Field Crops Research 121 (2011) 286–290

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

Partial collection of data on potato yield for experimental planning

Lindolfo Storck*

Department of Crop, Federal University of Santa Maria (UFSM), 97105-900 Santa Maria, RS, Brazil

ARTICLE INFO

Article history:
Received 16 August 2010
Received in revised form
16 December 2010
Accepted 18 December 2010

Keywords:
Solanum tuberosum
Experiment planning
Plot size
Sampled experimental area
Bootstrap resampling

ABSTRACT

The aim of this study was to estimate the number of blank experiments (BE) (i.e., a uniformity trial) required to estimate the optimum plot size for use in experiments involving potato crops. The study was based on data on the mass of potato tubers (Solanum tuberosum L.) from 3456 hills (i.e., 24 rows of 144 hills each) obtained from a BE. Using these data, BE of different sizes (i.e., 2 rows of 24, 36, 48 and 72 hills) were planned to estimate optimum plot size. For each BE, 11 plot sizes (*X*) were planned based on the sum of adjacent hills, and the mean, variance and coefficient of variation (CV) between plots of the same size were calculated. Regression models for CV were adjusted in terms of *X* to estimate the optimum plot size. For each BE size, a bootstrap resampling method was used to estimate the sufficient number of BE to enable precise estimates of optimum plot size, mean and other statistics. It was found that a sampling potato hill yield of 39% of subdivisions within an experimental area where a potato experiment is to be performed is sufficient to estimate optimum plot size for the experiment. Plots composed of one row of six hills are sufficient to estimate potato yield.

1. Introduction

Knowing the variability of agricultural crop yields in an experimental area is important in order to plan experiments effectively. This planning entails determining, for a given number of treatments, the experimental design, plot size and number of replications (Gomez and Gomez, 1984; Steel et al., 1997; Ramalho et al., 2000).

When attempting to control the experimental error generated by the heterogeneity of experimental units, researchers must dedicate special attention to certain aspects of estimation. So, it is necessary to perform blank experiments (i.e., a uniformity trial), with the aim of selecting appropriate block design, plot size and shape, the number of replications and treatments, and the level of precision for the chosen experimental area. The best combination for plot size and the number of replications and treatments are the main methods used to address the heterogeneity of crop yield in order to maximize the information obtained from the experimental area (Storck et al., 2006).

To understand variability in an experimental area, blank experiments may be performed that are specifically designed to identify the variability or level of heterogeneity for a given characteristic. In this methodology, data are collected on the yields of small plots called "basic units" (BU), and these BU are used to plan plots of different sizes, according to the grouping together adjacent BU. For each plot size, the number of plots of a given size together with means, variances and coefficients of variation are calculated. Based on the relationship between basic unit variances and planned plot size, the level of heterogeneity is estimated (Smith, 1938). Meier and Lessman (1971) showed that the optimum plot size can be obtained as the point of maximum curvature in the relationship between the coefficients of variation and the respective planned plot sizes. However, the use of blank experiments for an entire experimental area during planning is time-consuming and expensive. As a result, data obtained from experiments showing the effects of treatments in plots that have been subdivided into block, plot and subplot size or arranged in a lattice design may be used (Koch and Rigney, 1951) to make unbiased estimates of the level of heterogeneity (Hatheway and Williams, 1958).

Lin and Binns (1984) discussed the possibility of estimating the level of heterogeneity in an experimental area based on intraclass correlation for experiments using a randomized complete block design in which there are two plot sizes (i.e., blocks and plots). Applications for experiments involving common bean (Phaseolus vulgaris L.) cultivars (Storck et al., 2007) and corn (Zea mays L.) cultivars have also been discussed (Carvalho et al., 2009). Estimates of the level of heterogeneity based on the intraclass coefficient of correlation are not as precise as those obtained from blank experiments using Smith's (1938) method (Lin and Binns, 1984, 1986) or those obtained from experiments using lattice and subdivided plot arrangements (Koch and Rigney, 1951). However, due to the ease of obtaining estimates of the level of heterogeneity and the coefficients of variation with respect to a large number of experiments over the course of time and in different environments without

* Tel.: +55 55 3220 8451; fax: +55 55 3220 8899.
E-mail address: lindolfo@pq.cnpq.br

0378-4290/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.fcr.2010.12.018

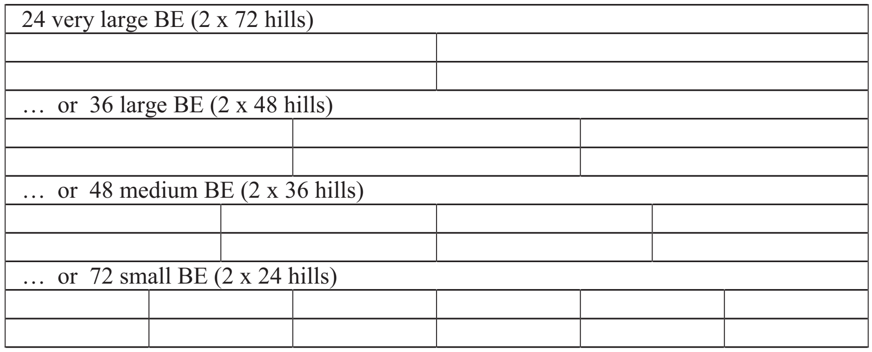

Fig. 1. Different sizes of blank experiments (BE) planned in the same experimental area.

additional costs. So, experimental precision may be monitored and improved using other methodologies, such as that proposed by Hatheway (1961), by varying the number of replications, treatments and plot size.

For some species cultivated in largely independent rows, such as corn (Zea mays L.), sorghum (Sorghum bicolor), sunflower (Helianthus annuus L.), cassava (Manihot esculenta) and potato (Solanum tuberosum), the possibility of using data from certain rows or groups of rows of a given length taken randomly from across the entire experimental area is a feasible alternative according to Lorentz et al. (2010). From this sample of data representing part of the experimental area, estimates may be obtained of the parameters necessary for experimental planning without loss of precision and at a lower cost in terms of financial and human resources.

As such, the aim of this work was to estimate the number of blank experiments required to estimate optimum plot size for use in experiments involving potato crops.

2. Materials and methods

2.1. Experimental data

This study was based on data on the mass of potato tubers (Solanum tuberosum L.) (g hill^{-1}) obtained from a blank experiment involving the Macaca cultivar that was conducted in the potato-growing area at the State Agricultural Research Foundation (*Fundação Estadual de Pesquisa Agropecuária*) in Júlio de Castilhos, Rio Grande do Sul, Brazil (29°12′S, 53°41′W; altitude 490 m). The area received uniform crop management and phytosanitary treatment (Pereira and Daniels, 2003). In all, 3456 hills were cultivated and distributed into 24 rows (width) of 144 hills (length). The rows were spaced 0.80 m apart, with 0.30 m space between hills. The crops were harvested and evaluated separately for all hills, which were the basic units (BU).

2.2. Sizes of plot and blank experiments

Using these data, plans were made for 72 small blank experiments (BE) (i.e., 2 rows of 24 hills), 48 medium BE (i.e., 2 rows of 36 hills), 36 large BE (i.e., 2 rows of 48 hills) and 24 very large BE (i.e., 2 rows of 72 hills), all using a total number of 3456 hills in the experimental area (Fig. 1).

For each of the 180 BE (i.e., 72 small BE, 48 medium BE, 36 large BE and 24 very large BE), 11 plots were proposed, varying the number of rows ($X1$) and the number of adjacent hills ($X2$) with the size of a plot (X) was calculated according to $X = X1 \times X2$. The proposed plots ($X1$;$X2$) were (1;1), (1;2), (1;3), (1;4), (1;6), (1;8), (1;12), (2;4), (2;6), (2;8) and (2;12). The number of replications (N) for each X was limited in accordance with the number of BE (Ne; Ne = 24 for very large BE, Ne = 36 for large BE, Ne = 48 for medium BE and Ne = 72 for small BE) for each BE size, i.e., $N = 3456/(X \times \text{Ne})$.

For each of the 180 BE and for each plot size X, the following information was determined: the variance ($V(x)$) between plots of size X; the variance per BU (VU(X)), calculated between plots of size X as $\text{VU}(x) = V(x)/X^2$; the coefficient of variation (CV(x)) between plots of X BU size; the mean ($M(x)$) of plots of size X; and the mean (M_1) of plots of one BU. Based on these calculations, for each of the 180 BE, the level of yield heterogeneity (b) was estimated using the relationship identified by Smith (1938), that is, $\text{VU}(x) = V_1/X^b$. Parameters V_1 and b were estimated based on the logarithmic transformation of the function and weighted by degrees of freedom (DF = $N - 1$) (Steel et al., 1997). The coefficient of determination (R^2) was also estimated. Similarly, estimates were made of parameters A and B for the function $\text{CV}(x) = A/X^B$, which was used to calculate the optimum plot size (X_0) following the modified maximum curvature method (Meier and Lessman, 1971), with $X_0 = [A^2B^2(2B+1)/(B+2)]^{1/(2B+2)}$.

2.3. A sufficient sample size for bootstrapping

For each BE size, the Ne estimates of the M_1, V_1, b, A, B, R^2 and X_0 statistics were used to estimate the sufficient number of BE or the adequate fraction of the experimental area with a total of 3456 hills. Considering that these statistics generally do not follow a known probability distribution, a bootstrap resampling methodology (Efron, 1979; Ferreira, 2005; Confalonieri et al., 2006) was adopted. Following this method, Ne observations for each statistic were used to generate 2000 samples with replacement (i.e., resampling), and the mean was calculated for each resample. These 2000 mean figures were ordered to identify the minimum (Min) and maximum (Max) values. The Min and Max values were considered to be an estimate for a confidence interval with a zero error rate. The same procedure involving 2000 samples with replacement values was performed for different sample sizes, that is, BE number (k = 3, 4, ..., Ne − 1). For each k value, the 2000 mean figures were ordered to identify the 0.025 quantile as the lower limit (LL(k)) and the 0.975 quantile as the upper limit (UL(k)). The LL(k) and UL(k) were used as estimates for a confidence interval with an error rate of 5% for Ne = k. The BE number was considered sufficient when LL(k) > Min and UL(k) < Max with an error rate of no more than 5%. This was used to determine the lowest value of k sufficient (k') to sample the population of Ne blank experiments.

Accordingly, for each BE size and statistic evaluated, the sufficient fraction of the experimental area to be evaluated was $p(k') = k'/\text{Ne}$.

To make the calculations, an Excel spreadsheet was used together with the SAEG (2007) application and an application created in Pascal language for bootstrap simulation.