Determination of the statistical power in yield experiments in corn

Jorge Claudio Vargas-Rojas[1][§]
Fernando García[2]

[1]University of Costa Rica-Guanacaste Regional Headquarters. Swept El Capulin, Liberia, Costa Rica. CP. 50101. [2]National University of Córdoba-Faculty of Economic Sciences. Bv. Enrique Barros, University City, Cordoba, Argentina. CP. X5000HRV.

[§]Corresponding author: jorgeclaudio.vargas@ucr.ac.cr.

Abstract

Prospective analysis of the statistical power of a hypothesis test should be one of the most important stages of any experiment; however, it is frequently omitted. In particular, for Costa Rica, no studies related to this topic were found for yield experiments in corn cultivation. The objective of this work was to determine the statistical power of a completely randomized design for yield experiments in the cultivation of corn (*Zea mays*) by simulating uniformity trials. To perform the calculations of power, the parameters of the spatial correlation process of a uniformity trial established in Santa Cruz, Costa Rica in 2018 were estimated. These estimates were used to perform 10 000 simulations of larger random fields, which allowed overlapping different number of repetitions and estimate the statistical power to detect a difference of 10% with respect to the mean in an experiment with a completely randomized design at a significance level of 5%. The 80% power was obtained with eight repetitions and it is concluded that, under the experimental conditions of this work, in yield trials in the cultivation of corn, to detect a difference of means of 10% at a level of significance 5%, eight or more repetitions should be used.

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

Introduction

In the planning of an agricultural experiment, it is often sought to compare treatments (genotypes of a cultivar, fertilizer dose, sowing density, formulations of an herbicide, among others). For this, these treatments are applied to experimental units, arranged according to a specific design, where one or more response variables are recorded. Then, with the use of statistical analysis techniques, the means of the response variable for each treatment are estimated and, commonly, hypotheses about them are tested (Robledo, 2015).

In a hypothesis test, two types of mistakes can be made. The first type, called type I error, corresponds to making the decision to reject the null hypothesis, when it is true. The second type, called type II error, corresponds to making the decision not to reject the null hypothesis, when in fact the alternative hypothesis is true (Montgomery, 2019). The statistical power of a hypothesis test is defined as the probability of rejecting the null hypothesis when it is false, that is, the probability of finding differences when they actually exist.

The study of statistical power is based on estimating the probability of making type II error, it is desired that this be small so that its complement (statistical power) is as high as possible (Lapeña *et al*., 2011). In other words, power is the probability that an effect of a given size can be distinguished from the intrinsic random variation of the variable (Gent *et al*., 2018). By convention, an acceptable power level has been set at 80% (Cohen, 1988).

The power function depends on four elements that relate to each other: 1) level of significance; 2) effect size; 3) variability; and 4) number of repetitions. The level of significance, symbolized by (α), is the probability of rejecting H_0 when it is true. The effect size is the difference that is expected to be found between a pair of treatments and is established according to biological, physical, economic, scientific or practical criteria (Kuehl, 2001), in a clearer way, it is the minimum difference considered significant that is desired to detect between treatments. Variability is the residual variance of the experiment. The number of repetitions is the number of independent experimental units per treatment that are required to reach a certain level of power (Cohen, 1992).

There is the problem that the statistical power works found in the literature assume some values for the parameters of the models in question. For example, for a classical linear model, the value of the variance of the mean difference is usually assumed to be known, because it is taken from other works or from previous experiences; however, these considerations may not represent well the actual conditions of the experiment.

Additionally, few journals publish measures of variation useful for the calculation of power, so the information necessary to design the size of the trial is not available or is unreliable (Stroup, 2002); however, this information can be obtained with simulation techniques (Lantuéjoul, 2002). Currently, in the area of geostatistics, the techniques of simulation or generation of random field realizations are frequent, since they play an important role as a tool that allows obtaining information or making inference when the required analytical results are difficult to obtain in practice (Diggle and Ribeiro, 2010).

The purpose of geostatistical simulations is to reproduce the inherent spatial variability of the regionalized variable. The simulation reproduces the expected value and variability of the model in terms of the variogram, which is parameterized by the 'nugget', the 'sill' and the 'range'. Then, it is possible to build many realizations from the same model; each realization will be different but will have the same expected value and the same covariance structure (Petitgas *et al*., 2017). Thus, with simulations of a uniformity trial, it is possible to study the effect of different designs and their variants on statistical power (Richter and Kroschewski, 2012).

The calculation of statistical power is not a usual part of the planning of agricultural experimentation (González-Lutz, 2008). Thus, it is not known if a number of repetitions that allow reaching the desired power are used, which can lead to the conclusions based on the hypothesis tests of some works are not reliable, particularly when no significant differences are found.

It is common for researchers to resort to arbitrary or traditional numbers without statistical or practical justification to define the repetitions that a trial should have. Most of the studies in the Costa Rican agricultural field focus on parametric tests such as analysis of variance or regression and the approach that has been given to the power from a prospective approach is null. In the literature consulted, only the work of Vargas-Rojas (2021) was found, who estimated the number of repetitions for yield trials in rice cultivation.

The appropriate number of repetitions for a trial cannot be estimated without establishing the effect size, variance, significance level and treatment structure. These elements can vary from experiment to experiment depending on the objectives and conditions that the research will have, so it is necessary to study the statistical power for agricultural experiments. The objective of this work was to determine the statistical power of a completely randomized design for yield experiments in the cultivation of corn (*Zea mays*) by simulating uniformity trials in the area of Santa Cruz, Costa Rica.

Materials and methods

Generalities and conditions of the experiment

To carry out the simulations, the data used were those from a uniformity trial carried out during the months of June to September 2016, at the Experimental Farm of Santa Cruz (FESC, for its acronym in Spanish), owned by the University of Costa Rica (10° 17' 6.24" north latitude and 85° 35' 42.95" west longitude), which is located in the canton of Santa Cruz, district of Santa Cruz, province of Guanacaste, at 54 masl. FESC has an average rainfall of 1 834 mm/year with dry season from December to April and with rainy season from May to November (Cerdas, 2015), it has an average annual temperature of 27.9 ℃, with average daily evaporation of 6.8 mm and daily global solar radiation of 18.7 MJ (Instituto Meteorológico Nacional, 2011). The predominant soils have medium to low fertility, are clayey, with characteristics of vertisols by 2:1 expandable clay, are taxonomically classified as Vertic Haplustalfs linked to Typic Haplusterts and Typic Ustorthents (Vega and Salas, 2012).

White corn seed of the HS5G hybrid was used, which is of high yield and low variability in its production (CV= 6.2%) (Cerritos *et al*., 1994). Sowing was carried out manually, in furrows separated by 1 m and with a distance between plants of 0. 25 m, for a density of 40 000 plants ha^{-1}. The selected plot had flat topography, without any factor that could cause systematic variability.

The uniformity trial technique described by Rodríguez *et al*. (1993) was used. According to this method, a corn plot of 26 m × 26 m (676 m^2) was sown, in which a three-meter border was left around its entire perimeter; thus, an area of 20 m × 20 m (400 m^2) was obtained to carry out the uniformity trial. All base units were indexed according to a Cartesian coordinate system. Then, 110 days after sowing, the whole corncob (production in grams) belonging to each base unit was harvested.

Spatial variation

To estimate the parameters of spatial correlation ('nugget', 'sill' and 'range') of the uniformity trial, two of the most used spatial correlation models were fitted to the data: exponential and spherical (Cressie, 1993; Bivand *et al*., 2013), these two models assume that observations that are close are more likely to have similar magnitudes and model this spatial structure with distance functions (Fortin *et al*., 2016). The model of independent errors was also adjusted, this assumes that the distance does not affect the similarity between the observations.

In the case of models with correlation, a structure of means that included the fixed effect as a first- and second-order trend of the Cartesian coordinates was also modeled in order to discount, if any, large-scale trends. Therefore, the models shown in Table 1 were adjusted by restricted maximum likelihood (REML).

Table 1. Models adjusted to the uniformity trial of corn (*Zea mays*). Santa Cruz, Costa Rica 2018.

Model	Correlation structure	Nugget effect	First-order trend	Second-order trend
0	no	no	no	no
1	Spherical	yes	no	no
2	Spherical	yes	yes	yes
3	Spherical	yes	no	yes
4	Spherical	no	no	no
5	Spherical	no	yes	no
6	Spherical	no	no	yes
7	Exponential	yes	yes	no
8	Exponential	yes	yes	yes
9	Exponential	yes	no	yes
10	Exponential	no	no	no
11	Exponential	no	yes	no
12	Exponential	no	no	yes

Then, for the comparison between the adjusted models, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used. Lower values of AIC or BIC indicate better fit of the statistical model (West *et al*., 2015). To compare the models with and without fixed effect of the coordinates (different structure of means, but same covariance), the likelihood ratio test (LRT), based on estimates of maximum likelihood (ML), was used. Finally, the estimated values for the 'nugget', 'sill' and 'range' of the model, which, in comparative terms, had a better fit to perform the simulation of the respective random fields, were used.

Simulation of random fields

The uniformity trial of this work, due to the size, has limitations to overlap a set of experimental units for different number of repetitions of the experiment. The spatial correlation parameters estimated in the uniformity trial were then used to simulate larger plots (random fields), so that it could achieve greater flexibility to overlap more repetitions in the design of the experiment.

The geoR package (Ribeiro and Diggle, 2001) of the R language (R Core Team, 2017) was used to simulate 10 000 random fields, each corresponding to a grid of 160 m × 20 m, with a space between points of 1 m × 1 m. Each point of the grid was a realization of the random variable, in this case yield in grams, parameterized with the estimated parameters of the uniformity trial.

In each of the 10 000 random fields, experimental units of dimensions 8 m long by 4 m wide (32 m^2), size defined by (Vargas-Rojas and Navarro-Flores, 2020), were formed by aggregation of realizations of each point in the grid, resulting in the formation of 10 000 sets of 100 experimental units, arranged in five columns and 20 rows. Within each set, all experimental units were indexed with Cartesian coordinates that defined the spatial location of their centroid. Once these were formed, the average yield in kilograms per experimental unit was estimated and the presence or not of spatial correlation was evaluated.

Statistical power

Estimation of variance components

By having a set of 100 experimental units, arranged in 20 rows and five columns, it was possible to recreate different numbers of repetitions for the CRD. For example, to randomize five treatments with two repetitions each, over the 100 experimental units, the columns were considered as fixed, then two rows were taken, thus ten experimental units were obtained to randomize the treatments.

Each time a repetition was required, another row from the set of one hundred experimental units was taken. In each of the 10 000 simulated random fields, different numbers of repetitions (from two to 20) were recreated and the variance components, in this case only the residual variance, were estimated for the different plans. The procedure was performed with the nlme package (Pinheiro *et al*., 2016) of the R language (R Core Team, 2017). Once the components of variance

were estimated, the calculations of the statistical power for the hypothesis test of mean difference were made under the approach of the classical linear model, described below. The notation proposed by Stroup (2002) was followed.

Classical linear model

The classical linear model is defined as follows: y=Xβ+e   1). Where: y= is the vector of observations, of dimension n × 1; X= is the design matrix, of dimension n × p; β= is the vector of parameters, of dimension p × 1; e= is the vector of random errors, of dimension n × 1. Under this approach, the null hypothesis (H_0) is defined in the form $K'β=0$, where $K'β$ is an estimable function. This hypothesis can be tested with the use of the F statistic, defined below: F=

$$\frac{(K'\hat{β})'\left[K'(X'X)^{-1}K\right]^{-1}(K'\hat{β})/\text{range (K)}}{\hat{σ}^2/(n-\text{range}(X))} \qquad 2).$$

This statistic has a distribution $F_{[\text{range}(K),ν,λ]}$, where the range of K are the degrees of freedom of the numerator, -range (X) are the degrees of freedom of the denominator ($ν$) and $λ$ is the non-centrality parameter, which has the following expression: $λ=\dfrac{(K'β)'\left[K'(X'X)^{-1}K\right]^{-1}(K'β)}{σ^2}$ 3). Under H_0, $λ$ is equal to 0; when H_0 is false, $λ$ will take values greater than 0. Next, the power will be determined as $P\{F_{[\text{range}(k),ν,λ]}>F_{[\text{critical}]}\}$.

For the estimation of power, first, the critical value ($F_{[\text{critical}]}$) was calculated. Then, $λ$ was estimated, defined according to the structure of the design that matrix X determined; the vector of treatment means that β determined; the contrast vector that K defined and the estimated residual variance. Once 2 and 3 were obtained, the power was estimated as: $P\{F_{[\text{range}(k),ν,λ]}>F_{[\text{critical}]}\}$.

Each power estimate was made from two to 20 repetitions in each of the 10 000 sets of experimental units, with the components of variance previously estimated. In the non-centrality parameter defined in (3), it was necessary to specify the contrasts of interest in the matrix K, and the means of treatments in the vector of parameters β. The foregoing is operationally difficult since there are infinite possible contrasts. To facilitate the calculation of the non-centrality parameter, the contrast matrix K was restricted to a single contrast that consisted of the comparison of an (arbitrary) pair of means, as suggested by Kuehl (2001). As a result, equation (3) only had the contrast related to the selected pair of means.

Thus, if $K'β$ is replaced in the non-centrality parameter of the F distribution, the distribution of the test statistic under the alternative hypothesis is obtained, from where the power of the test can be calculated for a difference between pairs of means greater than or equal to the assumption for $K'β$, which generates a lower bound for lambda. The significance level (α) used was 0.05. The minimum difference to detect (effect size) was determined by a practical-statistical consensus with Agricultural engineers specialized in the crop in question; it was agreed that a difference of 10% with respect to the general mean was the most frequent situation they face in the practice of trials in Costa Rica. Due to the methodology used, the power estimates are invariant to the number of treatments in question, arbitrarily, five treatments were used as a reference.

Therefore, the matrices K and β were defined as follows: $K=\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$; $\beta=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{bmatrix}$. Each element μ_i of the

vector β corresponds to the mean of each treatment. Where: $\mu_i = \mu + \tau_i$; μ denotes the general mean and τ_i denotes the difference between the i-th treatment mean and μ. For this work, μ would be the general mean of the variable production; however, this was assigned a value of 0. Then, arbitrarily, τ_1 was defined as the effect of an induced treatment that generated a deviation of 0, that is, the size of the effect that is desired to be detected.

The power estimate was made for each of the 10 000 sets of simulated experimental units, so the power and statistics shown correspond to an average. All respective procedures were done with the R language (R Core Team, 2020).

Results

Inference

The models that had the best fit were 10, 11 and 12. The estimated spatial correlation parameters together with the AIC and BIC criteria for these models are shown in Table 2.

Table 2. Estimated parameters and goodness of fit criteria for adjusted models of the uniformity trial of corn (*Zea mays*). Santa Cruz, Costa Rica 2018.

Model	Estimate			Information criterium	
	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

Model 12 was the one that had the lowest AIC of all the adjusted models. In order to evaluate whether the estimation of more parameters was justified, this model was compared with models 10 and 11, the results are shown in Table 3. The comparison with model 11 was not significant ($p<$ 0.05), on the other hand, the comparison with model 10 was significant ($p<$ 0.05), which suggests that model 11 is the right one. Despite this, because the comparison with model 10 had a p-value close to the decision limit and since, relatively, a small neighborhood was used to simulate the uniformity trials, local seasonality was assumed. For these reasons, model 10 was considered the right one, also because it was the one that had the lowest BIC.

Table 3. Statistics and likelihood ratio test for the models adjusted with ML in corn (*Zea mays*) cultivation. Santa Cruz, Costa Rica 2018.

Model	Test	Reference	log L	AIC	BIC	Test df	-2log (L)	*p*-value
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			

Simulation of uniformity trials

Simulations of the largest 10 000 uniformity trials were performed based on the parameters estimated with model 10.

Experimental units

The evaluation of the spatial correlation structure, once the experimental units were formed, showed that the simulated experimental units can be considered independent, so the statistical power was estimated without taking into account the spatial correlation.

Statistical power

With eight repetitions, 80% power was reached. Table 4 presents the results of the estimation of the power and other statistics for the different numbers of repetitions.

Table 4. Statistical power reached for a given number of repetitions in a completely randomized design in yield trials with corn (*Zea mays*). Santa Cruz, Costa Rica 2018.

Repetitions	df	F	SEMD	λ	Power	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	2.66 E-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

df= degrees of freedom of the denominator; F= critical F; SEMD= standard error of the mean difference; λ= non-centrality parameter; σ^2_{res}= residual variance.

Discussion

Based on the results obtained in this research, if one wants to obtain a statistical power of 80% to detect a difference of means of 10% with respect to the general mean, at a level of significance of 5% in yield trials, it is recommended to use eight repetitions. With less than the recommended repetitions, the power of the trial may not be sufficient. In this sense, Cohen (1992); Murphy *et al*. (2014) mention that, if the probability of rejecting the null hypothesis is the same as that of not rejecting it, it will generate an inconsistency between the results of studies; a trial with a power close to or less than 50% should not be carried out.

On the other hand, Gent *et al*. (2018) argue that 90% could be a more appropriate power level, when the effects of the treatment are significant, as is the case with experiments where the yields of promising varieties are evaluated. It is worth noting that the recommendation made is applicable under experimental conditions similar to those of the present study and that this may vary. Each specific trial must be planned in an appropriate manner. However, the information that was generated can be used as a basis for other experiments, information that would otherwise be very difficult to obtain.

For example, if based on previous experiments or pilot studies, there is information on the spatial variability of a site, the estimates of 'sill' and 'range' can be compared with the estimated values of Table 2 and depending on the comparison, adapt the number of repetitions for that particular site. The estimation of the 'sill' parameter is an estimate of the residual variance of the regionalized variable (Diggle and Ribeiro, 2010; Guedes *et al*., 2020), therefore, if it is smaller, the number of repetitions recommended in this work could be reduced.

The same happens with the 'range', Lapeña *et al*. (2011) studied, through simulations, factors that affect statistical power in the presence of spatial correlation and determined that an increase in the spatial dependence of the data, reflected in the 'range', increases the variance of the distribution of the test statistic, which reduces the power. So, if one has a lower 'range', fewer repetitions could be used, otherwise, no. On the other hand, if hybrids with less variability in yield than that reported for the HS5G hybrid are to be evaluated, reducing the number of repetitions could be evaluated, in case of materials with greater variability, reducing the number of repetitions would not be an option.

The results presented in this paper highlight the importance of considering statistical power when testing hypotheses. Formal consideration of the statistical power of an experiment should therefore be a routine and indispensable component of experimental design prior to data collection and analysis. Another motivation for conducting a power analysis in the early stages of planning an experiment is to ensure that resources are available to ensure that the proposed objectives can be achieved. Erroneous conclusions can be avoided only if experiments, through prospective analysis, ensure adequate power (Gent *et al*., 2018).

Carrying out more studies on this subject would be beneficial for the Costa Rican agricultural sector, knowing in depth the conditions in which an experiment is going to be executed will allow it to be properly designed. It is necessary that with this information there can be a dialogue between

the developers of trials and the statisticians, so that the former communicate their needs and the latter can transfer them to the corresponding statistical terms, this helps to make clear the objectives and the true scope of an investigation.

As Quinn and Keough (2002) mention, one of the most valuable aspects of a priori power analysis is that, in order to make the calculations, the alternative hypothesis (this is the effect size) and, most importantly, the statistical model that will be applied to the data must be specified. Specifying the model makes one think about analysis before collecting the data, a recommended habit.

Conclusions

Within the framework of the conditions that this work was carried out, it is considered that: If it is desired to detect a difference of means of 10% at a level of significance 5%, it is not recommended to use less than eight repetitions for yield trials in corn. Fewer repetitions could be plausible depending on the conditions of the experiment. This work provides information that is little available, which can be taken as a basis for planning future works.

Cited literature

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.

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 Gałecki, A. T. 2015. Linear mixed models: a practical guide using statistical Software. 2nd (Ed.). CRC Press. 38-41 pp.