

Determinación del tamaño de muestra en semillas poliembriónicas de cítricos mediante métodos robustos

Elisa del Carmen Martínez-Ochoa¹
Julio César Buendía-Espinoza^{2,§}
Ángel Villegas-Monter³

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.

Resumen

El porcentaje de semillas poliembriónicas y el número de embriones por semilla NE son variables comúnmente utilizadas para estimar el grado de poliembriónia en cítricos, característica utilizada como indicador de propagación clonal o vegetativa de genotipos seleccionados. Sin embargo, muchos estudios omiten la descripción de los métodos empleados para determinar el tamaño de muestra adecuado en la cuantificación de estas variables. El objetivo del presente trabajo fue proponer una metodología estadística robusta para calcular el tamaño de muestra óptimo en estudios de poliembriónia en cítricos, utilizando el método de máxima distancia y la prueba F. Se evaluaron cinco genotipos de cítricos (Citrange C-35, limón Volkameriano, mandarino Amblicarpa, naranja Valencia y tangelo Mineola) durante dos ciclos de cosecha (2020 y 2021), contabilizando el número de embriones en 100 semillas por genotipo. Dado que la variable número de embriones por semilla, no presentó una distribución normal, se aplicó el método de máxima distancia usando coeficientes de variación calculados con la media winsorizada y un procedimiento bootstrap no paramétrico. Los resultados indicaron tamaños óptimos de muestra entre 26 y 28 semillas, con coeficientes de variación entre 6% y 11%, dependiendo del genotipo. Se concluyó que, para los genotipos evaluados, el análisis de muestras de hasta 50 semillas resulta suficiente para estimar con precisión el grado de poliembriónia.

Palabras clave:

apomixis, bootstrap no paramétrico, cítricos, método de máxima distancia.



Introducción

La poliembrionía en cítricos es una característica de gran relevancia agronómica, ya que permite la obtención de plantas genéticamente idénticas a la planta madre a través de semillas, lo que constituye una ventaja estratégica para la propagación comercial de genotipos seleccionados (Duarte *et al.*, 2013). Este tipo de reproducción asexual, conocida como apomixis adventicia esporofítica, se origina a partir del tejido nucelar del ovario y da lugar a la formación de embriones somáticos que coexisten con el embrión cigótico dentro de una misma semilla (Martínez-Ochoa *et al.*, 2012). Como resultado, la germinación de embriones nucleares permite conservar la identidad genética del genotipo, mientras que la emergencia ocasional de embriones sexuales puede generar segregación genética.

Además de su importancia genética, la propagación mediante semillas poliembriónicas ofrece ventajas operativas frente a otros métodos vegetativos, tales como la reducción del riesgo de transmisión de patógenos y la facilidad para el transporte y manejo del material vegetal entre regiones o países (Andrade-Rodríguez *et al.*, 2003; Kishore *et al.*, 2012; Villegas-Monter *et al.*, 2023). Por estas razones, la poliembrionía ha sido ampliamente estudiada en cítricos, particularmente a través de dos variables clave: el porcentaje de semillas poliembriónicas y el número de embriones por semilla, las cuales se utilizan como indicadores del potencial de propagación clonal de los genotipos evaluados (Martínez-Ochoa *et al.*, 2012; Duarte *et al.*, 2013).

No obstante, a pesar de la abundancia de estudios sobre poliembrionía en cítricos, la mayoría de las investigaciones no reportan de manera explícita la metodología empleada para determinar el tamaño de muestra utilizado en la cuantificación de estas variables. En la literatura se observan tamaños muestrales altamente variables, que oscilan entre 100 y 1 000 semillas, sin una justificación estadística clara (Andrade-Rodríguez *et al.*, 2003; Darlan-Ramos *et al.*, 2006; Kishore *et al.*, 2012). Esta omisión limita la reproducibilidad de los estudios y puede conducir a un uso ineficiente de los recursos experimentales.

Algunos trabajos han intentado atender esta problemática mediante la estimación del tamaño de muestra a partir del coeficiente de variación, utilizando el método de la curvatura máxima. Tal es el caso de Santos *et al.* (2015) en cítricos y Arellano-Durán *et al.* (2018) en mango. Sin embargo, estos enfoques se basan en supuestos de normalidad y en el uso de la media aritmética como estimador de tendencia central, lo cual puede resultar inadecuado cuando los datos presentan distribuciones asimétricas o valores atípicos, una condición frecuente en variables biológicas como el número de embriones por semilla (Leys *et al.*, 2013).

La determinación adecuada del tamaño de muestra constituye un elemento fundamental en el diseño experimental, ya que permite optimizar la precisión de las estimaciones y el uso de los recursos disponibles (Avanza *et al.*, 2010). Tradicionalmente, este cálculo se apoya en medidas de dispersión relativas a la media aritmética; sin embargo, cuando los datos no cumplen con el supuesto de normalidad, la media puede verse fuertemente influenciada por valores extremos, lo que distorsiona la estimación del coeficiente de variación y en consecuencia, del tamaño de muestra requerido.

Ante este escenario, el uso de métodos estadísticos robustos y no paramétricos representa una alternativa adecuada para el análisis de variables con comportamiento asimétrico. Entre estos métodos destacan la winsorización, que reduce la influencia de valores atípicos sobre los estimadores de tendencia central y dispersión (Barnett y Lewis, 1994), y el bootstrap no paramétrico, que permite estimar parámetros poblacionales sin asumir una distribución teórica específica de los datos (Efron y Tibshirani, 1986). Estas técnicas han demostrado su utilidad en estudios biológicos y agronómicos con alta variabilidad (Martínez-Ochoa *et al.*, 2012).

En este contexto, el objetivo del estudio fue proponer una metodología estadística robusta para estimar el tamaño de muestra óptimo en estudios de poliembrionía en cítricos, particularmente cuando la variable número de embriones por semilla no cumple con el supuesto de normalidad. Para ello, se evaluaron cinco genotipos de cítricos y se aplicó el método de máxima distancia, utilizando coeficientes de variación winsorizados estimados mediante bootstrap no paramétrico, complementados con una prueba F para la validación objetiva del tamaño de muestra seleccionado.

Materiales y métodos

Sitio de estudio, material vegetal y muestreo

Se recolectaron diez frutos maduros por genotipo provenientes de árboles con polinización abierta de los siguientes cítricos: naranja Valencia (*Citrus sinensis* (L.) Osbeck), tangelo Mineola (*C. paradisi* Macf. × *C. reticulata* Blanco), citrange C-35 (*C. sinensis* × *Poncirus trifoliata*), limón Volkameriano (*C. volkameriana* Pasq.) y mandarino Amblicarpa (*C. amblycarpa* (Hassk.) Ochse). La colecta se realizó durante dos ciclos productivos consecutivos (2020 y 2021) en huertos del vivero 'Cazones', Veracruz, México (20°42' latitud norte; 97°18' longitud oeste; 23 msnm).

De cada fruto se seleccionaron al azar diez semillas viables, integrando una premuestra de 100 semillas por genotipo y por ciclo. La unidad de muestreo fue una semilla. A cada semilla se le retiró manualmente la testa y el tegmen para exponer los embriones, los cuales se separaron y contaron individualmente para registrar el número de embriones por semilla (NE).

Evaluación de supuestos y análisis exploratorio

La normalidad de la variable NE se evaluó mediante la prueba de Kolmogorov-Smirnov y la presencia de valores atípicos se identificó mediante diagramas de caja. Dado que la variable no cumplió con el supuesto de normalidad y presentó valores extremos, se descartó el uso de la media aritmética como estimador de tendencia central para la determinación del tamaño de muestra. Todos los análisis estadísticos se realizaron con el software SAS (SAS Institute, 2003).

Estimación del coeficiente de variación mediante bootstrap no paramétrico

Se estimó el coeficiente de variación winsorizado por bootstrap (CV_w^B) a partir de la media winsorizada ($\hat{x}_{\alpha, 1-\alpha}^{*w}$) y la desviación estándar winsorizada (S_w^*). Para ello, se generaron $B = 200$ réplicas bootstrap no paramétricas, de acuerdo con Efron y Tibshirani (1986); Ramírez *et al.* (2013),

con el fin de calcular con precisión el CV_w^B en cada genotipo. 1) se generaron B muestras bootstrap independientes x_1^*, \dots, x_n^* , obtenidas con reemplazo a partir del conjunto de datos original x_1, \dots, x_n ; 2) se evaluó las réplicas bootstrap del estimador correspondiente a cada muestra: $\theta^{*(b)} = CV_w^{*(b)}$ desde $b = 1, \dots, B$; y 3) se CV_w^B como el promedio de las B réplicas del estimador, mediante estimó la siguiente expresión:

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

Donde la media winsorizada es:

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

$S_w^{*(b)}$ es la desviación estándar winsorizada ($S_w^* = \sqrt{S_w^{2*}}$); n = es el tamaño de la muestra; y # se refiere a una proporción de observaciones a winsorizadas en cada extremo o cualquier extremo (0.1, 0.2). Conforme a Erceg y Mirošević (2008), se aplicó una winsorización del 20%, sustituyendo el 10% de los valores en cada extremo por los valores más cercanos dentro del rango restante. En la siguiente expresión se aplicó este criterio para estimar la varianza winsorizada y en consecuencia, el coeficiente de variación:

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

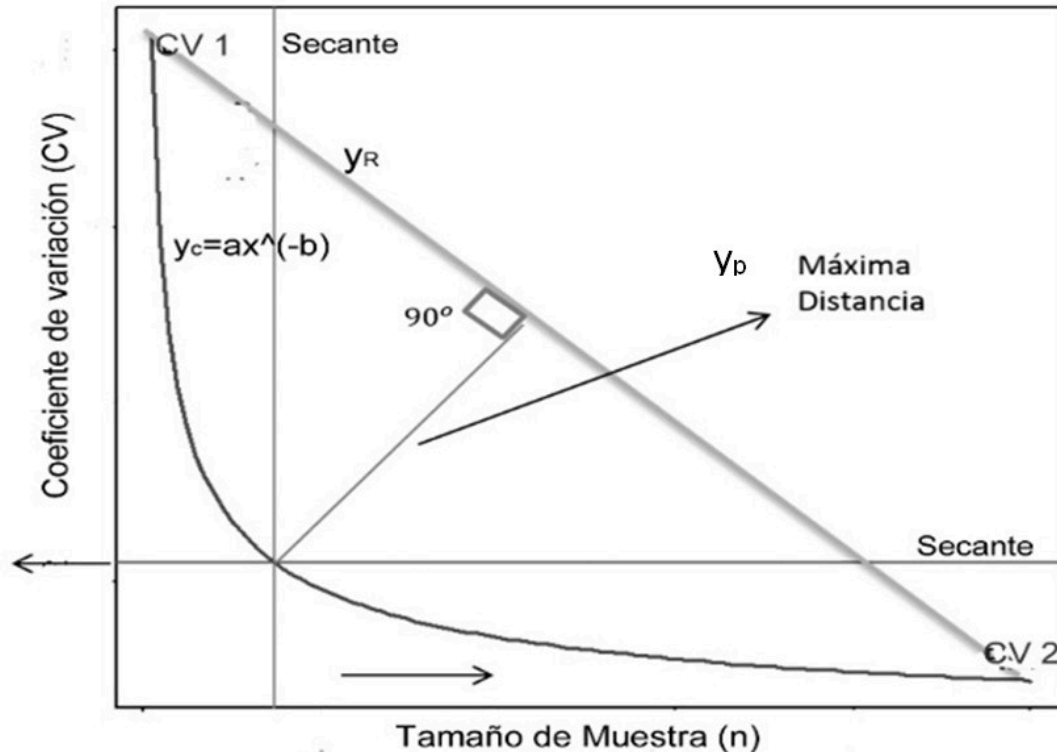
Donde: $S_w^{2*(b)}$ es la varianza winsorizada y B es el número de réplicas; cuando $B \rightarrow \infty$, el estimador bootstrap del coeficiente de variación winsorizado converge al estimador bootstrap ideal, lo cual implica que CV_w^B converge a CV_w^F .

Determinación del tamaño de muestra mediante el método de máxima distancia (MMD)

Los coeficientes de variación winsorizados (CV_w) se estimaron para diferentes tamaños de muestra y se representaron en función de n . Se ajustó un modelo potencial de la forma: $y = ax^{-b}$, propuesto por Lessman y Atkins (1963). Donde a y b son parámetros estimados. Con base en este ajuste, el criterio de curvatura máxima define como tamaño de muestra óptimo el punto de la curva en el cual incrementos adicionales de n no generan reducciones relevantes en el coeficiente de variación winsorizado (CV_w). La localización de este punto puede realizarse mediante inspección visual, procedimiento de carácter subjetivo o de manera objetiva mediante el método de máxima distancia (MMD) (Paranaíba *et al.*, 2009). En este estudio se aplicó el MMD conforme a lo descrito por Lorentz *et al.* (2012) (Figura 1).



Figura 1. Curvatura máxima con el método de máxima distancia entre la curva (y_c) y la recta (y_r) que se forma con el coeficiente de variación estimado para el menor (CV_1) y para el mayor (CV_2) tamaño de muestra.



El MMD identifica el tamaño de muestra óptimo como el punto de la curva de ajuste y_c , descrita por la función $y = ax^{-b}$; es decir, $y_c = ax^{-b}$, que máxima la distancia respecto a una recta secante y_r , trazada entre dos puntos de la misma curva. Para ello se define una recta perpendicular $y_p = ex + f$ a la secante y_r , la cual intersecta la curva y_c en el punto donde dicha distancia es máxima, correspondiente al tamaño de muestra óptimo.

La recta secante y_r se definió a partir de los puntos por (x_{cri}, y_{cri}) y por (x_{crf}, y_{crf}) . Su coeficiente angular c se calculó como: $c = (y_{crf} - y_{cri}) / (x_{crf} - x_{cri})$ y el coeficiente de linealidad d se obtuvo a partir de la ecuación de la recta $d = y_{cri} - cx_{cri}$ o $d = y_{crf} - cx_{crf}$. El coeficiente angular e de la recta perpendicular y_p se determinó mediante la condición de perpendicularidad entre ambas rectas: $e = -1/c$. Este procedimiento permitió identificar de manera objetiva el punto de máxima distancia entre la curva y_c y la recta y_r , el cual representa el punto de curvatura máxima y por tanto, el tamaño de muestra óptimo.

El coeficiente lineal f de la recta y_p se obtuvo mediante el procedimiento interactivo descrito por Lorentz *et al.* (2012), a partir de: $x_{rpj} = (f - d) / (c - e)$. La distancia entre los puntos (x_{cj}, y_{cj}) y (x_{rpj}, y_{rpj}) , ubicada sobre la recta y_{pj} , perpendicular a y_r , se calculó como:

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

Con base en lo anterior, el procedimiento para aplicar el MMD se desarrolló como sigue: 1) se tomaron valores sobre el eje x , entre el intervalo $[x_{cri}, x_{cfr}]$ a un espaciamiento $\Delta x = (x_{cfr} - x_{cri})/n$. Donde: n es el número de intervalos en los que se dividió el dominio de la función; 2) para cada valor x (con $j = 1, \dots, n$), se calculó $y_c(x_j)$ y hacer $y_c(x_j) = y_p(x_j)$; es decir, la recta y_p , perpendicular a la recta y_r , intercepta la curva y_c en x_j ; 3) los valores $y_c(x_j)$ y $y_p(x_j)$ se denominaron y_{cj} y y_{pj} , respectivamente; 4) el coeficiente f se obtuvo sustituyendo el punto (x_j, y_{pj}) en la ecuación de y_p , de manera que $f = y_{pj} - ex_j$; 5) se determinó el punto de intersección (x_{rpj}, y_{rpj}) entre las rectas y_r y y_{pj} , resolviendo el sistema $y_{rpj} = cx_{rpj} + d$; $y_{rpj} = ex_{rpj} + f$, cuya solución es: $x_{rpj} = (f - d)/(c - e)$; y 6) se calculó la distancia entre (x_{cj}, y_{cj}) , y (x_{rpj}, y_{rpj}) . Este procedimiento se repitió para $j = 1, \dots, n$ en todo el intervalo $[x_{ci}, x_{cf}]$; y se identificó el valor máximo $d_{c_{rm\acute{a}x}}$ y las coordenadas correspondientes $x_{c_{m\acute{a}x}}$, $y_{c_{m\acute{a}x}}$. El valor $d_{c_{rm\acute{a}x}}$ se consideró la distancia máxima entre la curva y_c y la recta y_r , medida sobre un segmento perpendicular a y_r . El punto $(x_{c_{m\acute{a}x}}, y_{c_{m\acute{a}x}})$ determinó el tamaño de muestra óptimo.

Validación del tamaño de muestra mediante la prueba F

Como criterio complementario, se aplicó una prueba F para evaluar la homogeneidad de varianzas entre tamaños de muestra consecutivos. Esta prueba se fundamentó en la comparación de varianzas absolutas, dado que en el diseño experimental el interés principal es minimizar el error experimental, representado por la varianza (Gómez y Gómez, 1984). El estadístico F se calculó como el cociente entre dos varianzas consecutivas. Las hipótesis planteadas fueron: $H_0: \text{Var}(x)_i = \text{Var}(x)_j$ vs $H_1: \text{Var}(x)_i \neq \text{Var}(x)_j$, para toda $i \neq j$. Con un nivel de significancia $\alpha = 0.05$, se rechazó H_0 cuando $p < 0.05$. En los casos en los que no se detectaron diferencias significativas entre varianzas consecutivas, se consideró que incrementos adicionales en el tamaño de muestra no producían reducciones significativas en la variabilidad.

Resultados y discusión

La variable número de embriones por semilla (NE), evaluada en cinco genotipos de cítricos durante los ciclos de cosecha 2020 y 2021, no cumplió con el supuesto de normalidad de acuerdo con la prueba de Kolmogorov-Smirnov ($p \leq 0.05$). En todos los genotipos se observaron distribuciones asimétricas positivas, caracterizadas por una alta dispersión del NE respecto a su media en cada genotipo. Este patrón de distribución es consistente con lo reportado para especies poliembriónicas, en las cuales la formación de embriones nucleares está regulada por uno o pocos genes, pero puede verse modulada por factores ambientales y fisiológicos que incrementan la variabilidad intraespecífica (Andrade-Rodríguez *et al.*, 2003; Kepiro y Roose, 2010). En particular, en el genotipo tangelo Mineola se registraron valores máximos de hasta 13 embriones por semilla, aun cuando la media se situó alrededor de cuatro embriones, lo que evidencia una marcada dispersión respecto a la media aritmética.

El análisis exploratorio permitió identificar valores atípicos en todos los genotipos evaluados. En Amblicarpa, C-35 y Mineola se consideraron atípicas las semillas con más de nueve embriones, mientras que en Valencia y Volkameriano este umbral correspondió a ocho y siete embriones, respectivamente. La presencia sistemática de estos valores extremos confirma que la media

aritmética no constituye un estimador robusto del NE, tal como lo señalan Leys *et al.* (2013) para variables con distribuciones asimétricas.

Como consecuencia de esta estructura de datos, los coeficientes de variación calculados a partir de la media aritmética, sin aplicar métodos robustos, fueron elevados y oscilaron entre 35% y 63% entre genotipos y ciclos de cosecha. Estos valores reflejan una alta variabilidad relativa del NE y limitan su utilidad como criterio confiable para la determinación del tamaño de muestra. La aplicación de la winsorización permitió reducir de manera sustancial la influencia de los valores extremos sobre la estimación de la variabilidad. Al emplear la media winsorizada, los coeficientes de variación se redujeron de forma consistente en todos los genotipos. Por ejemplo, en Valencia los coeficientes de variación disminuyeron de 35% y 50% a valores entre 19% y 28%; en Mineola de 41-43% a 34-38%; en C-35 de 44% a 24-28%; en Volkameriano de 52-63% a 24-30% y en Amblicarpa de 41-46% a 25-26%, considerando ambos ciclos de evaluación. Estos resultados concuerdan con lo señalado por Barnett y Lewis (1994), quienes destacan la utilidad de los estimadores robustos para controlar la variabilidad inducida por valores atípicos.

Los coeficientes de variación winsorizados del número de embriones por semilla (NE) se estimaron mediante bootstrap no paramétrico con reemplazo, utilizando 200 réplicas, y se presentan en los Cuadros 1 y 2 para los ciclos 2020 y 2021, respectivamente; este enfoque permitió cuantificar la variabilidad de NE sin asumir una distribución teórica, lo cual es apropiado para variables biológicas con comportamiento asimétrico (Efron y Tibshirani, 1986; Davison y Hinkley, 1997).

Cuadro 1. Coeficientes de variación winsorizados asociados a cada tamaño de muestra (n), con su valor F, en los genotipos Amblicarpa, Volkameriano, Valencia, C-35 y Mineola para el ciclo 2020.

n	Mandarina Amblicarpa		Limón Volkameriano		Naranja Valencia		Citrange C-35		Tangelo Mineola	
	CV _w ^B	Valor F	CV _w ^B	Valor F	CV _w ^B	Valor F	CV _w ^B	Valor F	CV _w ^B	Valor F
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

*= significativo ($p \leq \alpha = 0.05$); ns= no significativo.

Cuadro 2. Coeficientes de variación winsorizados asociados a cada tamaño de muestra (n), con su valor F, en los genotipos Amblicarpa, Volkameriano, Valencia, C-35 y Mineola para el ciclo 2021.

n	Mandarina Amblicarpa		Limón Volkameriano		Naranja Valencia		Citrange C-35		T. Mineola	
	CV _w ^B	Valor F	CV _w ^B	Valor F	CV _w ^B	Valor F	CV _w ^B	Valor F	CV _w ^B	Valor F
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

*= significativo ($p \leq \alpha = 0.05$); ns= no significativo.

La elección de 200 réplicas se justificó por la variación del coeficiente de variación entre genotipos y ciclos, y se encuentra dentro del rango recomendado para obtener estimaciones estables del error estándar y de los estimadores de varianza en contextos de alta variabilidad (Efron y Tibshirani, 1986; Ramírez *et al.*, 2013). En todos los genotipos se observó una disminución progresiva del coeficiente de variación a medida que aumentó el tamaño de muestra (Cuadro 1 y 2), con reducciones pronunciadas hasta aproximadamente 20 semillas y una tendencia asintótica a partir de tamaños de muestra cercanos a 60-70 semillas. Este comportamiento es característico de los ensayos de uniformidad empleados para definir tamaños óptimos de muestreo en estudios agronómicos (Lessman y Atkins, 1963; Paranaiba *et al.*, 2009).

La aplicación del (MMD) permitió identificar de manera objetiva el punto de curvatura máxima de la relación entre el coeficiente de variación winsorizado y el tamaño de muestra, evitando la subjetividad asociada a la inspección visual del método de curvatura máxima tradicional (Paranaiba *et al.*, 2009; Arellano-Durán *et al.*, 2018). Con base en este criterio, el tamaño de muestra óptimo estimado para los cinco genotipos, considerando ambos ciclos de cosecha, se ubicó en un rango de 26 a 28 semillas. En dichos tamaños de muestra, los coeficientes de variación oscilaron entre 6.01% y 11%, lo que indica una variabilidad controlada y una precisión adecuada en la estimación del NE.

En estudios previos que emplearon el método de curvatura máxima para determinar el tamaño de muestra en especies poliembriónicas, Santos *et al.* (2015) reportaron un rango de 10 a 23 semillas en 15 genotipos de cítricos, mientras que Arellano-Durán *et al.* (2018) recomendaron analizar ocho semillas poliembriónicas en cultivares de mango Manila y Ataulfo. No obstante, en ambos casos

no se especifica si los datos cumplieron con el supuesto de normalidad, condición necesaria para justificar el uso del coeficiente de variación como criterio de estimación.

Adicionalmente, la prueba F aplicada como criterio complementario para la validación del tamaño de muestra mostró que, en la mayoría de los genotipos, no se detectaron diferencias significativas entre varianzas consecutivas a partir de tamaños de muestra entre 30 y 50 semillas, dependiendo del genotipo y del ciclo de cosecha. En particular, en Valencia no se observaron diferencias significativas a partir de 30 semillas, en Volkameriano a partir de 50 semillas y en Amblicarpa, C-35 y Mineola a partir de 40 semillas. Estos resultados indican que incrementos adicionales en el tamaño de muestra no producen reducciones significativas en la variabilidad absoluta, en concordancia con lo señalado por Gómez y Gómez (1984).

En conjunto, el MMD permitió identificar un tamaño de muestra óptimo entre 26 y 28 semillas, mientras que la prueba F sugirió rangos ligeramente más amplios (30-50 semillas), en función del genotipo; esta convergencia entre criterios respalda la solidez del enfoque y ofrece un marco objetivo para seleccionar tamaños de muestra en estudios de poliembrionía en cítricos. Con base en ambos resultados y considerando la estabilidad observada en dos ciclos de cosecha, se recomienda analizar 27-40 semillas en Amblicarpa, 26-50 en Volkameriano, 26-30 en Valencia y 26-40 en C-35 y Mineola. En este sentido, el método propuesto representa una alternativa robusta para variables que no cumplen normalidad, ya que definir tamaños de muestra adecuados mejora la precisión de los estimadores, fortalece la confiabilidad de las inferencias y puede extrapolarse a otras especies y variables cuantitativas con distribuciones asimétricas, ampliando su utilidad en distintos contextos experimentales.

Conclusiones

Los resultados de este estudio muestran que la winsorización, combinada con un procedimiento bootstrap no paramétrico, constituye una estrategia eficaz para estimar el tamaño de muestra en estudios de poliembrionía en cítricos, particularmente cuando la variable número de embriones por semilla presenta asimetría y valores atípicos. La aplicación del método de máxima distancia, complementado con la prueba F, permitió definir de manera objetiva que un tamaño de muestra entre 26 y 50 semillas es suficiente para obtener estimaciones precisas y con variabilidad controlada en los cinco genotipos evaluados. En conjunto, la metodología propuesta contribuye a la estandarización del muestreo en semillas poliembriónicas de cítricos y ofrece un procedimiento replicable para otras especies y variables cuantitativas con distribuciones no normales.

Bibliografía

- 1 Andrade-Rodríguez, M.; Villegas-Monter, A. y García-Velázquez, A. 2003. Características morfológicas del fruto y poliembrionía 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. Poliembrionia 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 Mirosevich, 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. <https://dx.doi.org/10.5772/intechopen.105994>.

Determinación del tamaño de muestra en semillas poliembriónicas de cítricos mediante métodos robustos

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: Artículos

Palabras clave:

Palabras clave:

apomixis
bootstrap no paramétrico
cítricos
método de máxima distancia

Counts

Figures: 1
Tables: 2
Equations: 4
References: 20