

Formulas for series of experiments in a Latin square design in a split-plot arrangement

Delfina de Jesús Pérez-López¹

Artemio Balbuena-Melgarejo¹

Jesús Hernández-Ávila¹

J. Ramón Pascual Franco-Martínez¹

José Antonio Rodríguez-González²

Andrés González-Huerta^{1,§}

1 Centro de Investigación y Estudios Avanzados en Fitomejoramiento-Facultad de Ciencias Agrícolas-Campus Universitario 'El Cerrillo'-Universidad Autónoma del Estado de México. El Cerrillo Piedras Blancas, Toluca de Lerdo, Estado de México, México. AP. 435. Tel. 722 29 65518, ext. 60382. (djperel@uaemex.mx; jhernandez@uaemex.mx; jrfrancom@uaemex.mx; abalbuenam@uaemex.mx).

2 Programa de Maestría en Ciencias Agropecuarias y Recursos Naturales-Facultad de Ciencias Agrícolas-Universidad Autónoma del Estado de México. El Cerrillo Piedras Blancas, Toluca de Lerdo, Estado de México, México. AP. 435. Tel. 722 29 65518. (jrodriguez008@alumno.uaemex.mx).

Autor para correspondencia: agonzalezh@uaemex.mx.

Abstract

Series of experiments in time, space, or combining both environments in a split-plot or split-split-plot arrangement in a Latin square design have not been frequently used. In this research, a statistical model and formulas are constructed to obtain degrees of freedom and sum of squares using quadratic or matrix forms and least squares, when the components of time and space are confused in environments, as a prerequisite to extend their analysis with balanced subsampling. In relation to the main plot, it is assumed that there is no interaction between rows, columns and levels of factor A, and rows and columns with the environments; this restriction also applies between rows and columns or both, with factors A, B and their interaction, but it is also indicated how to reach the same results by introducing the principle of crossing and nesting, particularly if a statistical package will be applied; emphasis is also placed on the formulas to directly calculate degrees of freedom and sum of squares for errors a and b, as well as for those corresponding to main plots and subplots.

Keywords:

matrices, statistical models, years and localities confused in environments.



Introduction

The analysis and discussion of the series of experiments in agricultural and forestry sciences, among other disciplines, have been frequently addressed from the construction of their statistical-genetic models, the obtaining of an analysis of variance, and the application of a methodology for the comparison of means between levels of two or more factors (Sahagún, 1993, 1994, 2007).

Statistical processing of data from split-plot (SP) or split-split-plot (SSP) arrangements can be performed easily and reliably using InfoStat and InfoGen (Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini and Di Rienzo, 2016) and SAS, as can be the one corresponding to the series of experiments across years, localities, or both (González *et al.*, 2019; Pérez *et al.*, 2022; González *et al.*, 2024b). The most commonly used experimental designs are completely randomized (CRD) and randomized complete blocks (RCBD) (Martínez, 1988; Gomez and Gomez, 1984; Little and Hills, 2008; Montgomery, 2009), but Ledolter (2010) analyzed and discussed SP arrangements for factorial assays and for fractional factorials, with emphasis on the CRD and RCBD experimental designs.

The series of experiments for an LSD, in SPs or SSPs, could be fundamental to help detect important differences between years, between localities or between both when choosing different technological packages or different agronomic managements, aimed at the generation, validation, application or transfer of technology to farmers' land, with the purpose of increasing crop productivity and improving the quality of the raw material derived from them and used in Mexican agribusiness.

Genetic improvement, seed production, or intra- and interspecific hybridization could also be better documented based on these statistical methodologies (González *et al.*, 2019; Pérez *et al.*, 2022; González *et al.*, 2024a, 2024b). Farmers' lands that are very heterogeneous also present this type of random variability, which is undesirable to establish an adequate experiment, such environmental heterogeneity could be controlled more efficiently using SPs or SSPs in an LSD (González *et al.*, 2019; Pérez *et al.*, 2022; Rodríguez *et al.*, 2025).

Likewise, few reports on statistical models, analysis of variance, and comparison of means for the LSD have been found in factorial experiments (Tirado and Tirado, 2017; González *et al.*, 2019), particularly when an SP or SSP arrangement is used with balanced subsampling and with the application of a statistical package Proc Anova: Latin Square Split Plot :: SAS/STAT(R) 9.22 User's Guide. (Gomez and Gomez, 1984; Martínez, 1994; Montgomery, 2009).

In the previous context, this research built a statistical model and generated the formulas to calculate DF and SS with two methodologies, as a prerequisite for its validation with some software, such as InfoStat, InfoGen or SAS, among others.



Materials and methods

Symbology

This research will use the same terminology used by Mendenhall (1987); Sahagún (2007a); Montgomery (2009); Pérez *et al.* (2022); González *et al.* (2023, 2024a, 2024b). The five classification factors that will now be considered are environments, rows, columns, A and B, which will henceforth be identified as E, H, C, A and B, respectively; the subscripts and their levels will also correspond to: $m=1, 2, 3, \dots$; $e; i=1, 2, 3, \dots$; $h; j=1, 2, 3, \dots$; $c; k=1, 2, 3, \dots$; $t; l=1, 2, 3, \dots$; b .

Statistical model

The statistical model and formulas for calculating degrees of freedom and sum of squares in an LSD in SPs for a single trial was proposed by Rodríguez *et al.* (2025) and in the most practical situation, their series of experiments can be constructed assuming the absence of interaction between rows (H) and columns (C), between each of these or both with the factors E, A, B and of H and C with the interactions that can be generated with the latter three.

With five classification factors, E, H, C, A and B, there will be a total of 32 possible interactions, which can be obtained by applying combinations of five factors taken from zero, one, two, three, four and five times at the same time

$$(C_0^5 + C_1^5 + C_2^5 + C_3^5 + C_4^5 + C_5^5)$$

but of these, the first two are not viable because it is impossible to form combinations between any of them or between those corresponding to a single factor. Therefore, only the following are possible:

$$\left(\frac{1}{120}\right)$$

$$n(n-1)\{60 + (n-2)\{20 + (n-3)[5 + (n-4)]\}\} =$$

$$\left(\frac{1}{120}\right) 5(4) \{60 + 3\{20 + 2(5+1)\}\} = \left(\frac{1}{120}\right) 20(60+96) = 26$$

interactions, but this number is easier to calculate as: $32 - 1 - 6 = 26$ (first, second, third, and fourth order interactions).

With this background, it is easier to build the reference model, particularly if the guide published by Sahagún (1998) is correctly applied, who, before introducing the nesting principle, suggests building a preliminary model assuming that these factors only have crossover relationships. Thus, the proposed model is: $Y_{ijklm} = \mu + E_m + H_i + C_j + A_k + (EA)_{mk} + (EHA)_{i(mk)} + B_l + (EB)_{ml} + (AB)_{kl} + (EAB)_{mkl} + \varepsilon_{ijklm}$.

Where: Y is the variable that will be analyzed; μ is the general arithmetic mean; E_m , H_i , C_j , A_k , and B_l are the main effects caused by the factors identified as environments, rows, columns, A and B,

respectively; $(EHA)_{i(mk)}$ is the error a; $(AB)_{kl}$, $(EA)_{mk}$, $(EB)_{ml}$ and $(EAB)_{mkl}$ are first and second order interactions originated by the combination of two or three of the factors previously described; ϵ_{ijklm} is the error b or residual of the model.

Experimental design

To form a series of experiments in SPs in an LSD, with years and locations confused in environments, for each trial the levels of factor A are distributed in the main plots based on a Latin square design, as suggested by Smith (1951); Martínez (1994); Tirado and Tirado (2017) and the levels of factor B will be randomized independently in each subplot.

Recommended software

To obtain a randomization plan for each experiment in this type of experimental design, it is possible to use the statistical package Statistical Tools for Agricultural Research (Star) of the International Rice Research Institute (IRRI), based in Los Baños, Philippines [Digital Tools | International Rice Research Institute (irri.org)].

Likewise, SAS (Statistical Analysis System; SAS OnDemand for Academics | SAS) or other statistical packages that have this type of tool to generate them can be employed. To perform manual calculations with quadratic or matrix forms, it is possible to use the scientific calculator freely available at <https://matrixcalc.org/es/>. To analyze data, the statistical packages referenced above or others, such as InfoGen and InfoStat, could be used (Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini and Di Rienzo, 2016).

Results

In main plots, the number of levels for rows (H), columns (C) and factor A is equal ($h = c = t$); the combination ht or ct is equivalent to t^2 . For an RCBD in SPs, H or C, but not both could be considered as r , which would be the number of replications for each combination AB chosen for the experiment. The series of experiments is generated by adding E, the number of environments to be evaluated. The results presented below are an extension of the case proposed by Rodríguez *et al.* (2025).

Formulas for degrees of freedom (DF)

DF total = $et^2b - 1$; DF E = $e - 1$; DF H = $h - 1 = t - 1$; DF C = $c - 1 = t - 1$; DF A = $t - 1$; DF ExA = $(e - 1)(t - 1)$; DF error a = $t[e(t - 1) - 2] + 2$; DF main plots (MP) = DF E + DF H + DF C + DF A + DF ExA + DF error a. For verification: DF MP = $et^2 - 1$; DF B = $b - 1$; DF AxB = $(t - 1)(b - 1)$; DF ExB = $(e - 1)(b - 1)$; DF ExAxB = $(e - 1)(t - 1)(b - 1)$; DF error b = $et(t - 1)(b - 1)$; DF subplots (SUB) = DF total - DF E - DF H - DF C - DF A - DF ExA - DF error a - DF B - DF AxB - DF ExB - DF ExAxB - DF error b. Also: DF SUB = DF B + DF AxB + DF ExB + DF ExAxB + DF error b. For verification: DF SUB = $et^2(b - 1)$; DF MP + DF SUB = $(et^2 - 1) + et^2(b - 1) = et^2b - 1 =$ DF total.

Formulas for sum of squares (SS)

In the denominator of the following formulas, h or c is null; E, H, C, A, and B will be related to the subscripts m , i , j , k , and l , respectively. Quadratic or matrix forms will be written as in González *et al.* (2023); González *et al.* (2024a, 2024b), in these, before their application, sums or totals made on the subscript or subscripts that are not shown in their numerator will be used, the matrix J is square and is only formed by ones and has et^2b rows and columns.

$$SS \text{ Tot} = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2 b e} \right)^2$$

$$= Y'Y - \left(\frac{1}{t^2_{be}} \right) Y'JY; SS H$$

$$= \left(\frac{1}{t_{be}} \right) \sum_{i=1}^h Y_{i...}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2$$

$$= \left(\frac{1}{t_{be}} \right) Y'_{i...} Y_{i...} - \left(\frac{1}{t^2_{be}} \right)$$

$$SS C = \left(\frac{1}{t_{be}} \right) \sum_{j=1}^c Y_{.j...}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2$$

$$= \left(\frac{1}{t_{be}} \right) Y'_{.j...} Y_{.j...} - \left(\frac{1}{t^2_{be}} \right)$$

$$Y'JY; SS = \left(\frac{1}{t_{be}} \right) \sum_{k=1}^t Y_{..k..}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2$$

$$= \left(\frac{1}{t_{be}} \right) Y'_{..k..} Y_{..k..} - \left(\frac{1}{t^2_{be}} \right) Y'JY$$

$$SS E = \left(\frac{1}{t^2_{be}} \right) \sum_{m=1}^e Y_{....m}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2$$

$$= \left(\frac{1}{t^2_{be}} \right) Y'_{....m} Y_{....m} - \left(\frac{1}{t^2_{be}} \right) Y'JY$$

SS TREAT1= SS E+SS A+SS ExA; Thus: SS ExA= SS TREAT1-SS E-SS A. In addition:

$$SS TRAT1 = \left(\frac{1}{t_b} \right) \sum_{k=1}^t \sum_{m=1}^e Y_{..k.m}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2$$

$$= \left(\frac{1}{t_b} \right) Y'_{..k.m} Y_{..k.m} - \left(\frac{1}{t^2_{be}} \right) Y'JY$$

SS main plots (MP)=

$$\left(\frac{1}{b} \right) \sum_{i=1}^h \sum_{k=1}^t \sum_{m=1}^e Y_{i.k.m}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2$$

$$= \left(\frac{1}{b} \right) Y'_{i.k.m} Y_{i.k.m} - \left(\frac{1}{t^2_{be}} \right) Y'Y$$

error a=

$$\begin{aligned} & \left(\frac{1}{b} \right) \sum_{i=1}^h \sum_{k=1}^t \sum_{m=1}^e Y_{i.k.m}^2 - \left(\frac{1}{t_{be}} \right) \sum_{i=1}^h Y_{i...}^2 \\ & - \left(\frac{1}{t_{be}} \right) \sum_{j=1}^c Y_{j...}^2 - \left(\frac{1}{t_b} \right) \sum_{k=1}^t \sum_{m=1}^e Y_{..k.m}^2 + 2 \\ & \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2 \end{aligned}$$

Also:

$$SS \text{ error a} = \left(\frac{1}{b} \right) Y'_{i.k.m} Y_{i.k.m} - \left(\frac{1}{t_{be}} \right) Y'_{i...} Y_{i...}$$

$$- \left(\frac{1}{t_{be}} \right) Y'_{j...} Y_{j...} - \left(\frac{1}{t_b} \right) Y'_{..k.m} Y_{..k.m} + \left(\frac{2}{t^2_{be}} \right) Y'Y$$

$$SS B = \left(\frac{1}{t^2_e} \right) \sum_{l=1}^b Y_{...l}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2$$

$$= \left(\frac{1}{t^2_e} \right) Y'_{...l} Y_{...l} - \left(\frac{1}{t^2_{be}} \right) Y'Y$$

To calculate SS AxB, SS TREAT2 must first be obtained as follows: SS TREAT2= SS A+SS B+SS AxB; So: SS AxB= SS TREAT2-SS A-SS B. Where:

$$\begin{aligned} SS \text{ TREAT2} &= \left(\frac{1}{t_e} \right) \sum_{k=1}^t \sum_{l=1}^b Y_{..kl}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2_{be}} \right)^2 \\ &= \left(\frac{1}{t_e} \right) Y'_{..kl} Y_{..kl} - \left(\frac{1}{t^2_{be}} \right) Y'Y \end{aligned}$$

As SS TREAT3= SS E+SS B+SS ExB; SS ExB= SS TREAT3- S E-SS B; but

$$SS\ TREAT3 = \left(\frac{1}{t^2} \right) \sum_{l=1}^b \sum_{m=1}^e Y_{...lm}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2 b e} \right)^2$$

$$= \left(\frac{1}{t^2} \right) Y'_{...lm} Y_{...lm} - \left(\frac{1}{t^2 b e} \right) Y' J Y$$

SS TREAT4= SS E+SS A+SS B+SS ExA+SS ExB+SS AxB+SS ExAxB. Thus: SS ExAxB= SS TREAT4-SS E-SS A-SS B-SS ExA-SS ExB-SS AxB.

For direct calculation, the following formula can be used:

$$SS\ ExAxB = \left(\frac{1}{t} \right) \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{.klm}^2 - \left(\frac{1}{tb} \right) \sum_{k=1}^t \sum_{m=1}^e Y_{.k.m}^2$$

$$- \left(\frac{1}{t^2} \right) \sum_{l=1}^b \sum_{m=1}^e Y_{...lm}^2 + \left(\frac{1}{t^2 b} \right) \sum_{m=1}^e Y_{...m}^2$$

$$- \left(\frac{1}{te} \right) \sum_{k=1}^t \sum_{l=1}^b Y_{.kl.}^2 + \left(\frac{1}{tbe} \right) \sum_{k=1}^t Y_{.k..}^2$$

$$+ \left(\frac{1}{t^2 e} \right) \sum_{l=1}^b Y_{...l.}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2 b e} \right)^2$$

$$= \left(\frac{1}{t} \right) Y'_{.klm} Y_{.klm} - \left(\frac{1}{tb} \right) \sum_{k=1}^t \sum_{m=1}^e Y_{.k.m}^2$$

$$- \left(\frac{1}{t^2} \right) Y'_{...lm} Y_{...lm} + \left(\frac{1}{t^2 b} \right) Y'_{...m} Y_{...m} - \left(\frac{1}{te} \right) Y'_{.kl.} Y_{.kl.}$$

$$+ \left(\frac{1}{tbe} \right) Y'_{.k..} Y_{.k..} + \left(\frac{1}{t^2 e} \right) Y'_{...l.} Y_{...l.} - \left(\frac{1}{t^2 b e} \right) Y' J Y$$

Finally: SS error b= SS total-SS E-SS H-SS C-SS A-SS ExA-SS error a-SS B-SS AxB-SS ExB-SS ExAxB.

The following is also valid:

$$SS\ error\ b = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}^2$$

$$- \left(\frac{1}{b} \right) \sum_{i=1}^h \sum_{k=1}^t \sum_{m=1}^e Y_{i.k.m}^2$$

$$-\left(\frac{1}{t}\right) \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{..klm}^2$$

$$+\left(\frac{1}{tb}\right) \sum_{k=1}^t \sum_{m=1}^e Y_{..k.m}$$

$$=Y'Y - \left(\frac{1}{b}\right) Y'_{i.k.m} Y_{i.k.m} - \left(\frac{1}{t}\right) Y'_{..klm} Y_{..klm} + \left(\frac{1}{tb}\right) Y'_{..k.m} Y_{..k.m}$$

SS subplots (SUB) = SS total - SS main plots (MP);

$$SS \text{ SUB} = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}^2$$

$$-\left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2 b e}\right)^2$$

$$\left\{ \left(\frac{1}{b}\right) \sum_{i=1}^h \sum_{k=1}^t \sum_{m=1}^e Y_{i.k.m}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}}{t^2 b e}\right)^2 \right\}$$

$$= \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^e Y_{ijklm}^2 - \left(\frac{1}{b}\right) \sum_{i=1}^h \sum_{k=1}^t \sum_{m=1}^e Y_{i.k.m}^2$$

$$Y'Y - \left(\frac{1}{b}\right) Y'_{i.k.m} Y_{i.k.m}$$

Discussion

When an experiment or a series of trials is designed and analyzed across years (A) localities (L) or in their combinations (AxL), at least three fundamental stages are implicit: a) the building or selection of a statistical model; b) the generation of an analysis of variance; and c) the application of a test for the comparison of arithmetic means of two different levels, in one, two or more factors under consideration (Sahagún, 1993, 1998, 2007).

In relation to this strategy, it is also advisable to apply one or more statistical packages to save time during the statistical analysis of the data. In the previous context, choosing the right software is also of great relevance to generate the required outputs easily and reliably (González *et al.*, 2023; González *et al.*, 2024a, 2024b).

The series of experiments that have been most frequently addressed in agronomic research are those that correspond to the completely randomized design (CRD) and randomized complete block design (RCBD), for combinatorial arrangements, in split plots (SPs), split-split plots (SSPs) or split blocks or strips (SSts) (Sahagún, 1993, 1994, 2007; Gomez and Gomez, 1984; Ledolter, 2010; Tirado and Tirado, 2017), but the LSD in SPs, SSPs or SSts is not well documented.

In the published literature, little evidence has been found in relation to the statistical reference model, as well as for an Analysis of Variance (Anova) and a comparison of means (Ledolter, 2010; Tirado and Tirado, 2017; <https://biometrics.ilri.org/Publication/Full%20Text/chapter20.pdf>), especially for subsampling (Gomez and Gomez, 1984; Martínez, 1988) Proc Anova: Latin Square Split Plot :: SAS/STAT(R) 9.22 User's Guide.

Martínez's (1994) study only presented a table with the sources of variation (SV) and the degrees of freedom (DF) that could be calculated without subsampling. Gomez and Gomez (1984) showed two tables with the SV and DF for an RCBD and for an SP arrangement with balanced subsampling, but for a single trial; additionally, they described the procedures to obtain an Anova in both situations, with emphasis on the estimation of the sampling (SE) and experimental (EE) errors, which in the present study make up the joint error (JE).

In the Statistical Analysis System (SAS), only the data provided by Smith (1951; Proc Anova: Latin Square Split Plot :: SAS/STAT(R) 9.22 User's Guide), but without subsampling, and a code to obtain an Anova with three types of errors were presented; in relation to the present research and for a single trial, CxHxA is equal to error a, and the sum of HxB and the residual of the model gives rise to error b. Tirado and Tirado (2017) presented the statistical model for SPs in an LSD without subsampling, and an example to generate their Anova; they also generalized this type of analysis for an arrangement in split-split plots in an LSD without subsampling.

In previous studies, Sahagún (1993; 1994; 2007) addressed several situations in which he analyzed the trial series for the CRD and RCBD experimental designs and the latter in SP arrangement, this author analyzed their models, the expectations of the mean squares to perform the relevant hypothesis tests in an analysis of variance, as well as their efficiency, with and without restriction in these models. In addition, he considered two cases: a) when there is crossover between years, localities and genotypes; and b) when the years are nested within the localities, so that the genotypes may or may not be nested in the AxL interaction.

Rodríguez *et al.* (2025) findings considered the statistical model and formulas to calculate degrees of freedom and sum of squares by applying two methodologies to the specific case of a trial conducted in an experimental design in a Latin square in a split-plot arrangement, as a prerequisite to extend their analysis to a series of experiments with the application of software, when years and localities are confused within contrasting environments.

Therefore, the present research is an extension of their work, as a prerequisite to analyze data from this type of experiments when balanced subsampling is applied within the experimental units that make up each trial. Nevertheless, this type of trial will still need to be analyzed and discussed when, as Sahagún (1993, 1994, 2007) suggested, years and localities are considered as cross-factors or when years are nested within localities, particularly in the case of SPs, SSPs and SSTs in an LSD.

A complementary situation to what was previously considered was addressed by González *et al.* (2024b), who fractionated each of the replications in an arrangement of 'g' groups of balanced complete blocks for an RCBD, confusing years and localities in different environments, they extended the case presented by Gomez and Gomez (1984), for a single trial, for the analysis of 45 cultivars of rice (*Oryza sativa* L.) classified into three groups with 15 cultivars within each of these.

They also considered the possibility of pooling genetic material based on two criteria: a) that the groups be as heterogeneous as possible; and b) that, within each group, there were such small differences in physiological maturity in the fraction of material being considered, if there were no significant differences between groups of cultivars, the statistical analysis of the data can be performed simply as an RCBD or as a series of trials in this type of experimental design.

In the series of experiments that is being discussed in the present study, for an arrangement in SPs in an LSD without subsampling, it is assumed that the statistical model was built considering that years and localities are confused in contrasting environments (E) and the latter do not show interaction with rows (H), columns (C) or both, but E is crossed with factors A and B. Additionally, H, C or both do not interact with A, B or AB either. Finally, when InfoStat is applied, error a will be represented by the interactions ExHxA or ExCxX due to the restriction imposed on the main plots

($h=c=t$), related to the double blocking that is applied to the experimental units when the levels of factor A are randomized.

For a series of experiments in an RCBD, Sahagún (1993) considered three situations in relation to replications: a) they are nested within localities; b) they are nested within years and localities and c) they are nested within a factor C, corresponding to the combinations between A and L. In all three cases, it was recommended to consider replications as a random factor, just like A, L or C. In relation to genotypes, for selection purposes, he suggested considering them as a fixed-effect factor, but if the purpose is to estimate variance components, this should be defined as a random factor.

In the previous context, applying the guide proposed by Sahagún (1998), the same model will be built if the following conditions are established: a) rows (H) nested within columns (C) or vice versa; H, C or both nested within environments (E); b) E, A and B or their interactions are crossed; c) H, C or both are nested within B or AB; d) E is a random factor, but A and B can be fixed factors and e) the residual of the model is nested in all components of the model. In addition to the above, it should be considered whether or not there is a restriction on the components of this model (Sahagún, 1993).

For complementary purposes, it is suggested to consider the publications of Sahagún (1994 and 2007a, 2007b), who proposes to test the correct statistical hypotheses related to the components of the statistical models that he analyzed and discussed, using the appropriate mean squares. This last situation is also a critical point in the topic that is being considered in this research or in others that are now being used to generate, validate, apply or transfer technology, because when a statistical package is used, analyses of variance with incorrect F-values could be generated, as well as for some test of comparison of means for main effects or interactions that could not be valid as well, as suggested by Sahagún (1993, 1994, 1998, 2007); Gomez and Gomez (1984); Montgomery (2009).

The methodologies presented in this research will make it possible to standardize formulas to calculate degrees of freedom and sum of squares in an easy and reliable way, as González *et al.* (2023); González *et al.* (2024a, 2024b) have shown for other studies; with their symbology and that described in Mendenhall (1987); Sahagún (2007 a); Montgomery (2009), the quadratic or matrix forms that will most frequently feed matrix calculators or SAS, Agrobases, Spss, StatGraphics, Star, PB Tools will be generated easily and reliably.

Conclusions

In the proposed statistical model, it was assumed that years and localities are confused in contrasting environments (E) and that the latter do not interact with rows, columns, or both, but E is crossed with factors A and B. Additionally, it was considered that H, C or both do not interact with A, B or AB. Finally, error a will be represented by the interactions $ExHxA$ or $ExCxA$. To verify degrees of freedom and sum of squares, in both models, the alternative formulas that were constructed for errors a and b, as well as those corresponding to main plots and subplots, can be applied for each or both methodologies.

Bibliography

- 1 Balzarini, M. G. y Di-Rienzo, J. A. 2016. InfoGen Versión 2016. FCA. Universidad Nacional de Córdoba, Argentina. <http://www.info-Gen.com.ar>.
- 2 Balzarini, M. G.; González, L.; Tablada, M.; Casanoves, F.; Di-Rienzo, J. A. y Robledo, C. W. 2008. Manual del usuario de InfoStat. Ed. Brujas, Córdoba, Argentina. 348 p.
- 3 Di Rienzo, J. A.; Casanoves, F.; Balzarini, M. G.; González, L.; Tablada, M. y Robledo, C. W. 2008. InfoStat, versión 2008. Grupo InfoStat, FCA. Universidad Nacional de Córdoba, Argentina. 336 p. <https://www.infostat.com.ar>.
- 4 Gomez, K. A. y Gomez, A. A. 1984. Statistical procedures for agricultural research. 2nd Ed. John Wiley & Sons, Inc. Printed in Singapore. 680 p.

- 5 González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R. F.; Franco, M. J. R. P. y Padilla, L. A. 2019. InfoStat, InfoGen y SAS para contrastes mutuamente ortogonales en experimentos en bloques completos al azar en parcelas subdivididas. *Revista Mexicana de Ciencias Agrícolas*. 10(6):1417-1431.
- 6 González, H. A.; Pérez, L. D. J.; Balbuena, M. A.; Franco, M. J. R.; Gutiérrez, R. F. y Rodríguez, G. J. A. 2023. Submuestreo balanceado en experimentos monofactoriales usando InfoStat y InfoGen: validación con SAS. *Revista Mexicana de Ciencias Agrícolas*. 14(2):235-249. Doi: <https://doi.org/10.29312/remexca.v14i2.3418>.
- 7 González, H. A.; Pérez, L. D. J.; Hernández, A. J.; Franco, M. J. R. P.; Rubí, A. M. y Balbuena, M. A. 2024a. Tratamientos anidados dentro de un arreglo en grupos de bloques completos balanceados. *Revista Mexicana de Ciencias Agrícolas*. 15(2):e3634. Doi: <https://doi.org/10.29312/remexca.v15i2.3634>.
- 8 González, H. A.; Pérez, L. D. J.; Hernández, A. J.; Franco, M. J. R. P.; Balbuena, M. A. y Rubí, A. M. 2024b. Serie de experimentos para tratamientos anidados en grupos en arreglo de bloques completos balanceados. *Revista Mexicana de Ciencias Agrícolas*. 15(7):1-15.
- 9 Ledolter, J. 2010. Split-plot design: discussion and examples. *International Journal of Quality Engineering and Technology*. 1(4):441-457. Doi: <https://doi.org/10.1504/IJQET.2010.035588>.
- 10 Little, T. M. y Hills, F. J. 2008. Métodos estadísticos para la investigación en la agricultura. Ed. Trillas, SA de CV. México, DF. ISBN:978-968-24-3629-1. 270 p.
- 11 Martínez, G. A. 1988. Diseños experimentales. métodos y elementos de teoría. Editorial. Trillas, 1ª Ed. México, DF. 756 p.
- 12 Martínez, G. A. 1994. Experimentación agrícola: métodos estadísticos. Primera Edición. Universidad Autónoma Chapingo (UACH). Chapingo, Estado de México. ISBN:9789 6888 42 867. 357 p.
- 13 Mendenhall, W. 1987. Introducción a la probabilidad y la estadística. Grupo Editorial Iberoamérica. 1ª Ed. México, DF. 626 p.
- 14 Montgomery, D. C. 2009. Design and analysis of experiments. 7th Ed. John Wiley & Sons, Inc. USA. 656 p.
- 15 Pérez, L. D.; Jasso, B. G.; Saavedra, G. C.; Franco, M. J. R. P.; Ramírez, D. J. F.; González, H. A. 2022. Uso de artificios en Opstat para analizar series de experimentos en dialéxico parcial. *Revista Mexicana de Ciencias Agrícolas*. 13(2):273-287. Doi: <https://doi.org/10.29312/remexca.v13i2.3130>.
- 16 Rodríguez, G. J. A.; Pérez, L. D. J.; Hernández, A. J.; Balbuena, M. A.; Franco, M. J. R. P. y González, H. A. 2025. Parcelas divididas en cuadro latino: modelos estadísticos y fórmulas, sin y con submuestreo. *Revista Mexicana de Ciencias*. 16(2):1-11. Doi: <https://doi.org/10.29312/remexca.v16i2.3926>.
- 17 Sahagún, C. J. 1993. Funcionalidad de cuatro modelos para las evaluaciones genotípicas en series de experimentos. *Revista Fitotecnia Mexicana*. 16(3):161-171.
- 18 Sahagún, C. J. 1994. Evaluación de genotipos en series de experimentos: diferencias en parámetros genéticos generados en dos modelos. *Revista Fitotecnia Mexicana*. 17(2):116-125.
- 19 Sahagún, C. J. 1998. Construcción y análisis de los modelos fijos, aleatorios y mixtos. Departamento de Fitotecnia. Programa Nacional de Investigación en Olericultura. Universidad Autónoma Chapingo. Boletín técnico núm. 2. 64 p.
- 20 Sahagún, C. J. 2007b. Evaluación de genotipos en heterogeneidad meteorológica intrarregional: confusión vs Anidamiento de años en localidades. *Revista Fitotecnia Mexicana*. 30(1):97-104.

- 21 Sahagún, C. J. 2007a. Estadística descriptiva y probabilidad: una perspectiva biológica. 2ª Ed. Universidad Autónoma Chapingo (UACH). México, DF. ISBN: 978-968-02-035-4. 282 p.
- 22 Smith, W. G. 1951. Dissertation notes on Canadian sugar factories Ltd, Taber, Alberta, Canada 7 p. <https://searcharchives.vancouver.ca/fire-insurance-plans-for-bc-sugar-canadian-sugar-factories-and-manitoba-sugar-co>.
- 23 Tirado, E. G. y Tirado, G. D. N. 2017. Tratado de estadística experimental. Primera Ed. Editorial Centro de Estudios e Investigaciones para el Desarrollo Docente (CENID, AC), Guadalajara, Jalisco, México. 265 p. ISBN: 978-607-8435-43-2.



Formulas for series of experiments in a Latin square design in a split-plot arrangement

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 01 July 2025
Date accepted: 01 September 2025
Publication date: 25 September 2025
Publication date: Aug-Sep 2025
Volume: 16
Issue: 6
Electronic Location Identifier: e3861
DOI: 10.29312/remexca.v16i6.3861

Categories

Subject: Articles

Keywords:

Keywords:

matrices

statistical models

years and localities confused in environments

Counts

Figures: 0

Tables: 0

Equations: 44

References: 23

Pages: 0