Agronomía Mesoamericana

UNIVERSIDAD DE COSTA RICA

Artículo científico
Volumen 32(1):196-208. Enero-abril, 2021
e-ISSN 2215-3608, doi:10.15517/am.v32i1.40870

Simulación de ensayos en blanco para determinar la potencia estadística de experimentos en arroz[1]

Uniformity trials simulation to determine the statistical power in yield rice trials

Jorge Claudio Vargas-Rojas[2]

[1] Recepción: 3 de marzo, 2020. Aceptación: 3 de setiembre, 2020. Este trabajo formó parte de la tesis de maestría en Estadística Aplicada del autor. La totalidad de este proyecto fue financiado por Hacienda Mojica, Bagaces, Guanacaste, Costa Rica.

[2] Universidad de Costa Rica, Sede Regional de Guanacaste. Liberia, Costa Rica. jorgeclaudio.vargas@ucr.ac.cr (autor para la correspondencia; https://orcid.org/0000-0002-1139-2148).

Resumen

Introducción. 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 Costa Rica, dentro de la bibliografía consultada, no se encontraron investigaciones relacionadas con este tema para experimentos de rendimiento en el cultivo de arroz. **Objetivo.** Simular ensayos en blanco para determinar la potencia estadística de un diseño completamente aleatorizado para experimentos de rendimiento de arroz en Bagaces, Costa Rica. **Materiales y métodos.** Se estimaron los parámetros del proceso de correlación espacial de un ensayo en blanco establecido en Bagaces, Costa Rica. Luego, las estimaciones se utilizaron para realizar 10 000 simulaciones de campos aleatorios de mayor tamaño, lo que permitió superponer diferente número de repeticiones y estimar la potencia lograda para detectar una diferencia del 10 % con respecto a la media en un experimento completamente aleatorizado a un nivel de significación del 5 %. **Resultados.** La potencia del 80 % se obtuvo con cinco repeticiones. **Conclusión.** En ensayos de rendimiento en arroz, para detectar una diferencia de medias del 10 % a un nivel de significación del 5 %, en esta investigación se requirió de cinco o más repeticiones.

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

Abstract

Introduction. Prospective analysis of the statistical power of a hypothesis test should be one of the most important stages in any experiment, however, it is frequently omitted. In Costa Rica, within the literature consulted, no research related to this topic was found for yield experiments in rice cultivation. **Objective.** To simulate uniformity trials to determine the power of a completely randomized design for rice yield experiments in Bagaces, Costa Rica. **Materials and methods.** The parameters of the spatial correlation process of a blank trial established in Bagaces, Guanacaste were estimated. Then, the estimates were used to perform 10 000 simulations of larger random fields, this allowed to superimpose different number of repetitions and estimate the power achieved to detect a difference of 10 % with respect to the mean in a completely randomized experiment at a significance level of 5 %. **Results.** The power

of 80 % was obtained with five repetitions. **Conclusion.** In rice yield trials, to detect a mean difference of 10 % at a significance level of 5 %, this investigation required five or more repetitions.

Keywords: test power, number of repetitions, geostatistical simulation, random fields.

Introducción

Cuando se planifica un experimento agrícola, frecuentemente, el objetivo principal es comparar tratamientos (genotipos de un cultivar, dosis de fertilizante, densidad de siembra, formulaciones de un herbicida, fecha de siembra, combinaciones de varios factores, etc.). Para esto, dichos tratamientos son aplicados a unidades experimentales, dispuestas según algún diseño de experimento, donde se registra una o varias variables de respuesta. Después, con el uso de metodologías estadísticas, se estiman las medias de la variable de respuesta para cada tratamiento y, comúnmente, se prueban hipótesis sobre estas (Robledo, 2015).

En el contexto de la mayoría de experimentos agrícolas, para comparar dos o más tratamientos, se recurre al análisis de varianza, cuya finalidad es contrastar hipótesis relacionadas con las medias de los tratamientos. El análisis de varianza se formula en términos de hipótesis nula (H_0) o de no efecto de tratamientos e hipótesis alternativa (H_A) que indica la existencia de efectos no nulos. Básicamente, es un procedimiento que utiliza los valores observados para decidir si hay suficiente evidencia para rechazar la hipótesis nula a favor de la hipótesis alternativa. Para tomar esta decisión en el enfoque del análisis de varianza se utiliza el estadístico F (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, se da cuando 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 este sea pequeño 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). Un nivel de potencia aceptable se ha establecido por convención en 80 % (Cohen, 1988).

Los estudios que tienen por objetivo determinar el número de repeticiones para una prueba de hipótesis utilizan el concepto de potencia de la prueba (Ahn et al., 2015). 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 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 que se desea detectar entre tratamientos. La variabilidad es la varianza residual del experimento. El número de repeticiones es la cantidad de unidades experimentales por tratamiento que se requieren para alcanzar cierto nivel de potencia (Cohen, 1992).

El nivel de significación establece el punto, para la distribución F, en el cual el área a su derecha es α. Si el tamaño del efecto es cero, el estadístico F tiene una distribución F central. Cuando el estadístico F es mayor que el valor crítico, H_0 se rechaza, lo que da lugar a una F no central. El área bajo la curva a la derecha del valor crítico,

bajo una F no central (cuyo parámetro de no centralidad se denomina lambda) es la probabilidad de rechazar H_0 cuando esta es falsa. Esto es, la potencia de prueba para un nivel de significación, un número de repeticiones tamaño de efecto deseado y una variabilidad dada (Gbur et al., 2012).

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 (Stroup, 1999). Por ello, la información necesaria para diseñar el tamaño del ensayo no está disponible o es poco confiable. Dicha información puede obtenerse con técnicas de simulación. 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, por lo que pueden verse como una realización de una variable aleatoria con un valor esperado y una estructura de covarianza. La simulación reproduce el valor esperado y la variabilidad esperada del modelo en términos del variograma.

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). Con base en los parámetros estimados del semivariograma de un ensayo en blanco se pueden realizar simulaciones de este proceso estocástico (Fagroud y Meirvenne, 2002). El ensayo en blanco es una parcela de extensión relativamente grande que es tratada en toda su superficie uniformemente en cuanto a fertilización, aplicación de agroquímicos y demás labores de cultivo y que a la hora de la cosecha se subdivide en parcelas pequeñas (unidades básicas) (Rosselló y Fernández, 1986). Así, con simulaciones de un ensayo en blanco 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 en la experimentación agrícola (González, 2008). 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. 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 muchos trabajos no sean confiables, particularmente cuando no se encuentran diferencias significativas. Además, la potencia no se puede aplicar para todas las condiciones por igual. Algunos protocolos de ensayos para evaluar variedades a nivel de campo establecen que el mismo debe tener tres o más repeticiones para asegurar la confiabilidad estadística del ensayo, que miden con los grados de libertad del error. No obstante, hacer este tipo de recomendaciones generales no es adecuado y la estimación de la potencia no es algo trivial, como menciona Stroup (1999), no es suficiente con plantear un análisis de varianza, enumerar las fuentes de variación o decir se usará una prueba de separación de medias; no se puede indicar 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 en la que tendrá lugar la investigación, por lo que se hace necesario el estudio de la potencia estadística para experimentos agrícolas. El objetivo de este trabajo fue simular ensayos en blanco para determinar la potencia estadística de un diseño completamente aleatorizado para experimentos de rendimiento de arroz (*Oryza sativa*).

Agron. Mesoam. 32(1):196-208, enero-abril, 2021
ISSN 2215-3608 doi:10.15517/am.v32i1.40870

Materiales y métodos

Generalidades

El ensayo se llevó a cabo durante los meses de junio a noviembre del año 2012, en la Hacienda Mojica, situada en el cantón de Bagaces, distrito de la provincia de Guanacaste, Costa Rica. La cual se encuentra a 80 msnm, con una precipitación que varía entre 1500 a 2500 mm anuales y temperatura promedio anual de 29 ˚C.

La siembra se realizó con semilla de arroz de la variedad Palmar 18 (Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria, 2008), por medio de siembra directa, con una sembradora mecánica a chorro, en surcos separados 17,6 cm y una cantidad de semilla entre 100 y 115 kg ha^{-1}.

Se empleó la técnica del ensayo en blanco descrita por Rodríguez et al. (1993). De acuerdo con este método se seleccionó de la plantación comercial de la finca una parcela de 20 m × 20 m, por lo que la parcela total fue de 400 m^2. Alrededor de esta parcela se dejó una franja de dos metros de borde para todo el perímetro. Veinte días después de la siembra se diseñó una cuadrícula sobre la parcela; para esto se empleó estacas de bambú y cuerdas, de modo que se identificaron claramente las 400 microparcelas (unidades básicas), de 1 m^2 cada una. A cada unidad básica se le asignó coordenadas cartesianas, de manera que todas fueran ubicadas e identificadas en el terreno; ambas coordenadas estuvieron dadas por distancias en metros a ejes cartesianos (X fue el ancho e Y el largo de la parcela).

La cosecha (120 días después de siembra) se realizó por separado en cada una de las microparcelas. Se cortaron a nivel del suelo todas las plantas de arroz procedentes de cada unidad básica y se colocaron en un saco previamente identificado con el número correspondiente a la unidad básica cosechada, según el sistema de coordenadas cartesianas. Posteriormente, el grano y la paja se separaron y el grano se trasladó a bolsas de papel, igualmente identificadas. Los granos contenidos en cada bolsa se secaron al sol hasta que alcanzaron un promedio de humedad entre 13 % y 15 %; para obtener dicho promedio se midió el porcentaje de humedad a cuarenta bolsas seleccionadas al azar con un medidor de humedad Motomco®. Finalmente, se pesó cada una de las bolsas y se obtuvo la producción en gramos.

Variación espacial

Para estimar los parámetros de correlación espacial del ensayo en blanco, a los datos se les ajustó dos de los modelos de correlación espacial más utilizados: exponencial y esférico (Bivand et al., 2013; Cressie, 1993), además del modelo que supone errores independientes (Cuadro 1).

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, tendencia a gran escala. Por tanto, se ajustaron, por máxima verosimilitud restringida (REML), los siguientes modelos:

- **Modelo 0:** errores independientes,
- **Modelo 1:** correlación espacial esférica, con efecto *nugget* y sin tendencia.
- **Modelo 2:** correlación espacial esférica, con efecto *nugget* y con tendencia de primer orden.
- **Modelo 3**: correlación espacial esférica, con efecto *nugget* y con tendencia de segundo orden.
- **Modelo 4:** correlación espacial esférica, sin efecto *nugget* y sin tendencia.
- **Modelo 5:** correlación espacial esférica, sin efecto *nugget* y con tendencia de primer orden.
- **Modelo 6:** correlación espacial esférica, sin efecto *nugget* y con tendencia de segundo orden.
- **Modelo 7:** correlación espacial exponencial, con efecto *nugget* y sin tendencia.
- **Modelo 8:** correlación espacial exponencial, con efecto *nugget* y con tendencia de primer orden.

Cuadro 1. Modelos teóricos para funciones de semivariograma: errores independientes, esférico y exponencial.

Table 1. Theoretical models for semivariogram functions: independent errors, spherical, and exponential.

Modelo	Parametrización
Errores independientes	$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 & h > 0 \end{cases}$
Correlación espacial esférica*	$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C\left\{\frac{3h}{2R} - \frac{1}{2}\left(\frac{\lvert h\rvert}{R}\right)^3\right\} & 0 < h < R \\ C_0 + C & h \geq R \end{cases}$
Correlación espacial exponencia*	$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C\left\{1 - exp\left\{-\frac{\lvert h\rvert}{R}\right\}\right\} & h \neq 0 \end{cases}$

* Fuente/Source: Bivand et al. (2013); Cressie (1993).

- **Modelo 9:** correlación espacial exponencial, con efecto *nugget* y con tendencia de segundo orden.
- **Modelo 10:** correlación espacial exponencial, sin efecto *nugget* y sin tendencia.
- **Modelo 11:** correlación espacial exponencial, sin efecto *nugget* y con tendencia de primer orden.
- **Modelo 12:** correlación espacial exponencial, sin efecto *nugget* y con tendencia de segundo orden.

Para la comparación entre los modelos ajustados (Cuadro 1) se usó el criterio de información de *Akaike* (AIC, por sus siglas en inglés), el cual se basa en el logaritmo de la función de verosimilitud más una penalización según el número de parámetros estimados en el modelo de estructura de covarianza; y el criterio de información bayesiano (BIC, por sus siglas en inglés), al igual que el AIC también se basa en el logaritmo de la verosimilitud, pero cobra una penalización mayor para el número de parámetros estimados. 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). Se utilizaron 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 en blanco de este trabajo, debido al tamaño, tuvo limitaciones para superponer un conjunto de unidades experimentales para distintos planes de diseño del experimento. Con plan del experimento se hace referencia a distinto número de repeticiones para el diseño completamente aleatorizado (DCA). Se utilizaron los parámetros obtenidos en la estimación de la correlación espacial y la media estimada de los ensayos en blanco para simular lotes de mayor extensión (campos aleatorios), de manera que pudiera lograr mayor flexibilidad a la hora de superponer un conjunto de unidades experimentales para los diferentes planes de diseño.

El procedimiento utilizado se basó en simular realizaciones de la forma, $Y= (Y_1...Y_n)$, de un modelo gaussiano

Agron. Mesoam. 32(1):196-208, enero-abril, 2021
ISSN 2215-3608 doi:10.15517/am.v32i1.40870

sobre una grilla con *n* puntos x_i, donde $x_i \in \mathbb{R}^2$, este algoritmo está implementado en el paquete geoR (Ribeiro y Diggle, 2001), del lenguaje R (R Core Team, 2017), que se utilizó para realizar las simulaciones de este trabajo. Sobre este tipo de simulaciones (Diggle y Ribeiro, 2010).

Se simularon 10 000 campos aleatorios, cada uno correspondió a una grilla de 15 m × 160 m, con un espacio entre puntos de 1 m × 1 m. Cada punto de la grilla sería una realización de la variable aleatoria, en este caso rendimiento en kilogramos, parametrizada con los parámetros estimados del ensayo en blanco. En cada uno de los quinientos campos aleatorios se conformaron, mediante la agregación de realizaciones de cada punto en la grilla, unidades experimentales de dimensiones 3 m de largo por 8 m de ancho (24 m^2)*; lo que resultó en la formación de 10 000 conjuntos de cien unidades experimentales, dispuestas en cinco columnas y veinte filas.

Dentro de cada conjunto, todas las unidades experimentales se indexaron 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. Para esto, el promedio de producción de las unidades experimentales, en cada uno de los 10 000 conjuntos, se modeló con dos estructuras distintas de covarianza del modelo descrito en la ecuación 1.

$$y_i = \mu + e_i \qquad (1)$$

Donde:
y_i: es la respuesta del i-ésima unidad experimental.
μ: media general.
e_i: error aleatorio.

Las estructuras de covarianza para el modelo descrito en (1) fueron las siguientes:

1. $cov(e_i, e_j) = 0$
2. $cov(e_i, e_j) = \sigma^2_{res}\left\{1 - \left(\frac{3h}{2R}\right) + \frac{1}{2}\left(\frac{|h|}{R}\right)^3\right\}$

A la opción 1 se le llamó modelo base (MB) y supone independencia de los errores. A la opción 2 se le llamó modelo base con correlación esférica (MBCEsf), en esta última los errores se suponen correlacionados. Para la comparación entre MB y MBCEsf se usaron el AIC y la prueba de cociente de verosimilitud. Al sólo diferir en la estructura de covarianza, el ajuste de MB y MBCEsf se hizo con máxima verosimilitud restringida (REML, por sus siglas en inglés).

Potencia estadística

Estimación de componentes de varianza

Al tener un conjunto de cien unidades experimentales, dispuestas en veinte 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 cien, cincuenta unidades experimentales se consideraron las columnas como fijas, entonces se tomaron dos filas, así se obtuvieron diez unidades experimentales para aleatorizar los tratamientos.

* Tamaño definido por Vargas y Navarro (2019).

ISSN 2215-3608 doi:10.15517/am.v32i1.40870

Cada vez que se requirió una repetición se tomó 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 veinte) y se estimaron los componentes de varianza (en este caso solo la varianza residual) para los distintos planes, necesarios para el cómputo de la potencia estadística. Esto se realizó con el paquete nlme (Pinheiro et al., 2016) del lenguaje R (R Core Team, 2017). Para estimar la varianza residual se utilizó el modelo descrito en (1).

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 puede expresar de la siguiente manera (2):

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e} \qquad (2)$$

Donde:
y = es un vector de observaciones, de dimensión n × 1.
X = es la matriz de diseño, de dimensión n × p.
β = es un vector de parámetros, de dimensión p × 1.
e = error aleatorio, de dimensión n × 1.

Bajo este enfoque la hipótesis nula (H_0) se define de la forma K'β = 0, donde K'β es una función estimable. Esta hipótesis puede ser probada con el uso del estadístico F, definido en (3):

$$F = \frac{(\boldsymbol{K}'\hat{\boldsymbol{\beta}})'\left[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\right]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}})/rango\,(\boldsymbol{K})}{\hat{\sigma}^2/(n - rango(\boldsymbol{X}))} \qquad (3)$$

Este estadístico tiene una distribución $F_{[rango\,(\boldsymbol{K}),\nu,\lambda]}$, donde el rango de K son los grados de libertad del numerador, $n - rango(\boldsymbol{X})$ son los grados de libertad del denominador (ν) y λ es el parámetro de no centralidad, que tiene la siguiente expresión (4):

$$\lambda = \frac{(\boldsymbol{K}'\boldsymbol{\beta})'\left[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\right]^{-1}(\boldsymbol{K}'\boldsymbol{\beta})}{\sigma^2} \qquad (4)$$

Bajo H_0, λ es igual a 0; cuando H_0 es falsa λ tomará valores mayores a 0. Luego, la potencia quedará determinada como P $\left\{F_{[rango(k),\nu,\lambda]} > F_{[crítica]}\right\}$, donde el lado izquierdo de la inecuación corresponde a una F no central, mientras que el lado derecho es el cuantil (1-α) de una F central para un nivel de significación α determinado.

Los pasos para la estimación de la potencia fueron:

Se calculó el valor crítico $(F_{[crítica]})$ para el estadístico F como el cuantil 1-α de una $F_{[rango(k),\nu,0]}$.

Se estimó λ. Este quedó 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 1 y 2 se computó la potencia como:

Agron. Mesoam. 32(1):196-208, enero-abril, 2021
ISSN 2215-3608 doi:10.15517/am.v32i1.40870

$$P\left\{F_{[rango\,(k),v,\lambda]} > F_{[crítica]}\right\}.$$

Cada estimación de potencia se hizo desde dos hasta veinte repeticiones en cada uno de los 10 000 conjuntos de unidades experimentales con los componentes de varianza previamente estimados.

La matriz de diseño, *X*, de la ecuación (3) es una matriz binaria de dimensión n x p, donde n es el número de repeticiones y p el número de tratamientos. El rango de la matriz, *X*, en este caso por la reparametrización usada, siempre será igual a p. Entonces, conforme se aumente el número de repeticiones, los grados de libertad del denominador para la distribución F definida en (3) serán mayores. Luego, fijos los otros parámetros, el valor obtenido para el estadístico F tiene mayor probabilidad de caer en el área de rechazo, es decir, la potencia estadística de la prueba aumenta.

En el parámetro de no centralidad definido en (4) se debe especificar, en la matriz *K*, los contrastes de interés para el vector de parámetros β, en este caso, medias de tratamientos, lo cual 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 (4) solo tuvo el contraste relacionado al par de medias seleccionado. Así, si se reemplaza K'B en el parámetro de no centralidad de la distribución F, se obtiene la distribución del estadístico de la prueba bajo hipótesis alternativa. De allí es posible calcular la potencia de la prueba para una diferencia entre pares de medias mayor o igual al supuesto para K'B, lo que genera una cuota inferior para lambda.

El nivel de significancia (α) que se utilizó fue de 0,05. La diferencia mínima a ser detectada (tamaño del efecto), se determinó mediante un consenso práctico-estadístico con Ingenieros Agrónomos especialistas en el cultivo en cuestión, se convino que una diferencia del 10 % con respecto a la media y cinco tratamientos eran las situaciones más frecuentes a las que se enfrentan en la práctica de ensayos en Costa Rica. Por lo tanto, las matrices K y β quedaron definidas de la siguiente manera:

$$\boldsymbol{K} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \; ; \; \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{bmatrix}$$

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_i se definió como el efecto de un tratamiento inducido que generó un desvío de 0, o sea, el tamaño del efecto que quiere ser detectado. En general, las medias usadas no fueron de importancia, lo relevante es la diferencia entre medias de tratamientos que se intentó comparar, ya que fueron reflejo del tamaño del efecto (Stroup, 2002).

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 correspondieron a un promedio. Todos los procedimientos respectivos se hicieron con el lenguaje R (R Core Team, 2017).

ISSN 2215-3608 doi:10.15517/am.v32i1.40870

Resultados

Inferencia

Los parámetros estimados para el proceso de correlación espacial para los datos del ensayo en blanco en conjunto con los criterios de información de *Akaike* (AIC) y de información bayesiano (BIC) se presentan en el Cuadro 2. El modelo 4 fue el que presentó el menor AIC y el menor BIC de todos los modelos ajustados, además en la prueba de cociente de verosimilitud, las tendencias de primer y segundo orden en este modelo no fueron significativas (p>0,05). Por lo que, se determinó que el modelo 4 fue el más adecuado de todos los modelos comparados.

Cuadro 2. Parámetros estimados y criterios de bondad de ajuste para los modelos ajustados en el ensayo en blanco con arroz (*Oryza sativa*). Hacienda Mojica, Bagaces, Costa Rica. 2012.

Table 2. Estimated parameters and goodness of fit criteria for the fitted models adjusted in the rice blank (*Oryza sativa*) trial. Hacienda Mojica, Bagaces, Costa Rica. 2012.

Modelo	Estimación			Criterio de información	
	Nugget	Sill	Range	AIC	BIC
0	9,2E-03	0	0	-732,30	-716,30
1	0,01	1,20E-03	9,24	-741,50	-725,60
2	0,01	1,50E-03	9,83	-735,10	-711,20
3	0,00	8,60E-03	1,41	-742,50	-706,60
4	**0,00**	**9,10E-03**	**1,44**	**-746,40**	**-730,40**
5	0,00	9,20E-03	1,44	-737,80	-713,80
6	0,00	8,80E-03	1,41	-742,50	-706,60
7	0,01	1,90E-03	9,02	-741,50	-725,60
8	0,01	1,20E-01	536,30	-737,50	-713,60
9	0,01	9,00E-04	4,95	-736,60	-700,60
10	0,00	9,20E-03	0,43	-740,20	-724,20
11	0,00	9,20E-03	0,44	-731,70	-707,70
12	0,00	8,70E-03	0,36	-736,90	-701,00

Simulación de ensayos en blanco

Las simulaciones de los 10 000 ensayos en blanco de mayor tamaño se hicieron con base en los parámetros estimados con el modelo 4. La estructura espacial fue baja, por tanto, las simulaciones que se desprendieron de los parámetros estimados fueron reflejo de esta.

Unidades experimentales

La evaluación de la estructura de correlación espacial, una vez conformadas las unidades experimentales, indicó que solo 10,54 % de las simulaciones tuvieron un valor de AIC menor para MBCEsf, cuyo promedio fue de -460,68; el promedio de AIC para MB fue de -462,26. Además, tan solo un 3,88 % de las 10 000 simulaciones tuvo

Agron. Mesoam. 32(1):196-208, enero-abril, 2021
ISSN 2215-3608 doi:10.15517/am.v32i1.40870

un valor p menor a 0,05 (inclusive menor a la tasa de error tipo I) para la prueba de cociente de verosimilitud. Lo anterior indicó que las unidades experimentales simuladas se pudieron considerar independientes.

Potencia estadística

Con cinco repeticiones se alcanzó una potencia del 82 %. Las estimaciones de potencia presentadas, debido a la metodología utilizada, fueron invariantes desde el punto de vista práctico a la cantidad de tratamientos en cuestión. El número de tratamientos igual a 5 tuvo influencia únicamente en el tamaño de los campos aleatorios simulados. El Cuadro 3 presenta los resultados de la estimación de la potencia para el distinto número de repeticiones.

Cuadro 3. Potencia de prueba alcanzada para un número dado de repeticiones en un diseño completamente aleatorizado en ensayos con arroz (*Oryza sativ*a). Hacienda Mojica, Bagaces, Costa Rica. 2012.

Table 3. Test power achieved for a given number of repetitions in a completely randomized design in rice (*Oryza sativa*) trails. Hacienda Mojica, Bagaces, Costa Rica. 2012.

Repeticiones	gl	F	EEDM	λ	Potencia	σ^2_{Res}
2	5	6,61	2,18E-02	4,74	0,40	5,01E-04
3	10	4,96	1,80E-02	6,38	0,60	5,05E-04
4	15	4,54	1,57E-02	8,14	0,73	5,05E-04
5	20	4,35	1,41E-02	9,94	0,82	5,05E-04
6	25	4,24	1,29E-02	11,75	0,88	5,06E-04
7	30	4,17	1,19E-02	13,55	0,93	5,06E-04
8	35	4,12	1,12E-02	15,35	0,95	5,07E-04
9	40	4,08	1,06E-02	17,14	0,97	5,08E-04
10	45	4,06	1,00E-02	18,97	0,98	5,07E-04
11	50	4,03	9,56E-03	20,78	0,99	5,07E-04
12	55	4,02	9,16E-03	22,59	0,99	5,07E-04
13	60	4,00	8,80E-03	24,40	1,00	5,08E-04
14	65	3,99	8,49E-03	26,20	1,00	5,08E-04
15	70	3,98	8,20E-03	28,01	1,00	5,08E-04
16	75	3,97	7,94E-03	29,82	1,00	5,08E-04
17	80	3,96	7,71E-03	31,63	1,00	5,08E-04
18	85	3,95	7,49E-03	33,43	1,00	5,08E-04
19	90	3,95	7,30E-03	35,24	1,00	5,08E-04
20	95	3,94	7,11E-03	37,05	1,00	5,09E-04

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 / gl: denominator degrees of freedom; F: critic F; EEDM: standard error of the mean difference; λ: non-central parameter; σ^2_{res}: residual variance.

Discusión

Con base en los resultados obtenidos en esta investigación, para obtener una potencia estadística del 80 % o más, se recomienda usar, al menos, cinco repeticiones si se quiere detectar una diferencia de medias para ensayos de rendimiento, del 10 % a un nivel de significación del 5 %, en condiciones similares a las que tuvo el presente estudio. Con menos de las repeticiones recomendadas la potencia del ensayo podría no llegar a ser suficiente. Si

la probabilidad de rechazar la hipótesis nula es la misma que la de aceptarla, va a generar una inconsistencia entre los resultados de investigaciones; un ensayo con una potencia cercana o menor al 50 % no debería ser realizado (Cohen, 1992; Murphy, 2014). Además, se aduce que el 90 % podría ser un nivel de potencia más apropiado cuando los efectos del tratamiento son importantes, como pueden ser experimentos con variedades nuevas con rendimientos promisorios (Gent et al., 2018).

Los resultados que se presentan en este trabajo ponen de manifiesto la importancia de tomar en cuenta la potencia estadística cuando realizan pruebas de hipótesis. 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 bibliografía consultada, no se encontró ningún trabajo en esa línea. 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 puedan concretar los objetivos propuestos. Conclusiones erróneas podrán ser evitadas solo si los experimentos, mediante un análisis prospectivo, aseguran una potencia adecuada (Gent et al., 2018).

Los resultados obtenidos en esta investigación son para una situación en particular y pueden variar, cada ensayo específico debe ser planeado de una forma adecuada. Sin embargo, provee de información que puede ser utilizada como base para otros experimentos y que de otra manera sería muy difícil de conseguir. Las revistas científicas usualmente no se preocupan por proporcionar datos relevantes acerca de los elementos necesarios para realizar el análisis de potencia, lo que hace imposible contar con datos útiles que puedan ser usados para planificar futuros ensayos (Stroup, 1999).

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. Una de las partes más valiosas del análisis de potencia *a priori* es que, para hacer los cálculos, se debe especificar la hipótesis alternativa (esto es tamaño del efecto) (Quinn y Keough, 2002) 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. “Los estadísticos pueden realizar un gran servicio al familiarizar a la comunidad científica con la importancia de la potencia estadística en el diseño y al dejar en claro que las revistas que no brindan información adecuada sobre la variabilidad dificultan el diseño, lo que contribuye a la inflación del costo en la investigación” (p. 24) (Stroup, 1999).

Conclusiones

En el marco de las condiciones que fue realizado este trabajo se considera que no es recomendable usar menos de cinco repeticiones para ensayos de rendimiento en arroz, si se quiere detectar una diferencia de medias del 10 % a un nivel de significación del 5 %.

Este trabajo brinda información poco disponible que puede ser tomado como base para planificar futuros trabajos.

Agron. Mesoam. 32(1):196-208, enero-abril, 2021
ISSN 2215-3608 doi:10.15517/am.v32i1.40870

Referencias

Ahn, C., Heo, M., & Zhang, S. (2015). *Sample size calculations for clustered and longitudinal outcomes in clinical researc*h. CRC Press, Taylor & Francis.

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R. Springer*. https://doi.org/10.1007/978-1-4614-7618-4

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates Inc.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Cressie, N. (1993). Statistics for spatial data (2nd ed.). Wiley-Interscience. https://doi.org/10.1002/9781119115151

Diggle, P. J., & Ribeiro, P. J. (2010). *Model-based geostatistics*. Springer.

Fagroud, M., & Meirvenne, M. V. (2002). Accounting for Soil Spatial Autocorrelation in the Design of Experimental Trials. *Soil Science Society of America Journal*, *66*(4), 1134–1142. https://doi.org/10.2136/sssaj2002.1134

Gbur, E. E., Stroup, W. W., McCarter, K. S., Durham, S., Young, L. J., Christman, M., West, M., & Kramer, M. (2012). *Analysis of generalized linear mixed models in the agricultural and natural resources sciences* (Chapter 7). American Society of Agronomy, Soil Science Society of America, and Crop Science Society of America.

Gent, D. H., Esker, P. D., & 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, M. I. (2008). Potencia de prueba: La gran ausente en muchos trabajos científicos. *Agronomía Mesoamericana*, *19*(2), 309–313. https://doi.org/10.15517/am. v19i2.5015

Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria. (2008). *Manual de recomendaciones del cultivo de arroz*. Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria.

Kuehl, R. (2001). *Diseño de experimentos: Principios estadísticos de diseño y análisis de investigación* (2nd ed.). International Thomson.

Lapeña, B. P., Wijnberg, K. M., Stein, A., & Hulscher, S. J. (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. C. 2019. *Design and analysis of experiments* (8th ed.). John Wiley & Sons.

Murphy, K. R., Myors, B., & Wolach, A. H. (2014). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (4th ed.). Routledge.

Petitgas, P., Woillez, M., Rivoirard, J., Renard, D., and Bez, N. (2017). *Handbook of geostatistics in R for fisheries and marine ecology*. ICES Cooperative Research Report.

Pinheiro, J., Bates D., DebRoy S., & Sarkar, D. (2016). *nlme: Linear and nonlinear mixed effects models*. R package. http://CRAN.R-project.org/package=nlme.

Quinn, G., & Keough, M. (2002). *Experimental Design and Data Analysis for Biologists* (1st ed. Chapter 7). Cambridge University Press.

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ribeiro, P.J., & Diggle, P.J. (2001). GeoR: A Package for Geostatistical Analysis. *R-News*, *1*, 14–18.

Richter, C., & Kroschewski, B. (2012). Geostatistical Models in Agricultural Field Experiments: Investigations Based on Uniformity Trials. *Agronomy Journal*, *104*(1), 91–105. https://doi.org/10.2134/agronj2011.0100

Robledo, W. (2015). Diseño y análisis de experimentos a un criterio de clasificación. In: M. Balzarini, J. Di Rienzo, M. Tablada, L. González, C. Bruno, M. Córdoba, W. Robledo, & F. Casanoves (Eds.), *Estadística y biometría: Ilustraciones del uso de Infostat en problemas de agronomía* (2nd ed., pp. 257–285). Editorial Brujas.

Rodríguez, N., Sánchez, H., & Pacheco, P. (1993). Determinación de tamaño y forma óptimos de parcela para ensayos de rendimiento con café. *Revista Colombiana de Estadística*, *14*(27), 50–64.

Rosselló, J. M., & Fernández, M. (1986). *Guía técnica para ensayos de variedades en campo*. Organización de las Naciones Unidas para la Agricultura y la Alimentación.

Stroup, W.W. (1999). Mixed model procedures to assess power, precision, and sample size in the design of experiments. In American Statistical Association (Ed.) *Proceedings Biopharmaceutical Section American Statistical Association* (pp. 15–24). American Statistical Association; Biopharmaceutical Section.

Stroup, W. W. (2002). Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agricultural, Biological, and Environmental Statistics, 7*(4), 491–511. https://doi.org/10.1198/108571102780.

Vargas, J. C., & J. R. Navarro. (2019). Tamaño y forma de unidad experimental para ensayos de rendimiento de arroz (Oryza sativa), en Guanacaste, Costa Rica. *Cuadernos de Investigación*, *11*(3), 355–360. https://doi.org/10.22458/urj.v11i3.2653

West, B. T., Welch, K. B., & Gałecki, A. T. (2015). *Linear mixed models: A practical guide using statistical software* (2 ed.). Chapman & Hall/CRC.