

## Determination of sample size in polyembryonic citrus seeds using robust methods

Elisa del Carmen Martínez-Ochoa<sup>1</sup>  
Julio César Buendía-Espinoza<sup>2,5</sup>  
Ángel Villegas-Monter<sup>3</sup>

1 Departamento de Preparatoria Agrícola-Universidad Autónoma Chapingo. Carretera México-Texcoco km 38.5, Texcoco, Estado de México. CP. 56230.

2 Maestría en Agroforestería para el Desarrollo Sostenible-Departamento de Suelos-Universidad Autónoma Chapingo. Carretera México-Texcoco km 38.5, Texcoco, Estado de México. CP. 56230.

3 PREGEP-Fruticultura-Colegio de Postgraduado-Campus Montecillo. Carretera México-Texcoco km 36.5, Montecillo, Texcoco Estado de México. CP. 56230.

Autor para correspondencia: [jbuendiae@chapingo.mx](mailto:jbuendiae@chapingo.mx).

### Abstract

The percentage of polyembryonic seeds and the number of embryos per seed (NE) are variables commonly used to estimate the degree of polyembryony in citrus, a characteristic used as an indicator of clonal or vegetative propagation of selected genotypes. However, many studies omit the description of the methods used to determine the appropriate sample size in the quantification of these variables. The purpose of this work was to propose a robust statistical methodology for calculating the optimal sample size in polyembryonic studies in citrus using the maximum distance method and the F-test. Five citrus genotypes (C-35 citrange, Volkamer lemon, Amblicarpa mandarin, Valencia orange and Mineola tangelo) were evaluated over two harvest cycles (2020 and 2021) by counting the number of embryos in 100 seeds per genotype. Since the variable of the number of embryos per seed was not normally distributed, the maximum distance method was applied using coefficients of variation calculated with the winsorized mean and a nonparametric bootstrap procedure. The results indicated optimal sample sizes of 26 to 28 seeds, with coefficients of variation of 6% to 11%, depending on the genotype. It was concluded that, for the genotypes evaluated, analyzing samples of up to 50 seeds is sufficient to estimate the degree of polyembryony accurately.

### Keywords::

apomixis, citrus, maximum distance method, nonparametric bootstrap.



## Introduction

Polyembryony in citrus is a characteristic of great agronomic relevance, as it allows the obtaining of plants genetically identical to the mother plant through seeds, which constitutes a strategic advantage for the commercial propagation of selected genotypes (Duarte *et al.*, 2013). This type of asexual reproduction, known as sporophytic adventitious apomixis, originates from the nucellar tissue of the ovary and results in the formation of somatic embryos that coexist with the zygotic embryo within the same seed (Martínez-Ochoa *et al.*, 2012). As a result, the germination of nucellar embryos allows the genotype's genetic identity to be preserved, whereas the occasional emergence of sexual embryos can generate genetic segregation.

In addition to its genetic importance, propagation by polyembryonic seeds offers operational advantages over other vegetative methods, such as reducing the risk of pathogen transmission and facilitating the transport and handling of plant material between regions or countries (Andrade-Rodríguez *et al.*, 2003; Kishore *et al.*, 2012; Villegas-Monter *et al.*, 2023). For these reasons, polyembryony has been widely studied in citrus, particularly through two key variables: the percentage of polyembryonic seeds and the number of embryos per seed, which are used as indicators of the clonal propagation potential of the genotypes evaluated (Martínez-Ochoa *et al.*, 2012; Duarte *et al.*, 2013).

Nonetheless, despite the abundance of studies on polyembryony in citrus, most do not explicitly report the methodology employed to determine the sample size used for quantifying these variables. Highly variable sample sizes are observed in the literature, ranging from 100 to 1 000 seeds, without clear statistical justification (Andrade-Rodríguez *et al.*, 2003; Darlan-Ramos *et al.*, 2006; Kishore *et al.*, 2012). This omission limits the reproducibility of studies and can lead to inefficient use of experimental resources.

Some studies have tried to address this problem by estimating the sample size from the coefficient of variation, using the maximum curvature method. Such is the case of Santos *et al.* (2015) in citrus and Arellano-Durán *et al.* (2018) in mango. Nevertheless, these approaches are based on assumptions of normality and the use of the arithmetic mean as an estimator of central tendency, which may be inappropriate when the data present asymmetric distributions or outliers, a common condition in biological variables such as the number of embryos per seed (Leys *et al.*, 2013).

The proper determination of the sample size is a fundamental element of experimental design, since it allows optimizing the accuracy of the estimates and the use of available resources (Avanza *et al.*, 2010). Traditionally, this calculation is based on dispersion measures relative to the arithmetic mean; however, when the data do not meet the normality assumption, the meaning can be strongly influenced by extreme values, which distorts the estimation of the coefficient of variation and consequently, the required sample size.

In this scenario, the use of robust and nonparametric statistical methods represents a suitable alternative for analyzing variables with asymmetric behavior. These methods include winsorization, which reduces the influence of outliers on central tendency and dispersion estimators (Barnett

and Lewis, 1994) and the nonparametric bootstrap, which allows estimating population parameters without assuming a specific theoretical distribution of the data (Efron and Tibshirani, 1986). These techniques have demonstrated their usefulness in biological and agronomic studies with high variability (Martínez-Ochoa *et al.*, 2012).

In this context, the study aimed to propose a robust statistical methodology to estimate the optimal sample size in citrus polyembryonic studies, particularly when the variable of the number of embryos per seed does not meet the normal assumption. To this end, five citrus genotypes were evaluated and the maximum distance method was applied, using winsorized coefficients of variation estimated by a nonparametric bootstrap, complemented by an F-test to validate the selected sample size objectively.

## Materials and methods

### Study site, plant material and sampling

Ten mature fruits per genotype were collected from open-pollinated trees of the following citrus: Valencia orange (*Citrus sinensis* (L.) Osbeck), Mineola tangelo (*C. paradisi* Macf. × *C. reticulata* Blanco), C-35 citrange (*C. sinensis* × *Poncirus trifoliata*), Volkamer lemon (*C. volkameriana* Pasq.) and Amblicarpa mandarin (*C. amblycarpa* (Hassk.) Ochse). The collection was carried out during two consecutive production cycles (2020 and 2021) in groves of the 'Cazones' nursery, Veracruz, Mexico (20°42' north latitude; 97°18' west longitude; 23 masl).

Ten viable seeds were randomly selected from each fruit, forming a pre-sample of 100 seeds per genotype and per cycle. The sampling unit was a seed. Each seed had the testa and tegmen manually removed to expose the embryos, which were separated and counted individually to record the number of embryos per seed (NE).

### Assumption evaluation and exploratory analysis

The normality of the NE variable was assessed using the Kolmogorov-Smirnov test, and the presence of outliers was identified using box plots. Since the variable did not meet the assumption of normality and exhibited extreme values, the use of the arithmetic mean as an estimator of central tendency to determine the sample size was discarded. All statistical analyses were performed using SAS software (SAS Institute, 2003).

### Estimation of the coefficient of variation using nonparametric bootstrap

The winsorized coefficient of variation was estimated by bootstrap ( $CV_w^B$ ) from the winsorized mean ( $\bar{x}_{\alpha,1-\alpha}^{*w}$ ) and the winsorized standard deviation ( $S_w^*$ ). To this end,  $B = 200$  nonparametric bootstrap replicates were generated, according to Efron and Tibshirani (1986); Ramírez *et al.* (2013), in order to accurately calculate the ( $CV_w^B$ ) in each genotype. 1)  $B$  independent bootstrap samples  $x_1^*, \dots, x_n^*$  were generated, obtained with replacement from the original dataset  $x_1, \dots, x_n$ ; 2) the bootstrap replicates of the estimator corresponding to each sample were evaluated:  $\theta^{*(b)} = CV_w^{*(b)}$  from  $b = 1,$

...,  $B$ ; and 3)  $CV_w^B$  was estimated as the average of  $B$  replicates of the estimator, using the following expression:

$$CV_w^B = \frac{1}{B} \sum_{b=1}^B \frac{S_w^{*(b)}}{\bar{x}_{\alpha,1-\alpha}^{*(b)}} \cdot 100$$

Where:  $\bar{x}_{\alpha,1-\alpha}^{*(b)}$  is the winsorized mean;

$$\bar{x}_{\alpha,1-\alpha}^{*(b)} = \frac{1}{n} \left( k \cdot x_{k+1} + \sum_{i=k+2}^{n-k-1} x_i + k \cdot x_{n-k} \right)$$

$S_w^{*(b)}$  is the winsorized standard deviation ( $S_w^* = \sqrt{S_w^{2*}}$ );  $n$  is the sample size; and # it refers to a proportion of observations to winsorized at each end or either end (0.1, 0.2). According to Erceg and Mirošević (2008), a winsorization of 20% was applied, replacing 10% of the values at each end with the values closest within the remaining range. In the following expression, this criterion was applied to estimate the winsorized variance and consequently, the coefficient of variation:

$$S_w^{2*(b)} = \frac{k(x_{k+1} - \bar{x}_{\alpha,1-\alpha}^{*w})^2 + \sum_{i=k+2}^{n-k-1} (x_i - \bar{x}_{\alpha,1-\alpha}^{*w})^2 + k(x_{n-k} - \bar{x}_{\alpha,1-\alpha}^{*w})^2}{n-1}$$

Where:  $S_w^{2*(b)}$  is the winsorized variance and  $B$  is the number of replicates; when  $B \rightarrow \infty$ , the bootstrap estimator of the winsorized coefficient of variation converges to the ideal bootstrap estimator:

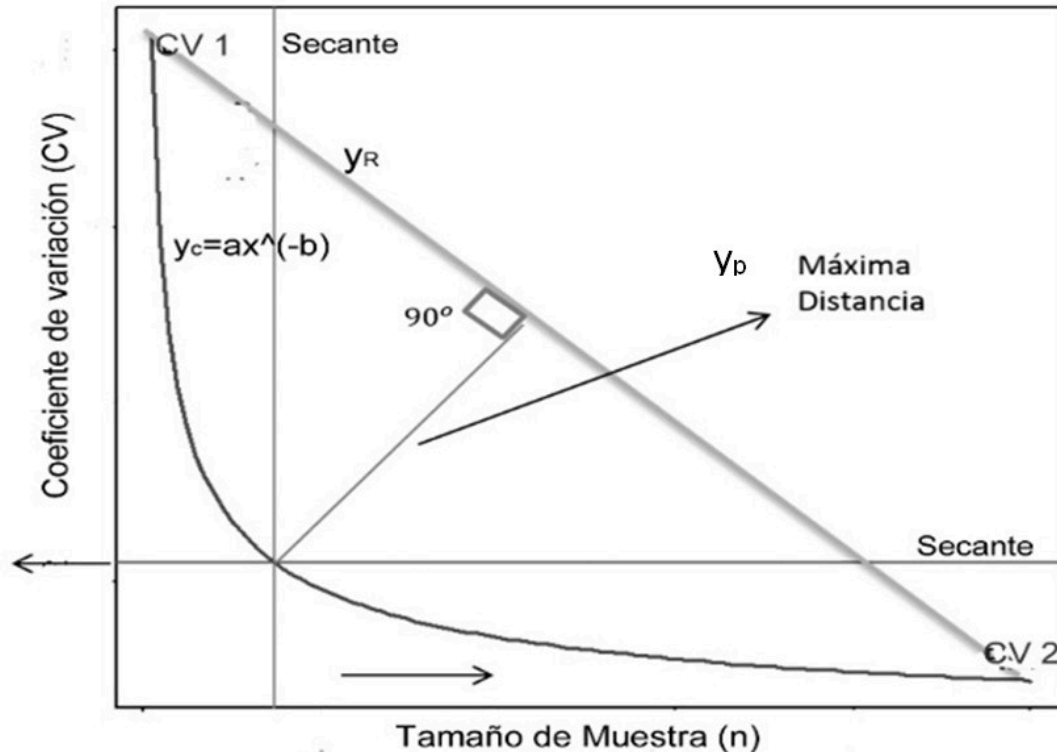
$$\lim_{B \rightarrow \infty} CV_w^B(\theta) = CV_w^F(\theta^*), \text{ implying that } CV_w^B \text{ converges to } CV_w^F.$$

### Determination of sample size using the maximum distance method (MDM)

Winsorized coefficients of variation ( $CV_w$ ) were estimated for different sample sizes and plotted as a function of  $n$ . A power model of the form:  $y = ax^b$ , proposed by Lessman and Atkins (1963), was fitted. Where  $a$  and  $b$  are estimated parameters. Based on this fit, the maximum curvature criterion defines the optimal sample size as the point on the curve at which additional increments of  $n$  do not generate relevant reductions in the winsorized coefficient of variation ( $CV_w$ ). The location of this point can be carried out by visual inspection, a subjective procedure, or objectively using the maximum distance method (MDM) (Paranaíba *et al.*, 2009). In this study, MDM was applied as described by Lorentz *et al.* (2012) (Figure 1).



Figure 1. Maximum curvature with the method of maximum distance between the curve ( $y_c$ ) and the line ( $y_r$ ) that is formed with the estimated coefficient of variation for the smallest ( $CV_1$ ) and for the largest ( $CV_2$ ) sample size.



The MDM identifies the optimal sample size as the point of the fit curve  $y_c$ , described by the function  $y = ax^{-b}$ ; that is  $y_c = ax^{-b}$ , that maximizes the distance from a secant line  $y_r$  drawn between two points of the same curve. To do this, a line  $y_p = ex + f$ , perpendicular to the secant line  $y_r$ , is defined, which intersects the curve  $y_c$  at the point where this distance is maximum, corresponding to the optimal sample size.

The secant line  $y_r$  was defined from the points by  $(x_{cri}, y_{cri})$  and by  $(x_{crt}, y_{crt})$ . Its angular coefficient  $c$  was calculated as:  $c = (y_{crt} - y_{cri}) / (x_{crt} - x_{cri})$ , and the linearity coefficient  $d$  was obtained from the equation of the line  $d = y_{cri} - cx_{cri}$  or  $d = y_{crt} - cx_{crt}$ . The angular coefficient  $e$  of the perpendicular line  $y_p$  was determined by the condition of perpendicularity between the two lines:  $e = -1/c$ . This procedure allowed us to objectively identify the point of maximum distance between the curve  $y_c$  and the line  $y_r$ , which represents the point of maximum curvature and, therefore, the optimal sample size.

The linear coefficient  $f$  of the line  $y_p$  was obtained using the interactive procedure described by Lorentz *et al.* (2012), based on:  $x_{rpj} = (f - d) / (c - e)$ . The distance between the points  $(x_{cj}, y_{cj})$  and  $(x_{rpj}, y_{rpj})$ , located on the line  $y_{pj}$ , perpendicular to  $y_r$ , was calculated as:

$$d_{cR} = \sqrt{(y_{cj} - y_{Rpj})^2 + (x_{cj} - x_{Rpj})^2}$$

Based on the above, the procedure for applying the MDM was as follows: 1) values were taken on the  $x$ -axis, between the interval  $[x_{cri}, x_{crl}]$  and a spacing  $\Delta x = (x_{crl} - x_{cri})/n$ . Where:  $n$  is the number of intervals into which the domain of the function was divided; 2) for each  $x$  value (with  $j = 1, \dots, n$ ),  $y_c(x_j)$  was calculated, setting  $y_c(x_j) = y_p(x_j)$ ; that is, the line  $y_p$ , perpendicular to the line  $y_r$ , intersects the curve  $y_c$  at  $x_j$ ; 3) the values  $y_c(x_j)$  and  $y_p(x_j)$  were named  $y_{cj}$  y  $y_{pj}$ , respectively; 4) the coefficient  $f$  was obtained by substituting the point  $(x_j, y_{pj})$  into the equation of  $y_p$ , so that  $f = y_{pj} - ex_j$ ; 5) the point of intersection  $(x_{rpi}, y_{rpi})$  between the lines  $y_r$  and  $y_{pj}$  was determined by solving the system  $y_{rpi} = cx_{rpi} + d$ ;  $y_{rpi} = ex_{rpi} + f$ , whose solution is:  $x_{rpi} = (f - d)/(c - e)$ ; and 6) the distance between  $(x_{cj}, y_{cj})$  and  $(x_{rpi}, y_{rpi})$  was calculated. This procedure was repeated for  $j = 1, \dots, n$  in the entire interval  $[x_{ci}, x_{cl}]$ ; and the maximum value  $d_{crm\acute{a}x}$  and the corresponding coordinates  $x_{cm\acute{a}x}$ ,  $y_{cm\acute{a}x}$  were identified. The value  $d_{crm\acute{a}x}$  was considered the maximum distance between the curve  $y_c$  and the line  $y_r$ , measured on a segment perpendicular to  $y_r$ . The point  $(x_{cm\acute{a}x}, y_{cm\acute{a}x})$  determined the optimal sample size.

### Sample size validation using the F-test

As a complementary criterion, an F-test was used to assess the homogeneity of variances across consecutive sample sizes. This test was based on comparing absolute variances, since in the experimental design, the primary interest is to minimize the experimental error, represented by variance (Gómez and Gómez, 1984). The F-statistic was calculated as the quotient of two consecutive variances. The hypotheses proposed were:  $H_0: \text{Var}(x)_i = \text{Var}(x)_j$  vs  $H_1: \text{Var}(x)_i \neq \text{Var}(x)_j$ , for all  $i \neq j$ . With a significance level  $\alpha = 0.05$ ,  $H_0$  was rejected when  $p < 0.05$ . In cases where no significant differences were detected between consecutive variances, further increases in sample size were not considered to produce significant reductions in variability.

### Results and discussion

The variable of the number of embryos per seed (NE), evaluated in five citrus genotypes during the 2020 and 2021 harvest cycles, did not meet the assumption of normality according to the Kolmogorov-Smirnov test ( $p \leq 0.05$ ). In all genotypes, positive asymmetric distributions were observed, characterized by a high dispersion of NE relative to its meaning in each genotype. This distribution pattern is consistent with that reported for polyembryonic species, in which the formation of nucellar embryos is regulated by one or a few genes, but may be modulated by environmental and physiological factors that increase intraspecific variability (Andrade-Rodríguez *et al.*, 2003; Kepiro and Roose, 2010). In particular, in the Mineola tangelo genotype, maximum values of up to 13 embryos per seed were recorded, even though the mean was around four embryos, indicating a marked dispersion relative to the arithmetic mean.

The exploratory analysis allowed us to identify outliers in all genotypes evaluated. In Amblicarpa, C-35 and Mineola, seeds with more than nine embryos were considered atypical, while in Valencia and Volkamer, this threshold corresponded to eight and seven embryos, respectively. The systematic presence of these extreme values confirms that the arithmetic mean does not constitute a robust estimator of NE, as pointed out by Leys *et al.* (2013) for variables with asymmetric distributions.

As a consequence of this data structure, the coefficients of variation calculated from the arithmetic mean, without applying robust methods, were high and ranged from 35% to 63% across genotypes and harvest cycles. These values reflect a high relative variability of NE and limit their usefulness as a reliable criterion for determining the sample size. The application of winsorization made it possible to substantially reduce the influence of extreme values on the estimation of variability. When using the winsorized mean, the coefficients of variation were consistently reduced in all genotypes. For example, in Valencia, the coefficients of variation decreased from 35% and 50% to values between 19% and 28%; in Mineola, from 41-43% to 34-38%; in C-35, from 44% to 24-28%; in Volkamer, from 52-63% to 24-30%, and in Amblicarpa, from 41-46% to 25-26%, considering both evaluation cycles. These results are consistent with the point made by Barnett and Lewis (1994), who highlight the usefulness of robust estimators to control for variability induced by outliers.

The winsorized coefficients of variation of the number of embryos per seed (NE) were estimated by nonparametric bootstrap with replacement, using 200 replicates, and are presented in Table 1 and 2 for the 2020 and 2021 cycles, respectively; this approach made it possible to quantify the variability of NE without assuming a theoretical distribution, which is appropriate for biological variables with asymmetric behavior (Efron and Tibshirani, 1986; Davison and Hinkley, 1997).

**Table 1. Winsorized coefficients of variation associated with each sample size (n), with their F-value, in the Amblicarpa, Volkamer, Valencia, C-35, and Mineola genotypes for the 2020 cycle.**

n	Amblicarpa mandarin		Volkamer lemon		Valencia orange		C-35 citrange		Mineola tangelo	
	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value
0	19.38		16.87		15.11		20.43		29.17	
5	15.18	1.74*	13.93	2.55*	10.93	1.99*	14.92	1.45*	22.32	2.22*
10	12.08	2.12*	10.24	2.07*	8.15	1.81*	10.88	2.22*	16.5	2.28*
30	8.99	1.81*	8.9	1.49*	6.88	1.68*	9.49	1.64*	13.01	1.46*
40	8.09	1.64*	7	1.51*	5.75	1.19ns	8.41	1.51*	11.07	1.55*
50	7.31	1.07ns	6.01	1.33*	5.7	1.02ns	7.18	1.07ns	9.67	1.22ns
60	7.34	1.1ns	5.64	1.15ns	5.12	1.07ns	6.26	1.2ns	8.2	1.15ns
70	5.5	1.23ns	4.6	1.29ns	4.47	1.04ns	6.32	1.27ns	8.67	1.19ns
80	6.24	1.28ns	4.33	1.25ns	3.89	0.98ns	5.96	1.16ns	7.8	1.27ns
90	6.01	1.24ns	4.32	1.21ns	3.67	1.28ns	5.25	1.17ns	6.95	1.14ns
100	5.18	1.01ns	4.25	1.26ns	3.7	0.92ns	5.55	1.18ns	7.89	0.99ns

\*= significant ( $p \leq \alpha = 0.05$ ); ns= not significant.

**Table 2. Winsorized coefficients of variation associated with each sample size (n), with their F-value, in the Amblicarpa, Volkamer, Valencia, C-35 and Mineola genotypes for the 2021 cycle.**

n	Amblicarpa mandarin		Volkamer lemon		Valencia orange		C-35 Citrange		Mineola t.	
	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value	CV <sub>w</sub> <sup>B</sup>	F-value
0	18.9		18.71		22.15		19.85		23.76	
5	14.1	2.1*	13.55	1.84*	18.49	1.52*	13.76	1.64*	14.85	2.36*
10	8.87	1.75*	10.14	1.92*	13.66	1.81*	9.19	1.57*	11.64	2.13*
30	7.45	1.46*	7.76	1.52*	9.83	1.42*	8.54	1.54*	9.56	1.69*
40	6.23	1.33*	7.42	1.54*	9.01	1.12ns	7.66	1.41*	8.9	1.7*
50	5.87	1.25ns	6.5	0.94*	8.07	1.16ns	6.17	1.16ns	7.19	1.19ns
60	5.35	1.01ns	5.59	1.18ns	7.64	1.14ns	5.84	1.3ns	6.29	1.12ns
70	5.46	1ns	5.94	0.93ns	6.82	0.8ns	5.42	0.88ns	5.73	0.88ns
80	4.59	1.19ns	5.3	1.19ns	6.25	0.98ns	4.37	1.05ns	5.73	1.09ns
90	4.9	1.12ns	4.75	1.07ns	5.84	1.03ns	4.9	1.24ns	5.48	1.03ns
100	4.22	1.02ns	4.75	1.12ns	5.24	0.96ns	4.72	1.09ns	4.7	1.15ns

\*= significant ( $p \leq \alpha = 0.05$ ): ns= not significant.

The choice of 200 replicates was justified by the variation in the coefficient of variation between genotypes and cycles and is within the recommended range to obtain stable estimates of standard error and variance estimators in contexts of high variability (Efron and Tibshirani, 1986; Ramírez *et al.*, 2013). A progressive decrease in the coefficient of variation was observed in all genotypes as the sample size increased (Table 1 and 2), with pronounced reductions to approximately 20 seeds and an asymptotic trend from sample sizes close to 60-70 seeds. This behavior is characteristic of uniformity trials used to define optimal sample sizes in agronomic studies (Lessman and Atkins, 1963; Paranaiba *et al.*, 2009).

The application of MDM allowed us to objectively identify the point of maximum curvature of the relationship between the winsorized coefficient of variation and the sample size, thereby avoiding the subjectivity associated with the visual inspection of the traditional maximum curvature method (Paranaiba *et al.*, 2009; Arellano-Durán *et al.*, 2018). Based on this criterion, the estimated optimal sample size for the five genotypes, considering both harvest cycles, was within a range of 26 to 28 seeds. In these sample sizes, the coefficients of variation ranged from 6.01% to 11%, indicating controlled variability and adequate accuracy in estimating NE.

In previous studies that employed the maximum curvature method to determine sample size in polyembryonic species, Santos *et al.* (2015) reported a range of 10 to 23 seeds in 15 citrus genotypes, whereas Arellano-Durán *et al.* (2018) recommended testing eight polyembryonic seeds in Manila and Ataulfo mango cultivars. However, in both cases, it is not specified whether the data complied with the assumption of normality, a necessary condition to justify the use of the coefficient of variation as an estimation criterion.

Additionally, the F-test applied as a complementary criterion for validating the sample size showed that, in most genotypes, no significant differences were detected between consecutive variances from sample sizes between 30 and 50 seeds, depending on the genotype and the harvest cycle. In particular, no significant differences were observed in Valencia from 30 seeds, in Volkamer from 50 seeds, and in Amblicarpa, C-35, and Mineola from 40 seeds. These results indicate that additional increases in sample size do not produce significant reductions in absolute variability, which aligns with Gómez and Gómez (1984).

Overall, the MDM allowed us to identify an optimal sample size between 26 and 28 seeds, while the F-test suggested slightly wider ranges (30-50 seeds), depending on the genotype; this convergence between criteria supports the robustness of the approach and provides an objective framework for selecting sample sizes in citrus polyembryonic studies. Based on both results and considering the stability observed in two harvest cycles, it is recommended to analyze 27-40 seeds in Amblicarpa, 26-50 in Volkamer, 26-30 in Valencia, and 26-40 in C-35 and Mineola. In this sense, the proposed method represents a robust alternative for variables that do not comply with normality, since defining adequate sample sizes improves the accuracy of the estimators, strengthens the reliability of the inferences, and can be extrapolated to other species and quantitative variables with asymmetric distributions, expanding its usefulness in different experimental contexts.

## Conclusions

The results of this study show that winsorization, combined with a nonparametric bootstrap procedure, is an effective strategy for estimating the sample size in citrus polyembryonic studies, particularly when the variable of the number of embryos per seed presents asymmetry and outliers. The application of the maximum distance method, complemented by the F-test, enabled us to objectively define that a sample size between 26 and 50 seeds is sufficient to obtain accurate estimates with controlled variability in the five genotypes evaluated. Overall, the proposed methodology contributes to the standardization of sampling in polyembryonic citrus seeds and offers a replicable procedure for other species and quantitative variables with non-normal distributions.

## Bibliography

- 1 Andrade-Rodríguez, M.; Villegas-Monter, A. y García-Velázquez, A. 2003. Características morfológicas del fruto y poliembriónia de tres portainjertos de cítricos. *Revista Chapingo Serie Horticultura*. 9(2):255-263.
- 2 Arellano-Durán, N.; Villegas-Monter, A. and Muñoz-Orozco, A. 2018. Optimum simple size in quantitative characteristics on seed of polyembryonic mango. *Revista Brasileira de Fruticultura*. 40(3):e-519. <https://doi.org/10.1590/0100-29452018519>.
- 3 Avanza, M. M.; Bramardi, S. J. y Mazza, S. M. 2010. Tamaño óptimo de muestra para evaluar el patrón de crecimiento de frutos de naranjo 'Valencia late'. *Revista Brasileira de Fruticultura*. 32(4):1154-1163. <https://doi.org/10.1590/S0100-29452010005000132>.
- 4 Barnett, V. D. and Lewis, T. 1994. *Outliers in statistical data*. 3rd Ed. Chichester, UK. John Wiley & Sons. 69-72 pp.
- 5 Darlan-Ramos, J.; Neto, S. E.; Castro, N. E. A.; Martis, P. C. C. y Correia, M. G. 2006. Poliembriónia e caracterização de frutos de citrumelo 'Swingle' e de *Poncirus trifoliata*. *Ciência e Agrotecnologia*. 30(1):88-91. <https://doi.org/10.1590/S1413-70542006000100012>.

- 6 Davison, A. C. and Hinkley, D. V. 1997. Bootstrap methods and their application. 1st Ed. UK: Cambridge University Press. 45-54 pp.
- 7 Duarte, F. E. V.; Barros, D. R.; Girari, E. A.; Soares-Filho, W. Dos S. e Passos, O. S. 2013. Poliembrionia e atributos morfológicos de sementes de porta-enxertos de citros. *Revista Brasileira de Fruticultura*. 35(1):246-254. <https://doi.org/10.1590/S0100-29452013000100028>.
- 8 Efron, B. and Tibshirani, R. J. 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*. 1(1):54-75.
- 9 Erceg-Hurn, D. M. and Miroseovich, V. M. 2008. Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist*. 63(7):591-601. <https://doi.org/10.1037/0003-066X.63.7.591>.
- 10 Kepiro, J. L. and Roose, M. L. 2010. AFLP markers closely linked to a major gene essential for nucellar embryony (apomixis) in *Citrus maxima* × *Poncirus trifoliata*. *Tree Genetics & Genomes*. 6(1):1-11. <https://doi.org/10.1007/s11295-009-0223-z>.
- 11 Kishore, K.; Monika, N.; Rinchen, D.; Lepcha B. and Pandey, B. 2012. Polyembryony and seedling emergence traits in Apomitic citrus. *Scientia Horticulturae*. 1(138):101-107.
- 12 Lessman, K. J. and Atkins, R. E. 1963. Comparisons of planning arrangements and estimates of optimum hill plot for grain sorghum yield tests. *Crop Science*. 3(1):477-481.
- 13 Leys, C.; Klein, O.; Bernard, P. and Licata, L. 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*. 49(4):764-766. <https://doi.org/10.1016/j.jesp.2013.03.013>.
- 14 Lorentz, H. L.; Erichsen, R. E. y Lucio, A. D. 2012. Proposta de método para estimação de tamanho de parcela para culturas agrícolas. *Revista Ceres*. 59(6):772-780.
- 15 Martínez-Ochoa, E. C.; Andrade, M. R.; Rocandio, M. R. and Villegas, M. A. 2012. Identification of zygotic and nucellar seedlings in polyembryonic mango cultivars. *Pesquisa Agropecuária Brasileira*. 47(11):1629-1636. <https://doi.org/10.1590/S0100-204X2012001100010>.
- 16 Paranaíba, P. F.; Ferreira, D. F. y De Moraes, A. R. 2009. Tamanho ótimo de parcelas experimentais: proposicao de metodos de estimacao. *Revista Brasileira de Biometria*. 27(2):255-268.
- 17 Ramírez, I. C.; Barrera, C. J. y Correa, J. C. 2013. Efecto del tamaño de muestra y el número de réplicas bootstrap. *Ingeniería y Competitividad*. 15(1):93-101.
- 18 Santos, C. Q.; De, J. E. A.; Girardi, E. L.; Vieira, C. A.; Da, S. L. y Soares-Filho, W. S. 2015. Tamanho ótimo de amostras de frutos e de sementes para determinação da poliembrionia em citros. *Revista Brasileira de Fruticultura*. 1(37):172-178. <https://doi.org/10.1590/0100-2945-063/14>.
- 19 SAS Institute. 2003. SAS (versión 9.4) [Software]. SAS Institute. <https://www.sas.com>.
- 20 Villegas-Monter, A.; Martínez-Ochoa, E. D. C.; Andrade-Rodríguez, M. and Villegas-Velásquez, I. 2023. Citrus polyembryony. *In: Gonzatto-Pereira, M. and Santos, J. S. (Eds.). Advances in citrus research horticultural and human health aspects*. London, UK. 1-15 pp. <http://dx.doi.org/10.5772/intechopen.105994>.

## Determination of sample size in polyembryonic citrus seeds using robust methods

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 01 November 2025
Date accepted: 01 February 2026
Publication date: 01 February 2026
Publication date: Feb-Mar 2026
Volume: 17
Issue: 2
Electronic Location Identifier: e4247
DOI: 10.29312/remexca.v17i2.4247

### Categories

Subject: Article

### Keywords:

**Keywords:**

apomixis

citrus

maximum distance method

nonparametric bootstrap

### Counts

Figures: 1

Tables: 2

Equations: 4

References: 20