

Determinación de la potencia estadística de experimentos de rendimiento en maíz

Jorge Claudio Vargas-Rojas^{1§}
Fernando García²

¹Universidad de Costa Rica-Sede Regional de Guanacaste. Barrió El Capulín, Liberia, Costa Rica. CP. 50101. ²Universidad Nacional de Córdoba-Facultad de Ciencias Económicas. Bv. Enrique Barros, Ciudad Universitaria, Córdoba, Argentina. CP. X5000HRV.

§Autor para correspondencia: jorgeclaudio.vargas@ucr.ac.cr.

Resumen

El análisis prospectivo de la potencia estadística de una prueba de hipótesis debería ser una de las etapas más importantes de cualquier experimento; sin embargo, se omite con frecuencia. En particular, para Costa Rica, no se encontraron investigaciones relacionadas con este tema para experimentos de rendimiento en el cultivo de maíz. El objetivo de este trabajo fue determinar la potencia estadística de un diseño completamente aleatorizado para experimentos de rendimiento en el cultivo de maíz (*Zea mays*) mediante la simulación de ensayos de uniformidad. Para realizar los cálculos de potencia se estimaron los parámetros del proceso de correlación espacial de un ensayo de uniformidad establecido en Santa Cruz, Costa Rica en el año 2018. Dichas estimaciones fueron utilizadas para realizar 10 000 simulaciones de campos aleatorios de mayor tamaño, lo que permitió superponer diferente número de repeticiones y estimar la potencia estadística para detectar una diferencia de 10% con respecto a la media en un experimento con un diseño completamente aleatorizado a un nivel de significación de 5%. Se obtuvo la potencia 80% con ocho repeticiones y se concluye que, bajo las condiciones experimentales de este trabajo, en ensayos de rendimiento en el cultivo de maíz, para detectar una diferencia de medias 10% a un nivel de significación 5%, se deben usar ocho o más repeticiones.

Palabras clave: campos aleatorios, ensayo blanco, número de repeticiones, potencia de prueba, simulaciones geoestadísticas.

Recibido: marzo de 2022

Aceptado: junio de 2022

Introducción

En la planificación de un experimento agrícola, con frecuencia se busca comparar tratamientos (genotipos de un cultivar, dosis de fertilizante, densidad de siembra, formulaciones de un herbicida, entre otros). Para esto, dichos tratamientos son aplicados a unidades experimentales, dispuestas según algún diseño específico, donde se registran una o varias variables de respuesta. Después, con el uso de técnicas de análisis estadísticas, se estiman las medias de la variable de respuesta para cada tratamiento y, comúnmente, se prueban hipótesis sobre estas (Robledo, 2015).

En una prueba de hipótesis se pueden cometer dos tipos de errores. El primer tipo, denominado error de tipo I, corresponde a tomar la decisión de rechazar la hipótesis nula, cuando esta es verdadera. El segundo tipo, denominado error de tipo II, corresponde a tomar la decisión de no rechazar la hipótesis nula, cuando en realidad la hipótesis alternativa es verdadera (Montgomery, 2019). La potencia estadística de una prueba de hipótesis se define como la probabilidad de rechazar la hipótesis nula cuando esta es falsa, esto es, la probabilidad de encontrar diferencias cuando realmente existen.

El estudio de la potencia estadística se basa en estimar la probabilidad de cometer error tipo II, se desea que esta sea pequeña de manera que su complemento (potencia estadística) sea lo más alto posible (Lapeña *et al.*, 2011). En otras palabras, la potencia es la probabilidad de que un efecto de tamaño dado se pueda distinguir de la variación aleatoria intrínseca de la variable (Gent *et al.*, 2018). Por convención, un nivel de potencia aceptable se ha establecido en 80% (Cohen, 1988).

La función potencia depende de cuatro elementos que se relacionan entre sí: 1) nivel de significación; 2) tamaño del efecto; 3) variabilidad; y 4) número de repeticiones. El nivel de significación, simbolizado por (α), es la probabilidad de rechazar H_0 cuando esta es verdadera. El tamaño del efecto es la diferencia que se espera encontrar entre un par de tratamientos y se establece en función de criterios biológicos, físicos, económicos, científicos o prácticos (Kuehl, 2001), de una manera más clara, es la mínima diferencia considerada significativa que se desea detectar entre tratamientos. La variabilidad es la varianza residual del experimento. El número de repeticiones es la cantidad de unidades experimentales independientes por tratamiento que se requieren para alcanzar cierto nivel de potencia (Cohen, 1992).

Existe la problemática de que los trabajos de potencia estadística que se encuentran en la bibliografía asumen algunos valores para los parámetros de los modelos en cuestión. Por ejemplo, para un modelo lineal clásico, el valor de la varianza de la diferencia de medias se suele asumir como conocida, porque es tomada de otros trabajos o de experiencias previas; sin embargo, estas consideraciones pueden no representar bien las condiciones reales del experimento.

Adicionalmente, pocas revistas publican medidas de variación útiles para el cálculo de la potencia, por ello la información necesaria para diseñar el tamaño del ensayo no está disponible o es poco confiable (Stroup, 2002); no obstante, dicha información puede obtenerse con técnicas de simulación (Lantuéjoul, 2002). Actualmente, en el área de la geoestadística, las técnicas de simulación o generación de realizaciones de campos aleatorios son frecuentes, ya que desempeñan un papel importante como una herramienta que permite obtener información o realizar inferencia cuando los resultados analíticos requeridos son difíciles de conseguir en la práctica (Diggle y Ribeiro, 2010).

El propósito de las simulaciones geoestadísticas es reproducir la variabilidad espacial inherente de la variable regionalizada. La simulación reproduce el valor y la variabilidad esperados del modelo en términos del variograma, que está parametrizado por el ‘nugget’, el ‘sill’ y el ‘range’. Luego, es posible construir muchas realizaciones a partir del mismo modelo; cada realización será diferente, pero tendrá el mismo valor esperado y la misma estructura de covarianza (Petitgas *et al.*, 2017). De manera que, con simulaciones de un ensayo de uniformidad es posible estudiar el efecto de distintos diseños y variantes de estos sobre la potencia estadística (Richter y Kroschewski, 2012).

El cálculo de la potencia estadística no forma parte usual en la planificación de la experimentación agrícola (González-Lutz, 2008). Así, no se conoce si utilizan un número de repeticiones que permitan alcanzar la potencia deseada, lo que puede llevar a que las conclusiones basadas en las pruebas de hipótesis de algunos trabajos no sean confiables, particularmente, cuando no se encuentran diferencias significativas.

Es común que los investigadores, acudan a números arbitrarios o tradicionales sin justificación estadística ni práctica para definir las repeticiones que debe tener un ensayo. La mayoría de los trabajos en el campo agrícola costarricense se centran en pruebas paramétricas como el análisis de varianza o regresión y es nulo el abordaje que se le ha dado a la potencia desde un enfoque prospectivo. En la literatura consultada, solo se encontró el trabajo de Vargas-Rojas (2021), quien estimó en número de repeticiones para ensayos de rendimiento en el cultivo del arroz.

No se puede estimar el número de repeticiones adecuado para un ensayo, sin establecer el tamaño del efecto, la varianza, el nivel de significación y la estructura de tratamientos. Estos elementos pueden variar de experimento a experimento según sean los objetivos y las condiciones que tendrá la investigación, por lo que se hace necesario el estudio de la potencia estadística para experimentos agrícolas. El objetivo de este trabajo fue determinar la potencia estadística de un diseño completamente aleatorizado para experimentos de rendimiento en el cultivo de maíz (*Zea mays*) mediante la simulación de ensayos de uniformidad en la zona de Santa Cruz, Costa Rica.

Materiales y métodos

Generalidades y condiciones del experimento

Para realizar las simulaciones se utilizaron datos de un ensayo de uniformidad que se llevó a cabo durante los meses de junio a septiembre del año 2016, en la Finca Experimental de Santa Cruz (FESC), propiedad de la Universidad de Costa Rica (10° 17' 6.24" latitud norte y 85° 35' 42.95" longitud oeste), la cual se ubica en el cantón de Santa Cruz, distrito de Santa Cruz, provincia de Guanacaste, a 54 msnm. La FESC posee una precipitación promedio de 1 834 mm año⁻¹ con estación seca de diciembre a abril y con estación lluviosa de mayo a noviembre (Cerdas, 2015), presenta temperatura media anual de 27.9 °C, con evaporación media diaria de 6.8 mm y radiación solar global diaria de 18.7 MJ (Instituto Meteorológico Nacional, 2011). Los suelos predominantes tienen fertilidad de media a baja, son arcillosos, con características de vertisoles por acillas 2:1 expandibles, se clasifican taxonómicamente como Vertic Haplustalfs coligados a Typic Haplusterts y Typic Ustorthents (Vega y Salas, 2012).

Se utilizó semilla de maíz blanco del híbrido HS5G que es de alto rendimiento y baja variabilidad en su producción (CV= 6.2%) (Cerritos *et al.*, 1994). La siembra se llevó a cabo de forma manual, en surcos separados por 1 m y con una distancia entre plantas de 0.25 m, para una densidad de 40 000 plantas ha⁻¹. La parcela seleccionada presentaba topografía plana, sin ningún tipo de factor que pudiera ocasionar variabilidad sistemática.

Se empleó la técnica del ensayo de uniformidad descrita por Rodríguez *et al.* (1993). De acuerdo con este método, se sembró una parcela de maíz de 26 m × 26 m (676 m²), en la cual se dejó tres metros de borde alrededor de todo su perímetro; así se obtuvo un área de 20 m × 20 m (400 m²) para ejecutar el ensayo de uniformidad. Todas las unidades básicas se indexaron según un sistema de coordenadas cartesianas. Luego, a los 110 días después de siembra, se cosechó el elote entero (producción en gramos) perteneciente a cada unidad básica.

Variación espacial

Para estimar los parámetros de correlación espacial ('nugget', 'sill' y 'range') del ensayo de uniformidad a los datos se ajustó dos de los modelos de correlación espacial más utilizados: exponencial y esférico (Cressie, 1993; Bivand *et al.*, 2013), estos dos modelos asumen que las observaciones que están cerca tienen mayor probabilidad de tener magnitudes similares y modelan esta estructura espacial con funciones de distancia (Fortin *et al.*, 2016). También se ajustó el modelo de errores independientes, este supone que la distancia no afecta la similitud entre las observaciones.

En el caso de los modelos con correlación también se modeló una estructura de medias que incluyera el efecto fijo como tendencia de primer y segundo orden de las coordenadas cartesianas con el fin de descontar, en caso de que existiera, tendencias a gran escala. Por tanto, se ajustaron, por máxima verosimilitud restringida (REML), los modelos que se presentan en el (Cuadro 1).

Cuadro 1. Modelos ajustados al ensayo de uniformidad de maíz (*Zea mays*). Santa Cruz, Costa Rica 2018.

Modelo	Estructura de correlación	Efecto nugget	Tendencia de primer orden	Tendencia de segundo orden
0	no	no	no	no
1	Esférica	si	no	no
2	Esférica	si	si	si
3	Esférica	si	no	si
4	Esférica	no	no	no
5	Esférica	no	si	no
6	Esférica	no	no	si
7	Exponencial	si	no	no
8	Exponencial	si	si	si
9	Exponencial	si	no	si
10	Exponencial	no	no	no
11	Exponencial	no	si	no
12	Exponencial	no	no	si

Luego, para la comparación entre los modelos ajustados se usó el criterio de información de *Akaike* (AIC) por sus siglas en inglés y el criterio de información bayesiano (BIC). Valores menores de AIC o BIC indican mejor ajuste del modelo estadístico (West *et al.*, 2015). Para comparar los modelos con y sin efecto fijo de las coordenadas (diferente estructura de medias, pero igual covarianza) se utilizó la prueba de cociente de verosimilitud (LRT) por sus siglas en inglés, basada en estimaciones de máxima verosimilitud (ML) por sus siglas en inglés. Finalmente, se utilizó los valores estimados para el ‘nugget’, ‘sill’ y ‘range’ del modelo, que, en términos comparativos, presentó un mejor ajuste para realizar la simulación de los respectivos campos aleatorios.

Simulación de campos aleatorios

El ensayo de uniformidad de este trabajo, debido al tamaño, tiene limitaciones para superponer un conjunto de unidades experimentales para distinto número de repeticiones del experimento. Entonces, se utilizaron los parámetros de correlación espacial estimados en el ensayo de uniformidad para simular parcelas de mayor extensión (campos aleatorios), de manera que pudiera lograr mayor flexibilidad para superponer más repeticiones en el diseño del experimento.

Se utilizó el paquete geoR (Ribeiro y Diggle, 2001) del lenguaje R (R Core Team, 2017) para simular 10 000 campos aleatorios, cada uno correspondió a una grilla de 160 m × 20 m, con un espacio entre puntos de 1 m × 1 m. Cada punto de la grilla fue una realización de la variable aleatoria, en este caso rendimiento en gramos, parametrizada con los parámetros estimados del ensayo de uniformidad.

En cada uno de los 10 000 campos aleatorios se conformaron, mediante la agregación de realizaciones de cada punto en la grilla, unidades experimentales de dimensiones 8 m de largo por 4 m de ancho (32 m²), tamaño definido por (Vargas-Rojas y Navarro-Flores, 2020), lo que resultó en la formación de 10 000 conjuntos de 100 unidades experimentales, dispuestas en cinco columnas y 20 filas. Dentro de cada conjunto, todas las unidades experimentales fueron indexadas con coordenadas cartesianas que definieron la ubicación espacial de su centroide. Conformadas estas, se estimó el promedio de rendimiento en kilogramos por unidad experimental y se evaluó la presencia o no de correlación espacial.

Potencia estadística

Estimación de componentes de varianza

Al tener un conjunto de 100 unidades experimentales, dispuestas en 20 filas y cinco columnas fue posible recrear distinto número de repeticiones para el DCA. Por ejemplo, para aleatorizar cinco tratamientos con dos repeticiones cada uno, sobre las 100 unidades experimentales, se consideraron las columnas como fijas, entonces se tomaron dos filas, así se obtuvieron diez unidades experimentales para aleatorizar los tratamientos.

Cada vez que se requería una repetición, se tomaba otra fila del conjunto de cien unidades experimentales. En cada uno de los 10 000 campos aleatorios simulados, se recreó distinto número de repeticiones (de dos hasta 20) y se estimaron los componentes de varianza, en este caso solo la varianza residual, para los distintos planes. El procedimiento se realizó con el paquete nlme (Pinheiro *et al.*, 2016) del lenguaje R (R Core Team, 2017). Una vez estimados los componentes

de varianza, los cálculos de la potencia estadística para la prueba de hipótesis de diferencia de medias se hicieron bajo el enfoque del modelo lineal clásico, descrito a continuación. Se siguió la notación propuesta por Stroup (2002).

Modelo lineal clásico

El modelo lineal clásico se define de la siguiente manera: $y = X\beta + e$ (1). Donde: y es el vector de observaciones, de dimensión $n \times 1$; X es la matriz de diseño, de dimensión $n \times p$; β es el vector de parámetros, de dimensión $p \times 1$; e es el vector de errores aleatorios, de dimensión $n \times 1$. Bajo este enfoque la hipótesis nula (H_0) se define de la forma $K'\beta = 0$, donde $K'\beta$ es una función estimable. Esta hipótesis puede ser probada con el uso del estadístico F , definido a

$$\text{continuación: } F = \frac{(K'\hat{\beta})' [K'(X'X)^{-1}K]^{-1} (K'\hat{\beta}) / \text{rango}(K)}{\hat{\sigma}^2 / (n - \text{rango}(X))} \quad (2).$$

Este estadístico tiene una distribución $F_{[\text{rango}(K), v, \lambda]}$, donde el rango de K son los grados de libertad del numerador, $n - \text{rango}(X)$ son los grados de libertad del denominador (v) y λ es el parámetro de

no centralidad, que tiene la siguiente expresión: $\lambda = \frac{(K'\beta)' [K'(X'X)^{-1}K]^{-1} (K'\beta)}{\sigma^2}$ (3). Bajo H_0 , λ es igual a 0; cuando H_0 es falsa λ tomará valores mayores a 0. Enseguida, la potencia quedará determinada como $P\{F_{[\text{rango}(k), v, \lambda]} > F_{[\text{crítica}]}\}$.

Para la estimación de la potencia, primero, se calculó el valor crítico ($F_{[\text{crítica}]}$). Luego, se estimó λ , definido según la estructura del diseño que determinó la matriz X ; el vector de medias de tratamientos que determinó β ; el vector de contrastes que definió K y la varianza residual estimada. Una vez obtenidos 2 y 3 se estimó la potencia como: $P\{F_{[\text{rango}(k), v, \lambda]} > F_{[\text{crítica}]}\}$.

Cada estimación de potencia se hizo desde dos hasta 20 repeticiones en cada uno de los 10 000 conjuntos de unidades experimentales, con los componentes de varianza previamente estimados. En el parámetro de no centralidad definido en (3) se debió especificar en la matriz K , los contrastes de interés y en el vector de parámetros β las medias de tratamientos. Lo anterior, resulta operativamente difícil ya que existen infinitos contrastes posibles. Para facilitar el cálculo del parámetro de no centralidad, se restringió la matriz de contrastes K , a un único contraste que consistió en la comparación de un par (arbitrario) de medias, según sugiere Kuehl (2001). Como resultado, la ecuación (3) sólo tuvo el contraste relacionado al par de medias seleccionado.

Así, si se reemplaza $K'\beta$ en el parámetro de no centralidad de la distribución F , se obtiene la distribución del estadístico de la prueba bajo hipótesis alternativa, de donde se puede calcular la potencia de la prueba para una diferencia entre pares de medias mayor o igual al supuesto para $K'\beta$, lo que genera una cota inferior para λ . El nivel de significancia (α) que se utilizó fue de 0.05. La diferencia mínima para detectar (tamaño del efecto) fue determinada mediante un consenso práctico-estadístico con Ingenieros Agrónomos especialistas en el cultivo en cuestión; se convino que una diferencia de 10% con respecto a la media general era la situación más frecuente a la que se enfrentan en la práctica de ensayos en Costa Rica. Debido a la metodología utilizada, las estimaciones de potencia son invariantes a la cantidad de tratamientos en cuestión, de forma arbitraria, se utilizó como referencia cinco tratamientos.

Por lo tanto, las matrices K y β quedaron definidas de la siguiente manera: $K = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$; $\beta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{bmatrix}$. Cada

elemento μ_i del vector β corresponde a la media de cada tratamiento. Donde: $\mu_i = \mu + \tau_i$; μ denota la media general y τ_i denota la diferencia entre la i -ésima media tratamiento y μ . Para este trabajo, μ sería la media general de la variable producción; no obstante, a esta se le asignó un valor de 0. Luego -arbitrariamente- τ_1 se definió como el efecto de un tratamiento inducido que generó un desvío del 0, o sea, el tamaño de efecto que quiere ser detectado.

La estimación de la potencia se hizo para cada uno de 10 000 conjuntos de unidades experimentales simuladas, por lo que la potencia y los estadísticos presentados corresponden a un promedio. Todos los procedimientos respectivos se hicieron con el lenguaje R (R Core Team, 2020).

Resultados

Inferencia

Los modelos de mejor ajuste fueron el 10, 11 y 12. Los parámetros de correlación espacial estimados junto con los criterios AIC y BIC para estos modelos se presentan en el Cuadro 2.

Cuadro 2. Parámetros estimados y criterios de bondad de ajuste para los modelos ajustados del ensayo de uniformidad de maíz (*Zea mays*). Santa Cruz, Costa Rica 2018.

Modelo	Estimación			Criterio de información	
	Nugget	Sill	Range	AIC	BIC
10	0	1.23 E-01	0.72	256.3	272.3
11	0	1.2 E-01	0.68	252.8	276.8
12	0	1.18 E-01	0.65	252.4	288.3

El modelo 12 fue el que presentó el menor AIC de todos los modelos ajustados. Con la finalidad de evaluar si se justificaba la estimación de más parámetros, este modelo se comparó con los modelos 10 y 11, los resultados se presentan en el Cuadro 3. La comparación con el modelo 11 no fue significativa ($p < 0.05$) por su parte, la comparación con el modelo 10 fue significativa ($p < 0.05$), lo que sugiere que el modelo 11 es el adecuado. A pesar de esto, debido que la comparación con el modelo 10 tuvo un valor p cercano al límite de decisión y como, se usó un vecindario pequeño para simular los ensayos de uniformidad, se asumió estacionalidad local. Por estas razones se consideró el modelo 10 como el adecuado, también porque fue el que presentó menor BIC.

Cuadro 3. Estadísticos y prueba de cociente de verosimilitud para los modelos ajustados con ML en el cultivo de maíz (*Zea mays*). Santa Cruz, Costa Rica 2018.

Modelo	Prueba	Referente	log L	AIC	BIC	gl prueba	-2log (L)	Valor p
10	10 vs 12	10	-124.61	257.21	273.18	5	11.67	0.04
11	11 vs 12	11	-118.77	254.92	278.87	3	5.38	0.14
12			-121.46	255.54	291.46			

Simulación de ensayos de uniformidad

Las simulaciones de los 10 000 ensayos de uniformidad de mayor tamaño se hicieron con base en los parámetros estimados con el modelo 10.

Unidades experimentales

La evaluación de la estructura de correlación espacial, una vez conformadas las unidades experimentales, arrojó que las unidades experimentales simuladas se pueden considerar independientes, por lo que la potencia estadística se estimó sin tomar en cuenta la correlación espacial.

Potencia estadística

Con ocho repeticiones se alcanzó una potencia 80%. El Cuadro 4 presenta los resultados de la estimación de la potencia y otros estadísticos para los distintos números de repeticiones.

Cuadro 4. Potencia estadística alcanzada para un número dado de repeticiones en un diseño completamente aleatorizado en ensayos de rendimiento con maíz (*Zea mays*). Santa Cruz, Costa Rica 2018.

Repeticiones	gl	F	EEDM	λ	Potencia	σ^2_{Res}
2	5	6.61	6.3 E-02	2.67	0.26	4.19 E-03
3	10	4.96	5.2 E-02	3.64	0.4	4.21 E-03
4	15	4.54	4.54 E-02	4.63	0.51	4.23 E-03
5	20	4.35	4.08 E-02	5.64	0.6	4.24 E-03
6	25	4.24	3.73 E-02	6.65	0.68	4.25 E-03
7	30	4.17	3.46 E-02	7.66	0.75	4.26 E-03
8	35	4.12	3.24 E-02	8.67	0.8	4.26 E-03
9	40	4.08	3.06 E-02	9.69	0.84	4.27 E-03
10	45	4.06	2.91 E-02	10.72	0.88	4.27 E-03
11	50	4.03	2.77 E-02	11.75	0.91	4.26 E-03
12	55	4.02	2.66E-02	12.76	0.93	4.27 E-03
13	60	4	2.55 E-02	13.78	0.94	4.27 E-03
14	65	3.99	2.46 E-02	14.79	0.96	4.27 E-03
15	70	3.98	2.38 E-02	15.82	0.97	4.27 E-03
16	75	3.97	2.3 E-02	16.84	0.98	4.27 E-03
17	80	3.96	2.24 E-02	17.87	0.98	4.27 E-03
18	85	3.95	2.17 E-02	18.89	0.99	4.27 E-03
19	90	3.95	2.12 E-02	19.92	0.99	4.27 E-03
20	95	3.94	2.06 E-02	20.93	0.99	4.28 E-03

gl= grados de libertad denominador; F= F crítica; EEDM= error estándar de la diferencia de medias; λ = parámetro de no centralidad; σ^2_{res} = varianza residual.

Discusión

Con base en los resultados obtenidos en esta investigación, si se quiere obtener una potencia estadística 80% para detectar una diferencia de medias 10% con respecto a la media general, a un nivel de significancia 5% en ensayos de rendimiento, se recomienda usar ocho repeticiones. Con menos de las repeticiones recomendadas la potencia del ensayo podría no llegar a ser suficiente. En este sentido Cohen (1992); Murphy *et al.* (2014) mencionan que, si la probabilidad de rechazar la hipótesis nula es la misma que no rechazarla, va a generar una inconsistencia entre los resultados de investigaciones; un ensayo con una potencia cercana o menor al 50% no debería ser realizado.

Por otro lado, Gent *et al.* (2018) aducen que 90% podría ser un nivel de potencia más apropiado, cuando los efectos del tratamiento son importantes, como es el caso de experimentos donde se evalúan los rendimientos de variedades promisorias. Vale la pena indicar que la recomendación realizada es aplicable en condiciones experimentales similares a las que tuvo el presente estudio y que esta puede variar. Cada ensayo específico debe ser planificado de una forma adecuada. Sin embargo, la información que se generó puede ser utilizada como base para otros experimentos y que de otra manera sería muy difícil de conseguir.

Por ejemplo, si con base en experimentos anteriores o estudios piloto se tiene información de la variabilidad espacial de un sitio, se puede comparar las estimaciones de ‘sill’ y ‘range’ con los valores estimados Cuadro 2 y en función de la comparación, adecuar el número de repeticiones para ese sitio en particular. La estimación del parámetro ‘sill’ es una estimación de la varianza residual de la variable regionalizada (Diggle y Ribeiro, 2010; Guedes *et al.*, 2020); por tanto, si es menor se podría disminuir el número de repeticiones recomendadas en este trabajo.

Así mismo sucede con el ‘range’, Lapeña *et al.* (2011) investigaron, a través de simulaciones, factores que afectan la potencia estadística en presencia de correlación espacial y determinaron que un aumento en la dependencia espacial de los datos, reflejada en el ‘range’, aumenta la varianza de la distribución del estadístico de prueba lo que reduce la potencia. Entonces, si se tiene un ‘range’ menor se podría usar menos repeticiones, caso contrario, no. Por otra parte, si se van a evaluar hídricos con menor variabilidad en el rendimiento que el reportado para el híbrido HS5G se podría evaluar disminuir el número de repeticiones, en caso de materiales con mayor variabilidad, disminuir el número de repeticiones no sería una opción.

Los resultados que se presentan en este trabajo ponen de manifiesto la importancia de tomar en cuenta la potencia estadística cuando se realizan pruebas de hipótesis. La consideración formal de la potencia estadística de un experimento debe, por lo tanto, ser un componente rutinario e indispensable del diseño experimental antes de la recopilación y análisis de datos. Otra motivación para llevar a cabo un análisis de potencia en las primeras etapas de planificación de un experimento es garantizar que haya recursos disponibles que aseguren que se puede concretar los objetivos propuestos. Conclusiones erróneas podrán ser evitadas sólo si los experimentos, mediante un análisis prospectivo, aseguran una potencia adecuada (Gent *et al.*, 2018).

Realizar más trabajos sobre esta temática resultaría beneficioso para el sector agrícola costarricense, conocer a fondo las condiciones en que se va a ejecutar un experimento permitirá diseñarlo adecuadamente. Es preciso que con esta información pueda existir un diálogo entre los

desarrolladores de ensayos y los estadísticos, de manera que los primeros comuniquen sus necesidades y los segundos las puedan trasladar a los términos estadísticos correspondientes, esto ayuda a dejar claro los objetivos y el verdadero alcance de una investigación.

Como mencionan Quinn y Keough (2002), uno de los aspectos más valiosos del análisis de potencia *a priori* es que, para hacer los cálculos, se debe especificar la hipótesis alternativa (esto es tamaño del efecto) y, lo más importante, el modelo estadístico que se aplicará a los datos. Especificar el modelo hace pensar en el análisis antes de recopilar los datos, un hábito recomendable.

Conclusiones

En el marco de las condiciones que fue realizado este trabajo se considera que: Si se quiere detectar una diferencia de medias 10% a un nivel de significación 5%, no se recomienda usar menos de ocho repeticiones para ensayos de rendimiento en maíz. Menos repeticiones podrían ser plausibles en función de las condiciones del experimento. Este trabajo brinda información poco disponible que puede ser tomada como base para planificar futuros trabajos.

Literatura citada

- Bivand, R. S., Pebesma, E. and Gómez-Rubio, V. 2013. Applied spatial data analysis with R. 2nd (Ed.). Springer New York. <https://doi.org/10.1007/978-1-4614-7618-4>. 405 p.
- Cerritos, G.; Gómez, F. y Palma, A. 1994. Lote demostrativo fact: introducción de una nueva metodología para evaluar híbridos de maíz en fincas de agricultores. Informe Anual de Investigación, 7(1):76-79. [https://bdigital.zamorano.edu/bitstream/11036/2455/1/206105-0167 - Copy.pdf](https://bdigital.zamorano.edu/bitstream/11036/2455/1/206105-0167-Copy.pdf).
- Cohen, J. 1988. Statistical power analysis for the behavioral sciences. 2nd (Ed.). Routledge. <https://doi.org/10.4324/9780203771587>. 1-17 pp.
- Cohen, J. 1992. A power primer. Psychological Bulletin. 112(1):155-159. <https://doi.org/10.1037/0033-2909.112.1.155>.
- Cressie, N. A. C. 1993. Statistics for spatial data. 2nd (Ed.). John Wiley y sons, Inc. <https://doi.org/10.1002/9781119115151>. 29-105 pp.
- Diggle, P. J. and Ribeiro, P. J. 2010. Model-based geostatistics. 1st (Ed.). Springer New York. <https://doi.org/10.1007/978-0-387-48536-2>. 227 p.
- Gent, D. H.; Esker, P. D. and Kriss, A. B. 2018. Statistical power in plant pathology research. Phytopathology. 108(1):15-22. <https://doi.org/10.1094/PHYTO-03-17-0098-LE>.
- González-Lutz, M. I. 2008. Potencia de prueba: la gran ausente en muchos trabajos científicos. Agron. Mesoam. 19(2):309-313.
- Guedes, L. P. C.; Bach, R. T. and Uribe-Opazo, M. A. 2020. Nugget effect influence on spatial variability of agricultural data. Engenharia Agrícola. 40(1):96-104. <https://doi.org/10.1590/1809-4430-ENG.AGRIC.V40N1P96-104/2020>.
- Kuehl, R. 2001. Diseño de experimentos: principios estadísticos de diseño y análisis de investigación. 2nd (Ed.). International Thomson. 1-66 pp.
- Lantuéjoul, C. 2002. Geostatistical simulation: models and algorithms. 1st (Ed.). Springer-Verlag. <https://doi.org/10.1007/978-3-662-04808-5>. 1-17 pp.

- Lapeña, B. P.; Wijnberg, K. M.; Stein, A. and Hulscher, S. J. M. H. 2011. Spatial factors affecting statistical power in testing marine fauna displacement. *Ecological Applications*. 21(7):2756-2769. <https://doi.org/10.1890/10-1887.1>.
- Montgomery, D. 2019. *Design and analysis of experiments*. 10nd (Ed.). John Wiley y Sons. 1- 125 pp.
- Murphy, K. R., Myors, B. y Wolach, A. H. 2014. *Statistical power analysis: a simple and general model for traditional and modern hypothesis tests* 4th (Ed.). Routledge. 229 p.
- Petitgas, P.; Woillez, M.; Rivoirard, J.; Renard, D. and Bez, N. 2017. Handbook of geostatistics in R for fisheries and marine ecology. *In: ICES cooperative research report*. Issue 338. <https://doi.org/10.17895/ices.98-107> pp.
- Pinheiro, J.; Bates, D.; DebRoy, S. and Sarkar, D. 2016. *Nlme: linear and nonlinear mixed effects models*. <http://cran.r-project.org/package=nlme>. 338 p.
- Quinn, G. P. and Keough, M. J. 2002. *Experimental design and data analysis for biologists*. 1st (Ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511806384>. 155-172 pp.
- R Core Team. 2020. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ribeiro, P. J. and Diggle, P. J. 2001. Geor: a package for geostatistical analysis. *R-News*. 1(2):15-18. <https://doi.org/10.1159/000323281>.
- Richter, C. and Kroschewski, B. 2012. Geostatistical models in agricultural field experiments: investigations based on uniformity trials. *Agron. J.* 104(1):91-105. <https://doi.org/10.2134/agronj2011.0100>.
- Robledo, W. 2015. *Diseño y análisis de experimentos a un criterio de clasificación. Estadística y biometría: ilustraciones del uso de Infostat en problemas de agronomía*. 2nd (Ed.). Editorial Brujas. 257-285 pp.
- Stroup, W. 2002. Power analysis based on spatial effects mixed models: a tool for comparing design and analysis strategies in the presence of spatial variability. *J. Agric. Biol. Environ. Statistics*. 7(4):491–511. <https://doi.org/10.1198/108571102780>.
- Vargas-Rojas, J. C. 2021. Simulación de ensayos en blanco para determinar la potencia estadística de de experimentos en arroz. *Agron. Mesoam.* 32(1):196-208. <https://doi.org/10.15517/am.v32i1.40870>.
- Vargas-Rojas, J. C. y Navarro-Flores, J. R. 2020. Determinación del tamaño y la forma de unidad experimental, con el método de regresión múltiple, para ensayos de rendimiento de maíz (*Zea mays*), guanacaste, Costa Rica. *InterSedes*. 21(43):1-10. <https://doi.org/10.15517/isucr.v21i43.41972>.
- West, B. T., Welch, K. B., y Galecki, A. T. 2015. *Linear mixed models: a practical guide using statistical Software*. 2nd (Ed.). CRC Press. 38-41 pp.