Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Article

Series of experiments for treatments nested in groups with balanced complete block arrangement

Andrés González-Huerta[1,§]
Delfina de Jesús Pérez-López[1]
Jesús Hernández-Ávila[1]
J. Ramón Pascual Franco-Martínez[1]
Artemio Balbuena-Melgarejo[1]
Martín Rubí-Arriaga[1]

1 Centro de Investigación y Estudios Avanzados en Fitomejoramiento (CIEAF)-Facultad de Ciencias Agrícolas-Universidad Autónoma del Estado de México-Campus Universitario "El Cerrillo". El Cerrillo, Piedras Blancas, Toluca de Lerdo, Estado de México, México. AP. 435. Tel. 722 2965518, ext. 148. (djperezl@uaemex.mx; jhernandeza@uaemex.mx; jrfrancom@uaemex.mx; mrubia@uaemex.mx; abalbuenam@uaemex.mx).

Autor para correspondencia: agonzalezh@uaemex.mx.

Abstract

The design and analysis of various series of experiments across years, locations, or both have been of great relevance in agronomic research. This study extrapolates, across environments, the case presented by Gomez and Gomez (1984) in relation to the grain yield recorded in 45 varieties of rice nested in three groups in a balanced complete block arrangement; its statistical model is built for an experimental design of randomized complete blocks, formulas to calculate the sums of squares are included and the procedure to generate an output if InfoGen is applied is proposed. The procedures are based on environments, groups, environments x groups, replications within environments, error a, environments x treatments within groups, treatments within groups, and error b; the first five components define the main unit, and the remaining ones are the subunit. In addition, other ways to calculate degrees of freedom for the main unit and the subunit, as well as those corresponding to the residual of the model or error b, are mentioned, which simplify manual calculations. The difference between a conventional analysis of variance and the one considered in this work, based on the sums of squares, is discussed; finally, it is indicated how to apply Tukey's test for the comparison of means of varieties within each group if this software, InfoStat or SAS are used.

Keywords:

InfoStat, random complete block design, sum of squares and quadratic formulas.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Introduction

The design, analysis, and interpretation of data from an experiment, or a series of experiments recorded in years, localities, or both, has become an essential tool in the agricultural sciences when using a completely randomized experimental design, random complete blocks, Latin square, or some type of lattice; nevertheless, the first two have been more used (Martínez, 1988; Sahagún, 1993; Sahagún, 2007).

In a conventional RCBD, in each trial, the treatments are randomly assigned independently in each replication so that all the blocks are perpendicular to the prevailing variability gradient, such as slope or fertility level in the soil. Generally, the blocks are the same size, and each treatment is assigned only once within each of them. Whenever possible, the differences within each block should be minimal, and the heterogeneity between them should tend to be maximized for this experimental design to be efficient (Gomez and Gomez, 1984; Martínez, 1988; Little and Hills, 2008; Montgomery, 2009).

In homogeneous experimental areas, it is not justifiable to use an RCBD; with two gradients of variability, one perpendicular to the other, a Latin square or a lattice would have to be chosen, depending on how large the number of treatments and repetitions to be evaluated are; in partially balanced lattices, more than 30 treatments are frequently evaluated (Cochran and Cox, 1954; Gomez and Gomez, 1984; Martínez, 1998).

In the above context, there would be two possibilities for an RCBD: without and with a balanced complete block arrangement. Gomez and Gomez (1984); Shikari *et al*. (2015); Maranna *et al*. (2021) address the concepts and application of the second possibility: each replication is subdivided into groups of treatments that share some similarity within them and that differ significantly between them.

In forming these groups, the biological cycle of the cultivars or their plant height could be considered; however, this grouping could also be done considering their genetic or geographical origin, as suggested by González *et al*. (2008, 2010, 2011), among others. According to González *et al*. (2024), for a single trial using a randomized complete block experimental design in a balanced complete block arrangement (RCBD-BCBA), the experimental area could be divided into main unit (MU) and subunit (SUB), in such a way that the former includes the groups (G), the replications (R), and the error a, which is equivalent to the GxR interaction, while the second must contain treatments nested within G, [T(G)] and the error b.

Due to the close relationship between a single-factor experiment and factorials using the same experimental design, this approach could also be validated for series of experiments in RCBD-BCBA, as can be inferred from González *et al*. (2024). In the previous reference framework, it is necessary to extrapolate this type of study to the case of years, localities, or both in a series of experiments across environments as a prerequisite to evaluate new material in a genetic improvement program, as well as to validate, apply, generate, or transfer technology to producers' fields. Thus, the main objective of this research was to generate the statistical model and formulas to calculate degrees of freedom and sums of squares for an arrangement of experimental units, such as the one mentioned above.

Materials and methods

Preliminary concepts

The plant breeder collects, evaluates, and identifies outstanding biological material in a plant breeding program by applying some appropriate genetic and experimental design across several locations and a few years to make the results more reliable. This situation must be realistically represented in a statistical model, which must be independent of the availability of electronic equipment infrastructure; the plant breeder or their advisory team is responsible for its correct construction.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

In this, it must be determined whether the factors are fixed or random and whether there is crossing or nesting between them. The relationship between the model's components, an analysis of variance, and a comparison of treatment means through the estimation of effects or variances should also be considered (Sahagún, 1998). Before defining the types of models that are frequently used in the different branches of agricultural sciences, consider the case of the Cacahuacintle corn breed.

A population of landrace varieties of this breed could be determined by the number of farmers who sow it in the municipality of Calimaya de Díaz González, in the State of Mexico. If there are 2 500 farmers in this municipality, then there will be 2500 landraces, assuming that each of them has a different variety. A model is fixed-effect if the researcher considers a random and representative sample of the 2 500 varieties, for example a sample of 30, and the researcher estimates the differences that exist only between them, but if from them, the researcher makes inferences towards the entire population of landraces, then the researcher will be choosing a model of random effects; in the first case, effects are estimated and in the second, variances.

Statistical models are built using both principles and these can be fixed, random, or mixed; the latter situation arises from the need to include both fixed and random components. As an example, consider the need to evaluate 30 varieties of Cacahuacintle in three locations in the Toluca Valley, in the State of Mexico, using three replications per treatment. If varieties and localities are considered fixed and random factors respectively, then one will have a mixed model.

In agronomic research, it is frequent to use models whose components are fixed and random, as in the series of experiments across time, space, or both that were discussed in Sahagún (1993, 2007), or as the one considered in Gomez and Gomez (1984); Maranna *et al*. (2021); González *et al*. (2024). Years, localities, or combinations between them generate random components in their statistical models (Sahagún, 1998).

In the present study, it is said that the factors E and G, used to identify environments and groups, respectively, are crossed when each level of E is combined with each level of G. The T-factor, which represents varieties or treatments, is nested in the G-factor if each T-level is combined with only one G-level. In the series of experiments across years, locations, or both, replications are also nested within them.

In addition, the plots or experimental units are nested in replications and localities. If one factor is nested in another, it is not possible to study their interaction. A dataset is balanced if the number of observations in each smaller cell that can be formed is constant (Sahagún, 1998). In completely randomized, complete randomized blocks, and Latin square experimental designs, there are balanced experiments when each treatment has the same number of replications and when there is the same number of observations in each plot or experimental unit; the latter situation is commonly related to subsampling in experimental designs.

Otherwise, there will be an unbalanced situation; without balance, the statistical analysis of the data is more complex. The mean square expectation is essential when one wants to rationalize the methodology used to test a hypothesis in an analysis of variance or to estimate variance components; these can be derived using results from the general linear model or generated directly. In this context, there are several publications that provide guides for the construction of statistical models or expectations of the mean square, such as in Sahagún (1998); Piepho *et al*. (2003); Restrepo (2007 a, b).

Statistical model

The model for a series of experiments to evaluate treatments nested within groups in a balanced complete block arrangement, in an experimental design of randomized complete blocks, was built based on the guide provided by Sahagún (1998); Piepho *et al*. (2003); Restrepo (2007 a). This is: $X_{ijkl}= \mu+ E_i + G_j + R_{k(i)} + (EG)_{ij} + (GR)_{jk(i)} + {}_{l(j)} + (ET)_{il(j)} + {}_{ijkl}$.

Where: X is the grain yield or any other quantitative variable, μ is the arithmetic mean of the ert data, E_i is the effect caused by the i-th environment, G_j is the effect caused by the j-th group, $R_{k(i)}$ is

DOI: https://doi.org/10.29312/remexca.v15i7.3831

the contribution of the k-th replication nested in the i-th environment, $(EG)_{ij}$ is the interaction of the ij levels of the factors E and G, $(GR)_{jk(i)}$ is the interaction of the j-th group with the k-th replication nested in the i-th environment, also called error a, $_{l(j)}$ is the effect caused by the l-th treatment nested within the j-th group, $(ET)_{il(j)}$ is the interaction between the levels of the factors E and T, the latter nested within groups, and $_{ijkl}$ is the residual of the model, also known as error b.

Symbology used to calculate the sum of squares

The classification variables in the previously constructed model are environments, groups, replications, and treatments, which have been identified with the subscripts i, j, k, l; their levels are e, g, r, t/g, respectively. In the present study, g= g, and both will be equivalent to s, the latter used by Gomez and Gomez (1984). The treatments are divided into g groups and the total observations are calculated as: er

$$(\frac{t}{g}+\frac{t}{g}+\frac{t}{g}+,\ldots,+\frac{t}{g})=\text{erg}(\frac{t}{g}) \quad = \quad ert$$

To simplify manual calculations and to standardize both methodologies, in some denominators of the formulas shown in the results section, g will be considered null, as suggested by González *et al*. (2023) when they applied subsampling within plots in single-factor trials in the completely randomized, randomized complete blocks, and Latino square experimental designs. In these formulas, the formal symbology described in Mendenhall (1987); Sahagún (2007); Montgomery (2009) was applied.

Software used

InfoGen is used to describe the procedure that will allow the application of the least squares technique to obtain the analysis of variance, but InfoStat (https://www.InfoStat.com.ar) or Sas (https://www.sas.com), among others, could also be used. The three statistical packages could be used to generate the comparison of intra-group treatment means with Tukey's test or honest least significant difference, and Opstat (http://14.139.232.166/opstat/default.asp) could also be applied for its validation, Sheoran *et al*. (1998).

Results

Formulas for calculating degrees of freedom (DF) and sum of squares (SS)

The formulas that will generate DF and SS in the analyses of variance of a series of experiments across environments for the type of experimental unit arrangement mentioned above are presented below and are an extension of those that were published for a single-factor experiment by González *et al*. (2024).

Formulas for calculating DF

DF total= ert -1. DF environments (E)= e-1. DF groups (G)= g-1. DF replications within E= e(r-1). DF E x G= (e-1) (g-1). DF error a= e(g-1) (r-1). DF treatments (T) nested in G= t-g. DF ExT(G)= (e-1) (t-g). DF error b= e(r-1) (t-g). If the experimental area is divided into main unit (MU) and subunit (SU) (González *et al*., 2024), their DFs would be, respectively, egr-1 and er(t-g). The sum of both is ert-1, which corresponds to DF total.

Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Formulas for calculating SS

$$SC\ \ total = \sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{ert} = Y'Y - \left(\frac{1}{ert}\right)Y'JY$$

$$\text{SC environments}(E) = \left(\frac{1}{rt}\right)\sum_{i=1}^{e} Y_{i...}^2 - \frac{\left(\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{ert} = \left(\frac{1}{rt}\right)Y'_{i...}Y_{i...} - \left(\frac{1}{ert}\right)Y'JY.$$

$$SC\ \ groups = \left(\frac{g}{ert}\right)\sum_{j=1}^{g} Y_{.j..}^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{ert} = \left(\frac{g}{ert}\right)Y'_{.j..}Y_{.j..} - \left(\frac{1}{ert}\right)Y'JY$$

$$SC\ \ ExG = \left\{\left(\frac{1}{rt}\right)\sum_{i=1}^{e}\sum_{j=1}^{g} Y_{ij..}^2 - \left(\frac{\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}}{ert}\right)^2\right\} - SS\ A - SS\ G.$$

$$SC\ \ R(E) = \left\{\left(\frac{1}{t}\right)\sum_{i=1}^{e}\sum_{k=1}^{r} Y_{i.k.}^2 - \left(\frac{\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}}{ert}\right)^2\right\} - SS\ A$$

$$SC\ \ error\ \ a = \left(\frac{g}{t}\right)\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r} Y_{ijk.}^2 - \left(\frac{1}{rt}\right)\sum_{i=1}^{a}\sum_{j=1}^{g} Y_{ij..}^2 - \left(\frac{1}{t}\right)\sum_{i=1}^{a}\sum_{k=1}^{r} Y_{i.k.}^2 + \left(\frac{1}{rt}\right)\sum_{i=1}^{a} Y_{i...}^2$$

In the previous formulas, (1/ert)Y'JY is equivalent to the correction factor used to adjust the sums of squares, Y will be a scalar formed by 270 rows and a column, Y', its transposed matrix, will be formed by a row and 270 columns, J will be a symmetrical matrix formed by 1s, built with 270 rows and 270 columns.

To calculate the first component of the previous formula, a table of double classification criteria must be constructed: the groups, identified with the subscript j, will be placed in the rows, and the environments and replications, represented with the subscripts i, k, respectively, will be placed in the columns. In this table, there will be ijk= egr= 2(3) (3)= 18 data, which implies adding over the subscript l, corresponding to each of the subsets of treatments that are being evaluated; the remaining five components must be calculated beforehand.

The subscript j, used to represent groups, should not be confused with the matrix of ones, identified as J; Y must also be differentiated as a variable from Y as a matrix. The SS of treatments nested within groups will be calculated as:

$$SS\ \ TREAT\ \ (G1) = \left(\frac{1}{er}\right)\sum_{l=1}^{t} Y_{.1.l}^2 - \left(\frac{g}{ert}\right)\left(\sum_{l=1}^{t} Y_{.1.l}\right)^2 = \left(\frac{1}{er}\right)Y'_{.1.l}Y_{.1.l} - \left(\frac{g}{ert}\right)Y'_{.1.l}\ J\ Y_{.1.l}$$

Revista Mexicana de Ciencias Agrícolas
DOI: https://doi.org/10.29312/remexca.v15i7.3831

$$SS\ \ TREAT\ \ (G2) = \left(\frac{1}{er}\right)\sum_{l=1}^{t} Y_{.2.l}^{2} - \left(\frac{g}{ert}\right)\left(\sum_{l=1}^{t} Y_{.2.l}\right)^{2} = \left(\frac{1}{er}\right) Y'_{.2.l} Y_{.2.l} \ - \ \left(\frac{g}{ert}\right) Y'_{.2.l}\ J\ \ Y_{.2.l}$$

$$SS\ \ TREAT\ \ (G3) = \left(\frac{1}{er}\right)\sum_{l=1}^{t} Y_{.3.l}^{2} - \left(\frac{g}{ert}\right)\left(\sum_{l=1}^{t} Y_{.3.l}\right)^{2} = \left(\frac{1}{er}\right) Y'_{.3.l} Y_{.3.l} \ - \ \left(\frac{g}{ert}\right) Y'_{.3.l}\ J\ \ Y_{.3.l}$$

The sum of squares of the g-th group will be calculated similarly:

$$SS\ \ TREAT\ \ (Gg) = \left(\frac{1}{er}\right)\sum_{l=1}^{t} Y_{.g.l}^{2} - \left(\frac{g}{ert}\right)\left(\sum_{l=1}^{t} Y_{.g.l}\right)^{2} \ = \ \left(\frac{1}{er}\right) Y'_{.g.l} Y_{.g.l} \ - \ \left(\frac{g}{ert}\right) Y'_{.g.l}\ J\ \ Y_{.g.l}$$

To verify that the previous calculations are correct, the total of all the SS of treatments within groups will be:

$$SS\ \ T(G): \left(\frac{1}{er}\right)\sum_{l=1}^{t} Y_{...l}^{2} - \frac{\left(\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^{2}}{ert} \ = \ \left(\frac{1}{er}\right) Y'_{...l} Y_{...l} \ - \ \left(\frac{1}{ert}\right) Y'JY$$

To define the sum of squares of the interaction between environments and treatments nested within groups, first establish the following relationship: SS Treat 5= SS E + SS G + SS ExG + SS T(G) + SS ExT(G). Where:

$$SS\ \ Treat\ \ 5 = \left(\frac{1}{r}\right)\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{l=1}^{t} Y_{ij.l}^{2} - \frac{\left(\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^{2}}{ert} = \ \left(\frac{1}{r}\right) Y'_{ij.l} Y_{ij.l} \ - \ \left(\frac{1}{ert}\right) Y'J\ \ Y$$

In this context:

SS E x T(G)= SS Treat 5-SS E-SS G-SS ExG-SS T(G).=

$$\left(\frac{1}{r}\right) Y'_{ij.l} Y_{ij.l} \ - \ \left(\frac{1}{rt}\right) Y'_{i...} Y_{i...} \ - \ \left(\frac{g}{ert}\right) Y'_{.j..} Y_{.j..} \ - \ \left(\frac{g}{rt}\right) Y'_{ij..} Y_{ij..} \ - \ \left(\frac{1}{er}\right) Y'_{...l} Y_{...l} \ + \ \left(\frac{3}{ert}\right) Y'\ \ J\ \ Y$$

Additionally: SS total= SS E + SS G + SS R(E) + SS ExG + SS error a + [SS TREAT (G1) + SS TREAT (G2) + SS TREAT (G3) + ,..., + SS TREAT (Gg)] + SS ExT(G) + SS error b.

So: SS error b= SS total- (SS E + SS G + SS R(E) + SS ExG + SS error a)- [SS TREAT (G1) + SS TREAT (G2) + SS TREAT (G3) +,..., + SS TREAT (Gg)] - SS ExT(G). If the experimental area is divided into main unit (PU) and subunit (SU) and, as proposed by González *et al*. (2024), it is defined that SS total= SS MU + SS SU, then the following expression will also be valid: SS MU= SS E + SS G + SS R(E) + SS ExG + SS error a. In the above context, the following equivalence is also correct:

$$SS\ \ MU = \ \ SS\ \ Treat\ \ 1 = \ \left(\frac{g}{t}\right)\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r} Y_{ijk.}^{2} - \frac{\left(\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^{2}}{ert} \ = \ \left(\frac{g}{t}\right) Y'_{ijk.} Y_{ijk.} \ - \ \left(\frac{1}{ert}\right) Y'JY$$

By difference: SS SU= SS total - SS MU. Therefore:

$$SS\ \ SU = \ \sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}^{2} - \left(\frac{g}{t}\right)\sum_{i=1}^{e}\sum_{j=1}^{g}\sum_{k=1}^{r} Y_{ijk.}^{2} = Y'Y - \left(\frac{g}{t}\right) Y'_{ijk.} Y_{ijk.}$$

DOI: https://doi.org/10.29312/remexca.v15i7.3831

It was observed that the sum of the SS MU and SS SUB should be equal to SS total, both for the least squares methodology and for quadratic or matrix forms. Both could be used to verify various manual calculations or to define an appropriate routine when applying several statistical packages.

Using InfoGen or InfoStat

The labels for the columns will be environments, groups, replications, treatments, and the response variable, which could be identified with E, G, R, and X, respectively. The 270 data will be captured in three groups, each with 15 varieties, for each of their three replications, in the same order in which Gomez and Gomez (1984) showed their data.

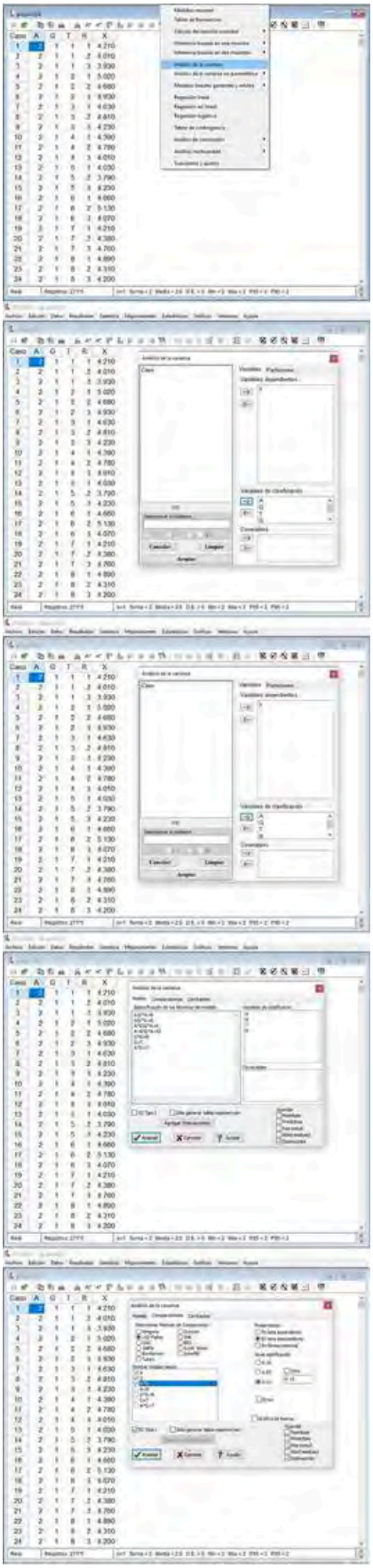
To build the database shown in InfoGen, some fictitious data were captured in order to show the procedure to be applied in this software (Balzarini *et al*., 2008; Di Rienzo *et al*., 2008; Balzarini and Di Rienzo, 2016). The statistical analysis must be generated in two stages: in the first, a general analysis of variance will be obtained with the division of effects into E, G, R(E), ExG, error a, T(G), ExT(G), and error b or residual of the model.

In the second stage, it will be indicated how to perform an analysis of variance by groups of treatments or for each environment. This same strategy will be applied to obtain the outputs corresponding to the comparison of means for each of the components of the linear model. Thus, the following was observed. Figure 1. Figure 2

Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Figure 1. General Anova and comparison of means for the main unit (MU) components.

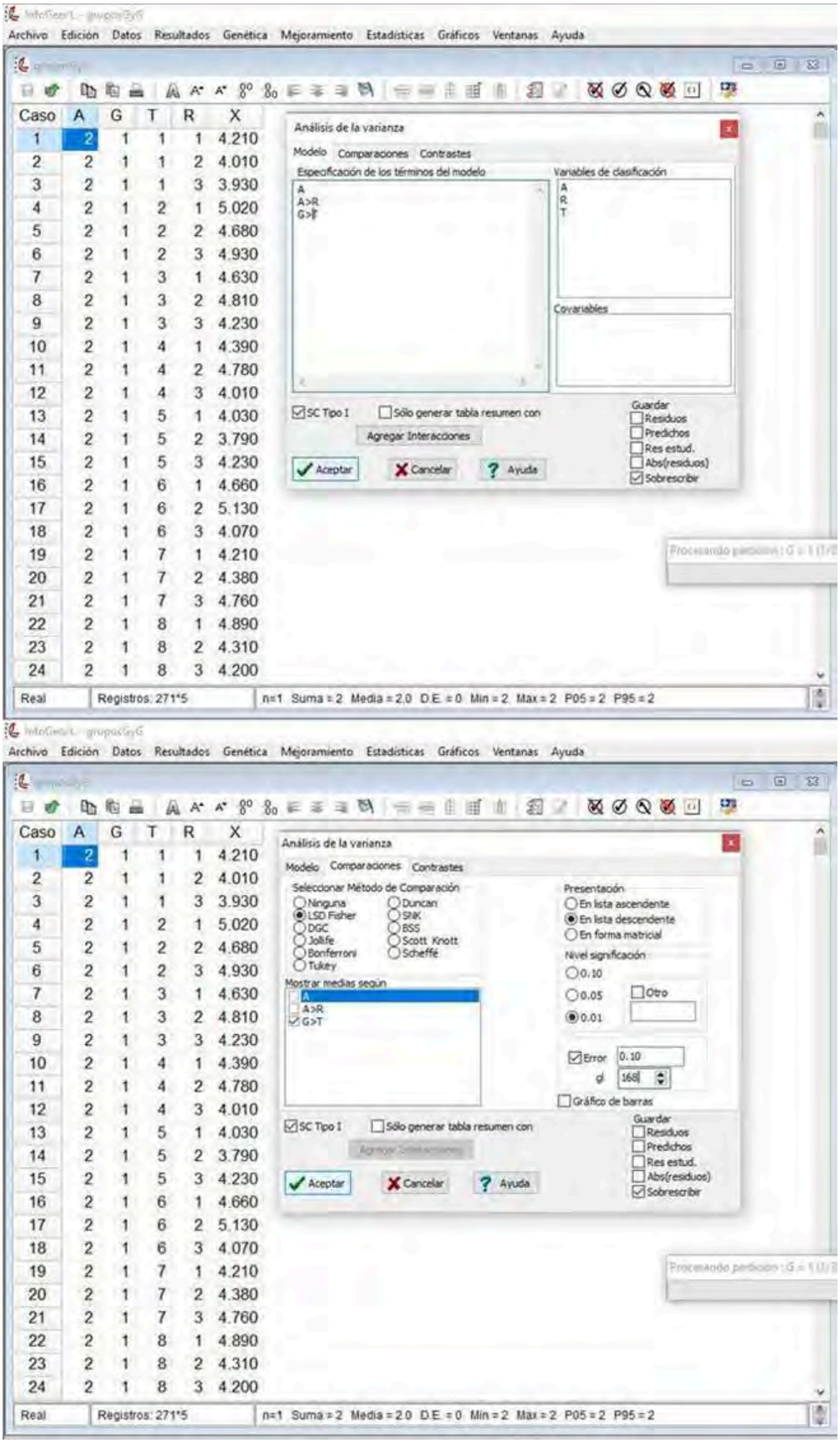


/

Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Figure 2. Anova and comparison of means for treatments nested within groups.



DOI: https://doi.org/10.29312/remexca.v15i7.3831

Discussion

The series of experiments across years, localities or both, in some of the basic experimental designs, have been of great relevance in agronomic research. The construction of their linear models for an experimental design in randomized complete blocks, the definition of the mathematical expectations of mean squares, and the appropriate hypothesis tests in the analysis of variance have been analyzed and discussed in Sahagún (1993); Sahagún (1994); Sahagún (2007).

Research conducted by Sahagún (1993) discussed the implications generated by the application of four linear models for the evaluation of various genotypes (G) across several years (Y) and localities (L) or when these are analyzed across environments generated with the combination of the levels of both classification factors, under a randomized complete block design (RCBD), recommended for use in annual crops.

In these models, he considered the years and localities to be random factors and defined the following: in model 1, Y, L, and G are crossed, and the replications (R) are nested within L; in model 2, R is nested within Y and L; in model 3, R is nested in Y and the latter is also nested in L; in model 4, he introduced the concept of confounding for factors Y and L, whose combination of levels generates another classification factor.

González *et al*. (2008, 2010, 2011) discussed some of the agronomic implications of the use of a series of experiments in RCBD to identify genetic material of high yield and phenotypic stability when evaluating a set of varieties and hybrids whose putative progenitors are the corn breeds: Cónico, Chalqueño, Cacahuacintle, Palomero Toluqueño, or racial complexes formed among some of these with germplasm of tropical or subtropical origin, from inbred lines formed by the International Maize and Wheat Improvement Center (CIMMYT, for its acronym in Spanish) and the National Institute of Forestry, Agricultural, and Livestock Research (INIFAP, for its acronym in Spanish), recommended for commercial sowing in the central region of Mexico.

In this context, the arrangement of balanced complete blocks for a randomized complete block design has also been relevant, which has been analyzed and discussed for a single trial by Gomez and Gomez (1984); Shikari *et al*. (2015); Maranna *et al*. (2021); González *et al*. (2024), among others, who suggested that this grouping could be carried out considering differences in plant height, biological cycle, grain yield or other important quantitative characteristic.

Cultivars could also be classified into subsets considering their geographic or genetic origin, as suggested by González *et al*. (2008); González *et al*. (2011), with and without arrangement in balanced complete blocks. In heterogeneous experimental areas, such as those that predominate throughout the Mexican Republic, with the proposal made by Gomez and Gomez (1984); Shikari *et al*. (2015); Maranna *et al*. (2021), statistical hypotheses for subsets of treatments would be tested more accurately compared to that conducted by González *et al*. (2008, 2010, 2011), other researchers who used an RCBD without BCBA.

Due to the existence of errors a and b: the first would be used to test hypotheses related to effects or variances for environments, groups, environments x groups, and replications within environments, while error b would be used to detect significant differences between treatments nested within groups and for the interaction of environments x treatments within groups.

In the above context, error a represents the interaction of groups x replications within environments, and error b is the residual of the linear model constructed and described in the present study. It could also be defined that error a is related to the main unit, and that error b is associated with the subunit, in the same way as proposed by González *et al*. (2024).

In González *et al*. (2008, 2010, 2011) or in multiple trials evaluating yield trials to assess the effects between treatments with another option, such as without or with the use of mutually orthogonal contrasts, these are tested with the residual of the model, which is equivalent to its experimental error; statistical significance in the F test for treatments within groups depends on whether or not there are statistical differences between and within groups in a BCBA-RCBD, both for one trial and for the series of experiments across environments.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

The summation and period notations have been very useful for manual calculations in various branches of statistics and probability and the analysis of agronomic experiments: their informal but easy and precise application can be consulted in Gomez and Gomez (1984); while Mendenhall (1987); Martínez (1988); Zamudio and Alvarado (1996); Sahagún (1998); Cochran and Cox (2004); Restrepo (2007a, b); Montgonery (2009), propose a more formal use to avoid confusion in their handling, particularly when the guidelines for the construction of fixed, random or mixed models will be applied or when the mathematical expectation of the mean squares will be defined to estimate components of variance.

Both notations are also very useful for homologating formulas generated with the least squares technique with those that can be derived from matrix or quadratic expressions (González *et al*., 2023; González *et al*., 2024). In addition to the symbology used, other aspects that cause confusion during calculations or in the handling of a statistical package are the absence of the linear model that was applied and the type of effects that are being evaluated; even though Gomez and Gomez (1984): Shikari *et al*. (2015); Maranna *et al*. (2021) did not present the linear model corresponding to a BCBA-RCBD trial, their results highlighted the relevance of this type of arrangement of experimental units in the design and analysis of agronomic experiments.

In this context, González *et al*. (2024) presented information complementary to that available in the previously referenced publications, and in the present study, a proposal is made to analyze the series of experiments across environments. Manual calculations are often considered a prerequisite for applying software.

In the previous context, in this study, the statistical model was homologated with the application of two methodologies to calculate degrees of freedom and sums of squares as a prerequisite to achieving the previously mentioned; InfoGen, InfoStat or SAS, among others, will be very useful to achieve this goal when the recommendations and suggestions made by Sahagún (1993); Sahagún (2007) are incorporated.

If the experimental area in the series of experiments conducted in BCBA-RCBD is divided into main unit (MU) and subunit (SU), as proposed by González *et al*. (2024), their degrees of freedom would be calculated as er g -1 and er (t-g), respectively, whose sum gives rise to the ert-1 degrees of freedom that correspond to the previous case and to a series of experiments conducted in an RCBD without BCBA. In addition, the total for degrees of freedom of treatments within groups would be the same as for each individual trial, that is,

$$\sum_{i=1}^{g}\left(\frac{t}{g}-1\right) = t - g$$

Thus, it will be easier to calculate the degrees of freedom for error b. González *et al*. (2019) fractionated the effects between treatments into groups in an RCBD without BCBA applying the technique of mutually orthogonal contrasts, but the precision with which the statistical hypotheses of interest are tested in a more heterogeneous experimental area could be more reliable using a BCBA-RCBD.

To verify the calculations related to the sums of squares (SS) that will be generated in the analysis of variance in a series of experiments with and without BCBA-RCBD, the outputs generated by this experimental design in both types of arrangements of experimental units must be compared. The SS T(G) plus the SS G must equal the SS of T. Also, the SS E*T(G) plus SS E*G will be equal to SS T*E, and, finally, the SS of the experimental error will be equal to the sum of the SS of errors a and b.

Authors such as Mendenhall (1987); Sahagún (1998); Montgomery (2009) pointed out that the analysis of variance is an essential part of facing the problem represented by the design and analysis of any experimental trial involving the calculation of degrees of freedom, sums of squares, and the construction of appropriate statistical tests considering the relationship that exists between

DOI: https://doi.org/10.29312/remexca.v15i7.3831

This problem has also been highlighted by Montgomery (2009); Restrepo (2007a, b); Piepho *et al*. (2003). González *et al*. (2023) emphasized correctly entering the instructions or procedures in the specification in the terms of the model in InfoStat, InfoGen or the SAS editor to adequately test statistical hypotheses related to experiments conducted, without and with subsampling within the experimental units, when applying the completely randomized and Latin square experimental designs; Zamudio and Alvarado (1996) made the same recommendation when they developed various codes for SAS to analyze the three experimental designs previously mentioned in balanced subsampling.

In the present study, the components of the MU will have to be tested using error a, while those corresponding to the SUB will use error b. For the comparison of means of varieties within groups, InfoStat and InfoGen are very flexible because, in both the database is automatically sorted, and both allow correcting the honest least significant difference or Tukey's test, but the degrees of freedom and the mean square of the error b, generated with all the data recorded in the series of experiments, must be captured manually; those corresponding to each trial will also be used to carry out this type of test independently.

If, in the series of experiments, the differences between treatment groups are not significant, InfoGen or InfoStat can generate an analysis of variance and a comparison of means with Tukey's test using the same database as when using an RCBA-RCBD. Their validation could be carried out with the Optat software available free of charge on its website, in which it is only necessary to capture the arithmetic means of each variety within each group, as well as the degrees of freedom and the mean square of the error b, which can be generated with any software or more easily, with a Microsoft Excel spreadsheet.

Conclusions

The statistical model considered in this study was built considering that environments and groups are crossed and that replications and treatments are nested within environments and groups, respectively. The formulas for calculating degrees of freedom and sums of squares will be simplified by dividing the experimental area into main unit and subunit: both contain errors a and b, respectively, the first is the interaction of groups x replications within environments, and the second is the residual of the model.

The least squares technique is easier to apply when using a statistical package, especially if the number of experiments and variables to be analyzed is large. InfoGen and InfoStat are very flexible when applying Tukey's test to treatments nested within groups, averaging the values over environments and replications, because they allow correcting the honest least significant difference when manually capturing the degrees of freedom and the mean square of error b.

If the treatment groups in the BCBA-RCBD are statistically the same, the data could be analyzed as a series of experiments in RCBD using the same file; as an option to generate the same results, one can use the Optat Software, available free of charge on its website, or any other statistical package, such as SAS or Agrobase, among others.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Bibliography

1 Balzarini, M. G. y Di Rienzo, J. A. 2016. InfoGen. 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. Editorial Brujas. Córdoba, Argentina. 348 p.

3 Cochran, W. G. y Cox, G. M. 2004. Diseños experimentales. Editorial Trillas, SA. de CV. 6ta. Reimpresión. México, DF. 661 p.

4 Di Rienzo, J. A.; Casanoves, F.; Balzarini, M. G.; González, L.; Tablada, M. and Robledo, C. W. 2008. InfoStat Versión 2008. Grupo InfoStat, FCA. Universidad Nacional de Córdoba. Argentina. (https://www.infostat.com.ar).

5 Gomez, K. A. and Gomez, A. A. 1984. Statistical procedures for agricultural research. 2nd Ed. John Wiley & Sons, Inc. Printed in singapore. 680 p.

6 González, H. A.; Vázquez, G. L. M.; Sahagún, C. J. y Rodríguez, P. J. E. 2008. Diversidad fenotípica de variedades e híbridos de maíz en el Valle Toluca, Atlacomulco, México. Revista Fitotecnia Mexicana. 31(1):67-76.

7 González, H. A.; Pérez, D. J.; Sahagún, C. J.; Franco, O.; Morales, E. J.; Rubí, A. M.; Gutiérrez, F. y Balbuena, A. 2010. Aplicación y comparación de métodos univariados para evaluar la estabilidad en maíces del Valle de Toluca, Atlacomulco, México. Revista Agronomía Costarricense. 34(2):129-143.

8 González, H. A.; Pérez, L. D. J.; Franco, M. O.; Nava, B. E. B.; Gutiérrez, R. F.; Rubí, A. M. y Castañeda, V. A. 2011. Análisis multivariado aplicado al estudio de las interrelaciones entre cultivares de maíz y variables agronómicas. Revista Ciencias Agrícolas Informa. 20(2):58-65.

9 González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R. F.; Franco, M. J. R. 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.

10 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.

11 González, H. A.; Pérez, L. D. J.; Hernández, A. J.; Franco, M. J. R.; Rubí, A. M. y Balbuena, M. A. 2024. Tratamientos anidados dentro de grupos en arreglo de bloques completos balanceados. Revista Mexicana de Ciencias Agrícolas. 15(2)e3634.

12 Little, T. M. y Hills, F. J. 2008. Métodos estadísticos para la investigación en la agricultura. Editorial Trillas SA. de CV. México. 270 p.

13 Maranna, S.; Nataraj, V.; Kumawat, G.; Chandra, S.; Rajesh, V.; Ramteko, R.; Manohar, P. R.; Ratnaparkhe, M. B.; Husain, S. M.; Gupta, S. and Khandekar, N. 2021. Breeding for higher yield, early maturity, wider adaptability and wáterlogging tolerance in soybean (Glycine max L.): a case study. Scientific reports. 11:22853. https://doi.org/10.1038/s41598-021-02064-x.

14 Martínez, G. A. 1988. Diseños experimentales. Métodos y elementos de teoría . Editorial Trillas, 1ra. Ed. México. 756 p.

15 Mendenhall, W. 1987. Introducción a la probabilidad y la estadística. Grupo Editorial Iberoamérica. 1ra. Ed. México. 626 p.

16 Montgomery, D. C. 2009. Design and analysis of experiments. 7th Ed. John Wiley (Sons, Inc. USA. 656 p.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

17 Piepho, H. P.; Büsche, A. and Enrich, K. 2003. A Hitchhiker's guide to mixed models for randomized experiments. J. Agron. Crop Sci. 189(2):310-322.

18 Restrepo, L. F. 2007a. Diagramas de estructuras en el análisis de varianza. Revista Colombiana de Ciencias Pecuarias. 20(1):202-208.

19 Restrepo, B. L. F. 2007b. La esperanza del cuadrado medio. Revista Colombiana de Ciencias Pecuarias. 20(2):193-201.

20 Sahagún, C. J. 1993. Funcionalidad de cuatro modelos para las evaluaciones genotípicas en series de experimentos. Revista Fitotecnia Mexicana. 16(3):161-171.

21 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.

22 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.

23 Sahagún, C. J. 2007. Evaluación de genotipos en heterogeneidad meteorológica intrarregional: confusión *vs* anidamiento de años en localidades. Revista Fitotecnia Mexicana. 30(1):97-104.

24 Sahagún, C. J. 2007. Estadística descriptiva y probabilidad: una perspectiva biológica. 2da. Ed. Universidad Autónoma Chapingo, México. 282 p.

25 Sheoran, O. P.; Tonk, D. S.; Kaushik, L. S.; Hasija, R. C. and Pannu, R. S. 1998. Statistical software package for agricultural research workers. Recent advances in information theory, statistical (computer applications by DS. Hooda (RC. Hasija Department of Mathematics Statistics, CCS HAU, Hisar. 139-143 pp.

26 Shikari, A. B.; Pourray, G. A.; Sofi, N. R.; Hussain, A.; Dar, Z. A. and Iqbal, A. M. 2015. Group balanced block design for comparisons among oilseed Brassicae. Academic Journals. 10(8):302-305. https://doi.org/10.5897/SRE2014.5792.

27 Zamudio, S. F. J. y Alvarado, S. A. A. 1996. Análisis de diseños experimentales con igual número de submuestras. 1ra. Ed. División de Ciencias Forestales. Universidad Autónoma Chapingo. Texcoco, Estado de México, México. 85 p.

Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Series of experiments for treatments nested in groups with balanced complete block 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 April 2024
Date accepted: 01 August 2024
Publication date: 01 November 2024
Publication date: Oct-Nov 2024
Volume: 15
Issue: 7
Electronic Location Identifier: e3831
DOI: 10.29312/remexca.v15i7.3831

Categories

Subject: Articles

Keywords:

Keywords:

InfoStat
random complete block design
sum of squares and quadratic formulas

Counts

Figures: 2
Tables: 0
Equations: 17
References: 27
Pages: 0