3	Linear plateau response method	36.82	21.72	37
E4	Geral method	8.60	30.18	9
E4	Spline method	12.27	26.78	13
E4	Perpendicular distance method	16.47	24.26	17
E4	Linear plateau response method	27.22	20.48	28
E5	Geral method	4.43	24.04	5
E5	Spline method	14.01	18.48	15
E5	Perpendicular distance method	17.97	17.46	18
E5	Linear plateau response method	33.79	15.12	34
E6	Geral method	3.63	26.57	4
E6	Spline method	14.93	20.77	15
E6	Perpendicular distance method	18.76	19.96	19
E6	Linear plateau response method	37.32	17.71	38

[†] 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.

those points. Only up from the sample numbers obtained through the perpendicular distance and linear plateau response methods, is $CI_{95\%}$ curve beginning to stabilize, which suggested that the values reached with those methods are representative enough sample sizes. Interestingly; although the perpendicular distance method recommended, at maximum, the sampling of 20 plants per experimental unit, and the linear plateau response reached a maximum of 41 plants. When analyzing $CI_{95\%}$, the precision gain obtained with the linear plateau response method is too little compared with the perpendicular distances', not being enough to justify the choice of the first over the latter. That way; although both methods are capable of obtaining sufficiently reliable sample size estimates to optimize the Bartlett test, we encourage the sampling of ≥ 17 to ≤ 20 plants per experimental unit, so that the test's estimates generate accurate results, enabling the verification of the meeting or violation of the homogeneity of variances assumption in an analysis of variance performed for soybean crop.

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 Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – Process number 313827/2021-4), to the FAPERGS/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

The authors declare no conflict of interest. Funding sponsors had no role in the study design, collection, analysis, and data interpretation; during the writing of this manuscript, and in the decision to publish the results.

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. Properties of sufficiency and statistical tests. **Proceedings of the Royal Society of London Series A**, v.160, p.268–282, 1937. Available from: <https://royalsocietypublishing.org/doi/10.1098/rspa.1937.0109>. Accessed: Jan. 29, 2022. doi: 10.1098/rspa.1937.0109.

BLANCA, M. J. et al. Non-normal data: Is ANOVA still a valid option?. **Psicothema**, v.29, p.552–57, 2017. Available from: <http://www.psicothema.com/pdf/4434.pdf>. Accessed: Jan. 29, 2022. doi: 10.7334/psicothema2016.383.

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: Jan 31, 2022. doi: 10.1111/aab.12734.

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: Sep. 18, 2021. doi: 10.1214/aos/1176344552.

FISHER, R. A. **Statistical methods for research workers**. Edinburgh: Oliver, 1925. 378p.

PIÑERA-CHAVEZ, F.J. et al. Optimizing phenotyping methods to evaluate lodging risk for wheat. **Field Crops Research**, v.258, 107933, 2020. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S037842902031217X>. Accessed: Feb. 02, 2022. doi: 10.1016/j.fcr.2020.107933.

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

SALVADORI, J. R. et al. **Technical indications for the soybean culture in Rio Grande do Sul and Santa Catarina, harvests of 2016/2017 and 2017/2018**. Passo Fundo: UPF, 2016. 128p.

SANTOS, H. G. et al. **Brasilian Soil Classification System**. Brasília: EMBRAPA, 2018. 303p.

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.

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. 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.

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

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.