

## Optimum Sample Size in the Germination of Atemoya Seeds (*Annona* × *atemoya* Mabb.)

Rafaela Lanças Gomes<sup>1</sup>, José Raimundo de Souza Passos<sup>2</sup>, Juliana Iassia Gimenez<sup>1</sup>, Marília Caixeta Sousa<sup>1</sup>,  
Mariana de Fatima De-Pieri-Oliveira<sup>1</sup>, Carolina Oville Mimi<sup>1</sup> & Gisela Ferreira<sup>1</sup>

<sup>1</sup> Department of Botany, 'Júlio de Mesquita Filho' São Paulo State University, Botucatu Campus, Botucatu, São Paulo, Brazil

<sup>2</sup> Department of Biostatistics, 'Júlio de Mesquita Filho' São Paulo State University, Botucatu Campus, Botucatu, São Paulo, Brazil

Correspondence: Rafaela Lanças Gomes, Department of Botany, IBB, 'Júlio de Mesquita Filho' São Paulo State University, Botucatu Campus, Rua Prof. Dr. Antonio Celso Wagner Zanin, s/n°, CEP: 18618-689, Botucatu, São Paulo, Brazil. Tel: 55-149-9759-3173. E-mail: r.lancasg@gmail.com

Received: February 11, 2019

Accepted: March 18, 2019

Online Published: May 31, 2019

doi:10.5539/jas.v11n7p239

URL: <https://doi.org/10.5539/jas.v11n7p239>

### Abstract

Atemoya currently has its seeds studied in several aspects, from the technological and physiological point of view. However, for the performance of the germination test, there is no standardization in relation to the number of seeds and replicates. Thus, this work aims to determine the optimal sample size for germination tests with atemoya seeds. A germination test was carried out with 5 treatments, considering 10, 20, 30 40 and 50 seeds per sampling unit with 50 replicates of each. Data were analyzed using logistic regression, non-linear Gompertz regression models, bootstrap simulation and graphs in the form of contour lines in order to be able to infer the best binomial of the number of seeds per sampling unit and the number of replicates. The expected germination percentage, maximum germination speed and times for the beginning of the germination process, maximum germination speed and interruption of the germination process were determined. The treatment of 20 seeds per sampling unit statistically differs from that of 40 seeds by the Tukey-Kramer test with 5% significance. It could be concluded that for germination tests with atemoya seeds to have statistical validity, the smallest sample size is 10 seeds per unit, regardless of number of replicates.

**Keywords:** non-linear gompertz regression, logistic regression, bootstrap, germination time

### 1. Introduction

Annonaceae is a botanical family in which some species economically stand out due to their fruits that have high commercial value and generate interest in both domestic and external market. Of these, atemoya, hybrid of *Annona cherimola* Mill. and *Annona squamosa* L., has large area of cultivation in Brazil, mainly in the Southern and Southeastern regions, already presenting cultivation area of over 1200 hectares (Lemos, 2014).

Its seeds, as well as the seeds of other Annonaceae, have been studied in several aspects, both from the technological point of view, for the production of seedlings and for the understanding of physiological processes, such as dormancy (Ferreira et al., 2016; Carvalho et al., 2018), desiccation tolerance (Corsato et al., 2013) and synthesis of specialized metabolites during germination (De La Cruz & González-Esquinca, 2012). However, for germination tests, there is no standardization in relation to the number of seeds per sampling unit and number of replicates used: Carvalho et al. (2018) used 5 replicates of 25 seeds in *Annona* × *atemoya* (cv Thompson and Gefner), Menegazzo et al. (2012) performed tests with 4 replicates of 20 seeds of *A. squamosa*. A smaller number of seeds were observed in the studies of Oliveira and Andrade (2005), who used only four replicates and 10 seeds of *Annona montana*.

To avoid the researcher to reach wrong conclusions due to sampling deficiency, a representative sampling methodology, based on solid and consistent methods that guarantee the quality of the scientific method, should be used. In germination tests, there are two factors to consider: number of seeds per sampling unit (pseudo-replicates) and number of sampling units to be used (replicates).

The number of replicates to be used will vary according to the study aims because it is closely linked to the

experimental design, and therefore also associated to the residue degrees of freedom. The number of seeds per sampling unit can be defined and fixed using statistical techniques.

In this way, characterizing the minimum number of seeds per sampling unit as well as the minimum number of sampling units is important to ensure the quality of statistical tests, as well as to avoid wasting seeds, time and resources. Thus, this work aims to determine the optimal sample size for germination tests with atemoya seeds (*Annona × atemoya* Mabb.).

## 2. Method

Atemoya fruits (*Annona × atemoya* Mabb.) ‘Thompson’ cultivar were collected in a commercial area in the municipality of Pardinho, São Paulo, in June 2017. Seeds (total of 7500) were collected from fruits, washed in running water, selected, seed surface was dried on a bench and treated by immersion in hypochlorite solution at 5% concentration and fungicide (1% Captan) for 15 minutes. Tetrazolium tests were carried out to evaluate the batch viability at the beginning of the experiment and at the end of the experiment to evaluate if seeds were dormant or dead (Gimenez et al., 2014) and water content (oven method at 105 °C).

### 2.1 Germination Test and Experimental Design

Seeds were soaked in 250 mg L<sup>-1</sup> gibberellic acid (GA<sub>3</sub>) solution for 24 hours with constant aeration, being then separated into 5 treatments with 50 replicates. Treatments consisted of the number of seeds (10, 20, 30, 40 and 50) per sampling unit (paper roll for germination-‘germiteste’ paper).

For the assembly of the experiment, sowing was carried out on paper roll for germination, using three leaves per roll. Leaves were moistened with 2.5 times the dry weight of the paper with deionized water. After sowing on rolls, they were placed in transparent polyethylene bags and randomly disposed in two germinators with alternating photoperiod, 16 hours of light at 30 °C and 8 hours of dark at 20 °C. Seeds with protrusion of at least 2 mm of primary root were considered germinated (Braga et al., 2010). Germinated seed counts were performed every 2 days during the 45-day period.

### 2.2 Data Analysis

#### 2.2.1 Logistic Regression

A generalized linear model with binomial distribution and logit - logistic regression link function (Nelder & Wedderburn, 1972) was used to study the effect of the number of seeds per sampling unit (s.u.) at maximum germination speed. The quality of the adjustment was initially performed by the analysis of deviations by degrees of freedom and later by the standardized Pearson residue graphs (Nelder & Wedderburn, 1972). For comparisons among treatments, the Tukey-Kramer test (Westfall et al., 2000) of the genmod procedure-statistical program SAS-Free Statistical Software, SAS University Edition, was used.

#### 2.2.2 Bootstrap Simulation

The bootstrap simulation technique was used to provide possible combinations of the binomial number of seeds per sampling unit (s.u.)—number of replicates for maximum accumulated germination percentage. For each treatment, 1000 random samples were obtained, with sample size ranging from 2 to 48 sampling units, varying from 2 to 2 units. Random samples were obtained using the survey select procedure of the SAS statistical software-Free Statistical Software, SAS University Edition.

From this sampling process, a graph in the form of contour lines of maximum germination percentages was constructed as a function of the number of seeds per sampling units and number of replicates. The smoothing method used was the spline, which consists of the use of a polynomial to generate a surface to minimize the curvature, which results in a smoothed surface, passing through the sampled points (Wahba, 1990). Thus, operational ranges can be empirically proposed, so that the sample size and the number of replicates can be chosen based on the desired germination percentage.

#### 2.2.3 Non-linear Gompertz Regression

Non-linear Gompertz regression models were adjusted to model the accumulated germination percentage in time, according to the number of seeds per s.u. The adjusted Gompertz model was as follows:

$$y_{ijk} = \alpha_{jk} \cdot \exp\{-\exp[\gamma_{jk}(\beta_{jk} - t_{ijk})]\} + u_{ijk} \quad (1)$$

where,  $y$  is the accumulated germination percentage (%);  $t$  is the time (days);  $\alpha$  is the asymptote (%);  $\beta$  is the time (days) for maximum accumulated germination speed;  $\gamma$  is a parameter related to the accumulated germination speed;  $u$  is the random error with normal distribution with zero mean and constant variance;  $\exp$  is the basis of Neperian logarithms. The  $i, j$  and  $k$  indexes are, respectively, observations in time, number of seeds

per s. u. and the number of sampling units.

After adjustment of models, the following functions of parameters were determined (Passos et al., 2012): maximum speed of crecimiento ( $v_{max} = \alpha\gamma/e$ ); lagtime ( $lagtime = \beta - 1/\gamma$ ): estimated time for the beginning of the germination process. The latter is determined by the interception of the tangent line to the Gompertz curve at the point of inflection with the time axis; interruption time ( $tempo_p = \beta + 1.72/\gamma$ ): time above which the gain in the y-axis variation (accumulated germination speed) is not statistically significant. This time is obtained by intercepting the tangent line to the Gompertz curve at the point of inflection with the asymptote, the straight line ( $y = \text{accumulated germination percentage} = \alpha$ ).

These models applied to the accumulated germination percentage in time for the different treatments allow us understanding the growth dynamics of the accumulated germination percentage over time and verifying the effect of treatment (number of seeds per s.u.) on the expected accumulated germination percentage, on the maximum germination speed and on times for: beginning of the germination process, maximum germination speed, and interruption of the germination process.

For each of the previously mentioned parameter functions, as well as for the estimated asymptote, generalized linear models were adjusted, taking treatments as factors. The quality of adjustments was initially performed by the analysis of the deviation by degrees of freedom and later by the residue graphs (Nelder & Wedderburn, 1972). For comparisons between number of seeds per s.u., the Tukey-Kramer test (Westfall et al., 2000) of the genmod procedure-statistical program SAS-Free Statistical Software, SAS University Edition, was used.

#### 2.2.4 Graphs in the Form of Contour Lines

A graph in the form of contour lines of the accumulated germination percentage according to the binomial time and number of seeds per s.u. was obtained. Thus, it is possible to propose, in an empirical way, regions in which this binomial produces the maximum accumulated germination percentage values and defines the time of experiment interruption as a function of time and number of seeds per s.u. In this graph we used spline smoothing method (Wahba, 1990).

### 3. Results

The water content and viability of seeds were respectively 8.27% and 88%.

The quality of adjustment of all models of this experiment can be considered good since the value of deviations by degrees of freedom is close to unity and the standardized Pearson analysis shows a random residue around zero and floating in its large majority, between -2 and +2, which is in agreement with (Nelder & Wedderburn, 1972).

#### 3.1 Logistic Regression

There was a significant effect of the number of seeds per s.u. (sampling unit) ( $p = 0.0231$ ), that is, there is treatment effect. Table 1 shows the means of the maximum germination speed, followed by the respective standard errors, as well as the result of the Tukey-Kramer test for comparison of means with 5% significance level. The mean T2 value (20 seeds per sampling unit) is statistically different at 5% significance level by the Tukey-Kramer test from the mean T4 value (40 seeds), with germination percentage of 79.9% and 72.7%, respectively. However, the mean germination percentage reached with these treatments (20 and 40 seeds per s.u.) was not different from the others (10, 30 and 50 seeds) at 5% significance level by the Tukey-Kramer test.

Table 1. Means of maximum accumulated germination percentage (%), standard error of the mean (in parentheses) according to the number of seeds per sampling unit for the effect of the number of seeds per sampling unit (s.u.-treatments) at maximum germination speed

Treatment	Number of seeds per s.u.	Means of maximum germination speed
1	10	77.40 <sup>(*)</sup> (2.33) AB
2	20	79.90 (1.72) A
3	30	76.13 (1.20) AB
4	40	72.70 (1.15) B
5	50	78.04 (1.22) AB

Note. (\*) Means followed by the same capital letter do not differ statistically by the Tukey-Kramer test ( $p < 0.05$ ).

### 3.2 Bootstrap Simulation

From the graph in the form of contour lines (Figure 1), it is possible to empirically infer that 20 seeds per s.u. (T2) is the treatment that generates the highest accumulated germination percentage, regardless of number of replicates.

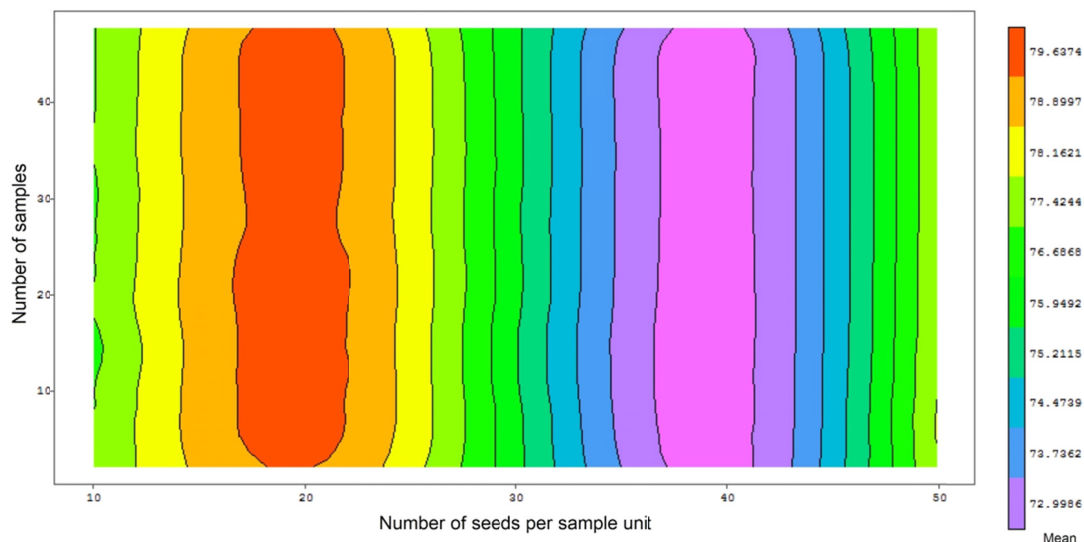


Figure 1. Contour lines of maximum accumulated germination percentages, number of seeds per sampling unit and number of sampling units, by the spline smoothing method

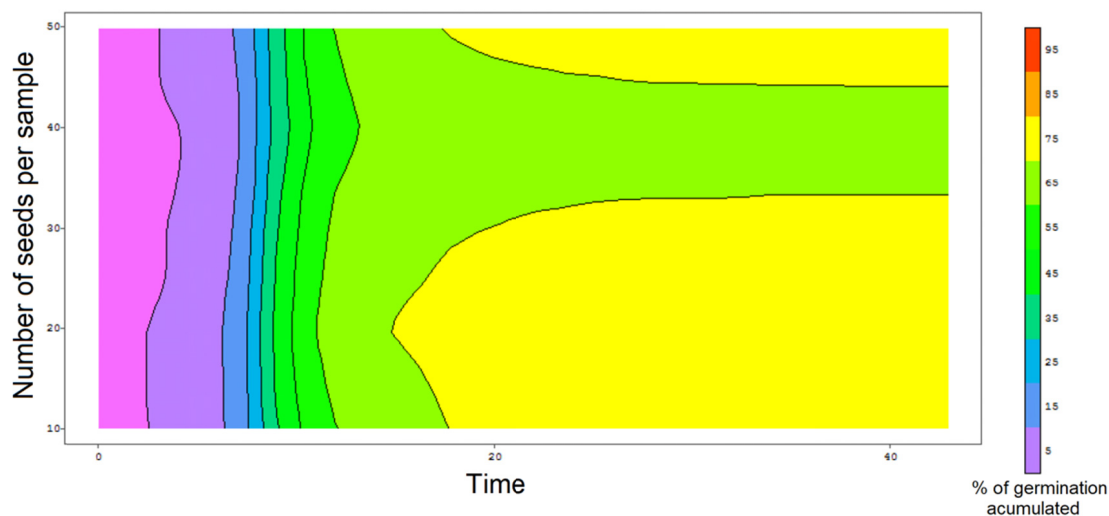


Figure 2. Contour lines of maximum accumulated germination percentages, number of seeds per sampling unit and time, by the spline smoothing method

### 3.3 Non-linear Gompertz Regression Models

For the estimated asymptote ( $\hat{\alpha}$ ), which is the expected theoretical accumulated germination, there is treatment effect ( $p = 0.0477$ ), that is, at least one treatment differs from the others (Table 2), and this difference agrees with results found by the generalized linear model and the bootstrap simulation. The time for the maximum germination speed ( $\hat{\beta}$ ), there is no treatment effect ( $p < 0.0965$ ), that is, the number of seeds per s.u. does not influence the time for the maximum germination speed. The  $\hat{\beta}$  mean for treatments was 7.94 days, and the standard error of the mean was 0.05. In relation to the maximum germination speed, there is treatment effect ( $p < 0.0001$ ), that is, number of seeds per paper roll influences the maximum germination speed. Treatments can be separated into 3 distinct groups according to the Tukey-Kramer test at 5% significance level applied to their means, and within groups, the means of treatments are not statistically different from each other (Table 2).

Table 2. Means of estimated asymptotes, maximum germination speed and time for interruption of the germination process (days), (standard error of the mean) according to the number of seeds per sampling unit

Treatment	Estimated asymptotes	Maximum germination speed	Interruption times
1	76.74 <sup>(*)</sup> (2.24) AB	29.56 <sup>(*)</sup> (4.95) A	11.45 <sup>(*)</sup> (0.42) A
2	79.15 (1.70) A	21.25 (1.57) BC	10.56 (0.19) AB
3	75.39 (1.17) AB	23.24 (1.76) AB	10.44 (0.20) B
4	72.18 (1.15) B	18.17 (1.42) BC	11.17 (0.22) AB
5	77.28 (1.18) AB	16.31 (0.87) C	11.37 (0.18) AB

Note. (\*) Means followed by the same capital letter in the columns do not differ statistically by the Tukey-Kramer test ( $p < 0.05$ ).

The highest mean group includes T1 and T3 (which means are not statistically different from each other), the intermediate mean group includes T2, T3 and T4 and the lowest mean group includes T2, T4 and T5. The mean T1 value is statistically different from the mean T2, T4 and T5 values; the mean T3 value is statistically different from the mean T5 values.

Considering the time for the beginning of the germination process (lagtime), there is no treatment effect ( $p = 0.3493$ ), that is, the number of seeds per s.u. does not influence the time for the beginning of the germination process. The mean lagtime was 6.17 days, and the standard error of the mean was 0.06. For the time for the interruption of the germination process, there is treatment effect, *i.e.*, the number of seeds per s.u. influences the germination process ( $p < 0.0076$ ).

According to the Tukey-Kramer test at 5% significance level, it could be inferred that the means of treatments can be separated into two distinct groups (Table 2). The group of treatments with the highest means, including T1, T2, T4 and T5, which are not statistically different from each other, and the group with the lowest means, including T2, T3, T4 and T5, which are not statistically different from each other. The mean T1 value is statistically different from the mean T3 value.

### 3.4 Graphs in the Form of Contour Lines

It could be inferred from the graph that T2 reaches the highest accumulated germination percentage faster than the other treatments, and that the maximum accumulated germination percentage of T4 is lower than those of the other treatments.

## 4. Discussion

Sampling sufficiency studies for germination tests are scarce, especially with little explored genera such as *Annona*. Works with this genus have standardization as to the number of seeds per sampling unit, so that some end up having their validity questioned due to the small amount of seeds used. Thus, this work elucidated which would be the minimum number of seeds that could be used for the test to have sampling sufficiency, but the choice of the amount to be used is the researcher's decision.

The different empirical and statistical approaches used in this work differ from the more common methods applied in the area of sampling sufficiency for germination tests, such as that proposed by (Cochran, 1977) (Bartlett et al., 2001), and methods developed later, such as (Ribeiro-Oliveira et al., 2016) which is the adaptation of the maximum curvature method. This study also used the combination of several techniques that converge on a common result and a very large sampling, avoiding that the results obtained are source of sample deficiency.

Germination tests aim to determine the maximum germination potential of a seed lot, which can be used to compare the quality of different seed lots and also to estimate the value for field sowing. When researches are concerned, tests can compare efficiency of treatments applied to seeds. Thus, accumulated germination percentage, germination speed and mean germination time are variables necessary in evaluations of the germination process. Thus, results such as those obtained in this experiment are important because they allow optimizing the experimentation time and reduce waste of resources.

The lagtime, which is the time for the beginning of the germination process, shows that seed evaluations can be started on the sixth day after the germination test is set up, and this time is not influenced by the number of seeds per sampling unit. The time for maximum germination speed shows that a large number of germinated seeds can be expected from 8 days, but the maximum germination speed to be reached will depend on the number of seeds per sampling unit, and in general, the smaller the amount of seeds per sampling unit, the faster the maximum

speed will be reached.

The time for the interruption of the germination process denotes the duration of the experiment, and this variable is important because it can result in resource savings and also minimize the risk of contamination of seeds by fungi and bacteria that can damage them due to long evaluation periods, which could compromise the results.

In the case of this experiment, although the interruption times are different from each other according to the number of seeds per sampling unit, this gain in time will not be significant for the researcher, since, as can be seen in Table 2, the real gain in time would be only 1 day. Thus, although treatment 1, of 10 seeds per sampling unit, is statistically different from T3, of 30 seeds per sampling unit, the gain in using greater number of seeds would be negligible compared to the disadvantage of using 3 times more seeds per sampling unit.

Based on the combination of empirical and statistical techniques, it could be inferred that the use of 10 seeds per sampling unit (T1) will already make this unit representative, which should be combined with number of replicates that fit the experimental design determined by the researcher. Considering the fact that often there is a restriction of the number of seeds available for germination tests, especially when it comes to native species, this result elucidates the viability for these works with the use of fewer seeds.

In this context, the results obtained in this work support what has been developed with the genus *Annona* in relation to the amount of seeds per sampling unit used in germination tests, such as Carvalho et al. (2018), who used 25 *Annona × atemoya* seeds, Mendonça et al. (2007) 15 *Annona muricata* L. seeds and Oliveira and Andrade (2005) 10 *Annona montana* seeds per sampling unit. In the same way, the number of replicates, since the bootstrap simulation technique is not a statistical analysis but rather an empirical method, allows us inferring that the number of replicates does not influence the accumulated germination percentage, which could be from 3 replicates, which already obtains maximum accumulated germination percentage. However, it should be emphasized that the number of replicates to be used in each new experiment should be determined at the time of the experimental planning, according to the basic precepts of statistics, such as respecting minimum values of residue degrees of freedom.

## 5. Conclusion

Based on the different descriptive and statistical approaches, it could be concluded that for germination tests with atemoya seeds (*Annona × atemoya* Mabb.) to have statistical validity, they should use at least 10 seeds per sampling unit. This statistical validity extends to tests with 3 replicates or more, taking into account the experimental design.

## References

- Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19, 43-50.
- Braga, J. F., Ferreira, G., Pinho, S. Z. de, Braga, L. F., & Sousa, M. P. (2010). Germination of atemoya (*Annona cherimola* Mill. × *A. squamosa* L.) Cv. Gefner seeds subjected to treatments with plant growth regulators. *International Journal of Science and Nature*, 1, 120-126.
- Carvalho, D., Deved U., Cruz, M. A. da, Osipi, E. A. F., Cossa, C. A., Colombo, R. C., & Sorace, M. A. F. (2018). Plant growth regulators on atemoya seeds germination. *Nucleus*, 15, 457-462. <https://doi.org/10.3738/1982.2278.2832>
- Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: John Wiley & Sons.
- Corsato, J. M., Ferreira, G., & Barbedo, C. J. (2012). Desiccation tolerance in seeds of *Annona emarginata* (Schldtl.) H. Rainer and action of plant growth regulators on germination. *Brazilian Journal Plant Physiology*, 24, 253-260. <https://doi.org/10.1590/S1677-04202012000400004>
- De La Cruz, C. I., González, E. A. R., Guevara, F. P., & Jiménez, G. L. F. (2011). Liriodenine, early antimicrobial defence in *Annona diversifolia*. *Zeitschrift für Naturforschung*, 66, 377-384. <https://doi.org/10.1515/znc-2011-7-809>
- Ferreira, G., De-La-Cruz-Chacón, I., & González-Esquinca, A. R. (2016). Overcoming seed dormancy in *Annona macrophyllata* and *Annona purpurea* using plant growth regulators. *Revista Brasileira de Fruticultura*, 38. <https://doi.org/10.1590/0100-29452016234>
- Gimenez, J. I., Ferreira, G., & Cavariani, C. (2014). Tetrazolium test for assessment of seed viability of atemoya (*Annona cherimola* Mill. × *A. squamosa* L.). *Journal of Seed Science*, 36, 357-361. <https://doi.org/>

10.1590/2317-1545v36n31004

- Lemos, E. E. P. de. (2014). The Production of Annona Fruits in Brazil. *Revista Brasileira de Fruticultura*, 36, 77-85. <https://doi.org/10.1590/S0100-29452014000500009>
- Mendonça, V., Ramos, J. D., Pio, R., Chaltein, T., & Gontijo, A. (2007). Superação de dormência e profundidade de semeadura de sementes de gravioleira. *Revista Caatinga*, 20, 73-78.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, 135, 370-384. <https://doi.org/10.2307/2344614>
- Oliveira, I., & Andrade, R. (2005). Influence of the temperature on seed germination of *Annona Montana*. *Revista Brasileira de Fruticultura*, 27, 344-345. <https://doi.org/10.1590/S0100-29452005000200041>
- Passos, J. R. de S., Pinho, S. Z. de, Carvalho, L. R. de, & Mischan, M. M. (2012). Critical points in logistic growth curves and treatment comparisons. *Scientia Agricola*, 69, 308-312. <https://doi.org/10.1590/S0103-90162012000500004>
- Ribeiro-Oliveira, J. P., Ranal, M. A., Garcia de Santana, D., & Pereira, L. A. (2016). Sufficient sample size to study seed germination. *Australian Journal of Botany*, 64, 295. <https://doi.org/10.1071/BT15254>
- Wahba, G. (1990). *Spline models for observational data*, book 59, Society for Industrial Mathematics. <https://doi.org/10.1137/1.9781611970128>
- Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., & Hochberg, Y. (2000). *Multiple Comparisons and Multiple Tests Using the SAS System* (1st ed.). Cary, USA: SAS Institute.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).