Article

Balanced subsampling in single-factor experiments using InfoStat and InfoGen: validation with SAS

Andrés González-Huerta[1][§]
Delfina de Jesús Pérez-López[1]
Artemio Balbuena-Melgarejo[1]
J. Ramón Pascual Franco-Martínez[1]
Francisco Gutiérrez-Rodríguez[1]
José Antonio Rodríguez-González[2]

[1]Center for Research and Advanced Studies in Plant Breeding-Faculty of Agricultural Sciences-Autonomous University of the State of Mexico. University *Campus* 'El Cerrillo', El Cerrillo Piedras Blancas, Toluca, State of Mexico, Mexico. AP. 435. Tel. 722 2965531, ext. 148. (djperezl@uaemex.mx; abalbuenam@uaemex.mx; jrfrancom@uaemex.mx; fgrfca@yahoo.com) [2]Postgraduate in Agricultural Sciences and Natural Resources-Autonomous University of the State of Mexico. (rodyof07@outlook.com).

[§]Corresponding author: agonzalezh@uaemex.mx.

Abstract

Even today there is little published information regarding the analysis of single-factor experiments when an equal number of subsamples are used within each experimental unit. This study analyzes male flowering data recorded in four varieties of corn (*Zea mays* L.) established under field conditions using four repetitions per treatment, 30 data were recorded within each experimental unit, but for the present study only three of these are considered. The experimental designs selected were completely randomized, randomized complete blocks and Latin square. The outputs were obtained with InfoStat and correspond to an analysis of variance and a comparison of means of treatments with the Tukey test ($p= 0.01$), and these can also be generated with InfoGen applying the same procedure. The data leading to both results were used for manual calculations and the results are validated with the statistical analysis system. Because the data are the same, the sampling error is common in the three experimental designs and it is shown how to obtain the joint error, the difference between the two generates the experimental error. To simplify the procedure on the personal computer, a single database is produced. Only for the case of the Latin square design, the matrix expressions that allow homologating the manual calculation with sums of squares in the analysis of variance are provided. If the secondary objective were to compare the three experimental designs, the statistical support generated by them would allow it, in a single run using SAS and individually for each design applying InfoStat and InfoGen.

Keywords: free statistical packages, matrix algebra, sampling error in experimental designs, sum of squares in trials with subsampling.

Introduction

When an experimental design is applied in the agricultural and forestry sciences, totals or arithmetic averages are used to test statistical hypotheses related to the questions that arise in relation to the structure of treatments being evaluated (Zamudio and Alvarado, 1996; Sahagún, 1998; Restrepo, 2007a, 2007b). Through analyses of variance, a partition of effects or variances related to the sources of variability that are implicit in the genetic-statistical models that are of interest to users is made (Sahagún, 1991; Sahagún, 1998; Piepho *et al.*, 2003; Restrepo, 2007a, 2007b).

In annual species, such as small-grain cereals, more than two plants or various of their parts are often measured and quantified within each experimental unit to improve or increase the accuracy with which these hypotheses are tested, the sample size, more repetitions, treatments or both, as well as better local control in the experimental area, among others, contribute to that purpose.

In this situation, users will be able to use each of the observations available within each experimental plot or unit (Gomez and Gomez, 1984; Martínez, 1988; Zamudio and Alvarado, 1996). In books, papers, theses, technical brochures or other sources of reliable information, it has been observed that little information exists when subsampling is carried out within the experimental units, especially when applying a statistical package (Martínez, 1988; Freund and Wilson, 1993; Zamudio and Alvarado, 1996). As the sums of squares can be obtained with two methodologies, the other drawback would be to present the homologation between the formulas that generate them, especially when they are more complex.

Zamudio and Alvarado (1996) presented an algorithm to perform subsampling in experimental designs through the application of matrix algebra, they emphasized the fact of selecting the correct statistical model and giving priority to the variability contained within the experimental unit, the variability that exists in each of the latter has two main components: one fixed and one random. They performed the statistical analysis in two stages through codes for SAS, to independently analyze the data to generate an analysis of variance in the completely randomized (CRD), randomized complete blocks (RCBD) and Latin square (LSD) experimental designs.

In the present study, the main objective was to apply InfoStat and InfoGen to analyze data from an experiment with balanced subsampling within each experimental unit in the CRD, RCBD and LSD designs, the secondary objective was to validate the manual calculations and the outputs generated with both statistical packages using SAS.

Materials and methods

Biological material

This study considered four varieties of corn (*Zea mays* L.) evaluated in the field in 2010 on a plot of land of the Faculty of Agricultural Sciences of UAEMéx: the cv. Ixtlahuaca, Cónico race, a variety of the Cacahuacintle race, a native population of the Palomero Toluqueño race and the Cóndor hybrid. These and other corn materials have been evaluated by González *et al*. (2008, 2010), but the data considered in this research were not published.

Experimental design and plot size

The trial was sown in the field in a 4 x 4 Latin Square experimental design. Each experimental plot (EU) consisted of 6 m in length, with a separation between furrows of 0.80 m. There were three rows of plants for each EU and within each of the latter, data were recorded in 30 plants, but only three observations will be considered.

Statistical models and experimental designs

For the construction of the conventional models, as well as those that include subsampling, described below, the guidelines provided by Sahagún (1998); Piepho *et al.* (2003); Restrepo (2007a, 2007b) can be consulted. The models are:

For CRD: $Y_{ikl} = \mu + \tau_k + \delta_{kl} + \varepsilon_{ikl}$
For RCBD: $Y_{ikl} = \mu + H_i + \tau_k + \delta_{kl} + \varepsilon_{ikl}$
For LSD: $Y_{ijkl} = \mu + H_i + C_j + \tau_k + \delta_{kl} + \varepsilon_{ijkl}$

Where: Y= the response variable; μ= the overall mean; τ_k= the effect caused by the k-th variety; H_i, C_j= the environmental heterogeneity that exists between rows and between columns. The δ's and ε's are the sampling and experimental errors, respectively: both will determine the joint error. Zamudio and Alvarado (1996) describe the first two models and how to estimate their components with matrix algebra.

Arithmetic-algebraic-matrix procedures

Analysis of variance (Anova)

The stages (E's) that will allow the verification of manual calculations are: E1) concentrate the data as shown in Table 1, obtain subtotals and totals.

Table 1. Data for male flowering considering the number of observations recorded in each treatment (number within parentheses), in each combination of row and column.

Row (i)	Sample (l)	Columns (j) 1	2	3	4	Total
1	1	95	85	93	105	
1	2	96(3)	86(4)	95(2)	104(1)	
1	3	98	83	96	105	
		289	254	284	314	1 141
2	1	85	92	104	94	
2	2	83(4)	94(3)	105(1)	92(2)	
2	3	84	96	106	95	
		252	282	315	281	1 130
3	1	93	104	87	95	

Row (i)	Sample (l)	Columns (j) 1	2	3	4	Total
3	2	90(2)	107(1)	85(4)	94(3)	
3	3	92	106	86	96	
		275	317	258	285	1 135
4	1	106	92	95	87	
4	2	105(1)	93(2)	93(3)	84(4)	
4	3	104	94	97	87	
		315	279	285	258	1 137
Subtotal		1 131	1 132	1 142	1 138	4 543

Totals for: 1= Cóndor= 1261; 2= Ixtlahuaca=1119; 3= Cacahuacintle= 1141; 4= Palomero Toluqueño= 1022.

E2) develop the format of Anova with subsampling. E3) calculate degrees of freedom (DF). In an LSD, the number of treatments (T), repetitions (R), rows (H), and columns (C) is equal; no T should be repeated in H or C. If S is the sample size recorded within each experimental unit, then: T= R= H= C= 4 and S= 3. The common DF in the three experimental designs are calculated as: Total DF= trs - 1= ths - 1= tcs - 1= 4 (4)(3) - 1= 47; DF T= DF R= DF H= DF C= t - 1= r - 1= h - 1= c - 1= 4-1= 3; DF sampling error (SE)= tr (s-1 = th(s-1)= tc (s-1)= 4(4)(3-1)= 32.

Now, what depends on the selected experimental design will be calculated. The joint error (JE) is the sum of the SE and the experimental error, the latter will be identified as EE, the following will be obtained:

CRD: DF JE= Total DF - DF T= 47 - 3= 44
DF EE= t(r-1)= t(h-1)= t(c -1)= 4(4-1)= 12
For verification: DF JE= DF SE + DF EE, thus: DF JE= 32 + 12= 44
RCBD: DF JE= total DF - DF R - DF T= 47 - 3 - 3= 41
DF EE= (t-1)(r-1)= (t-1)(h-1)= (t-1)(c-1)= 3(3) = 9. For verification: DF JE= DF SE + DF EE= 32 + 9 = 41
LSD: DF JE= Total DF - DF H - DF C - DF T= 47 - 3 - 3 - 3= 38
DF EE= (t-1)(t-2)= (h-1)(h-2)= (c-1)(c-2)= (4-1)(4-2)= 6. For verification: DF JE= DF SE + DF EE= 32 + 6= 38

E4) calculate sum of squares (SS). First, the SS that can be homologated in the three experimental designs are calculated. Correction factor (CF)= $\frac{(\sum_{i=1}^{h} \sum_{j=1}^{c} \sum_{k=1}^{t} \sum_{l=1}^{s} Y_{ijkl})^2}{hts}$ = $\frac{(95+95+95+,\ldots,+87)^2}{4(4)(3)}$ = 429 976.02; Total SS= $\sum_{i=1}^{h} \sum_{j=1}^{c} \sum_{k=1}^{t} \sum_{l=1}^{s} Y_{ijkl}^2$ -CF= $\left(95^2+95^2+95^2+,\ldots, +87^2\right)$- $\frac{(95+95+95+,\ldots,+87)^2}{4(4)(3)}$ = 432 483 - 429 976.02= 2506.98

As in the previous section, now H or C could be considered as R in a RCBD, in the denominator of the formulas shown in this section one of these will be null, but in the summations, these will be considered as appropriate. For this reason, the denominator of the CF includes h or c, but not both.

$$\text{SS T} = \frac{\sum_{k=1}^{t} Y^2_{..k.}}{hs} - CF = \frac{(1261^2 + 1119^2 + 1141^2 + 1022^2)}{4(3)} - \frac{(95+95+95+,...,+87)^2}{4(4)(3)} = 432\,387.25 - 429\,976.02 = 2411.23;$$

$$\text{SS R} = \text{SS H} = \frac{\sum_{i=1}^{h} Y^2_{i...}}{ts} - CF = \frac{(1141^2 + 1130^2 + 1135^2 + 1137^2)}{4(3)} - \frac{(95+95+95+,...,+87)^2}{4(4)(3)} = 429981.249 - 429\,976.02 =$$

$$5.2291; \text{SS SE} = \sum_{i=1}^{h} \sum_{j=1}^{c} \sum_{k=1}^{t} \sum_{l=1}^{s} Y^2_{ijkl} - \frac{\sum_{i=1}^{h} \sum_{k=1}^{t} Y^2_{i.k.}}{s}$$

To calculate this SS, Table 2 must be generated.

Table 2. Subtotals and totals to calculate SS SE.

Row (i)	Treatment (k)				Total
	1	2	3	4	
1	314	284	289	254	1 141
2	315	281	282	252	1 130
3	317	275	285	258	1 135
4	315	279	285	258	1 137
Total	1 261	1 119	1 141	1 022	4 543

With the data contained in Tables 1 and 2, the following is estimated: SS SE=
$(95^2 + 95^2 + 95^2 +,..., +87^2) - \frac{(314^2 + 284^2 + 289^2 +,...,+258^2)}{3} = 432\,483 - 432\,420.333 = 62.666$. Now, what is specific to each experimental design is calculated. For LSD: SS H = SS R; this was previously calculated. It was already commented that SS R could also be considered as SS C.

$$\text{SS C} = \frac{\sum_{j=1}^{c} Y^2_{j..}}{ts} - CF = \frac{(1131^2 + 1132^2 + 1142^2 + 1138^2)}{4(3)} - \frac{(95+95+95+,...,+87)^2}{4(4)(3)} = 429982.75 - 429976.02$$

= 6.73; SS JE = Total SS - SS H - SS C - SS T = 2506.98 - 5.2291 - 6.73 - 2411.23 = 83.7909
As: SS JE = SS EE + SS SE, then:
SS EE = SS JE - SS SE = 83.7909 - 62.666 = 21.1249
For RCBD: SS JE = Total SS - SS H - SS T = 2506.98 - 5.2291 - 2411.23 = 90.5209
Therefore: SS EE = SS JE - SS SE = 90.5209 - 62.67 = 27.8509
For: CRD: SS JE = Total SS - SS T = 2506.98 - 2411.23 = 95.75
Thus: SS EE = SS JE - SS SE = 95.75 - 62.67 = 33.08.

E5. The remaining calculations and hypothesis tests for H, C and T are conventionally obtained when considering arithmetic averages per experimental unit (see InfoStat outputs).

Matrix method

Total SS= $Y'Y - \left(\frac{1}{hts}\right) Y'JY$

SS H= $\left(\frac{1}{ts}\right) (Y'_i...)(Y_i...) - \left(\frac{1}{hts}\right) Y'JY$

SS C= $(\frac{1}{ts})$ (Y'.j..)(Y.j..) - $(\frac{1}{hts})$ Y'JY

SS T= $(\frac{1}{hs})$ (Y'..k.)(Y..k.) - $(\frac{1}{hts})$ Y'JY

SS SE= Y'Y - $(\frac{1}{s})$ (Y'i·k.)(Yi·k.)

SS JE= Y'Y - $(\frac{1}{ts})$ (Y'i...)(Yi...) - $(\frac{1}{ts})$ (Y'.j..)(Y.j..) - $(\frac{1}{hs})$ (Y'..k.)(Y..k.) + $(\frac{2}{hts})$ Y'JY

SS EE= $(\frac{1}{s})$ (Y'i·k.)(Yi·k.) + $(\frac{2}{hts})$ Y'JY -$(\frac{1}{ts})$ (Y'i...)(Yi...) - $(\frac{1}{ts})$ (Y'.j..)(Y.j..) - $(\frac{1}{hs})$ (Y'..k.)(Y..k.).

To simplify calculations, SS EE = SS JE - SS SE. In the matrices Y, Y', the three observations of treatment 1 in repetition 1, the three observations of treatment 1 in its repetition 2 and so on are captured. That is:

$$Y_{1x48} = \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \dots \\ 87 \\ 84 \\ 87 \end{bmatrix}$$

Y' is the transpose of the matrix Y; the matrix J, of 1's, has 48 rows and 48 columns. The remaining matrices are constructed with totals; their subscript(s) indicate the totals to be entered within them. In this section it might be useful to consult the publications of Jasso *et al*. (2022); Pérez *et al*. (2022), in which matrix calculations are performed, such as those shown below, with a statistical-genetic approach. Thus: total SS = Y'Y - $(\frac{1}{hts})$ Y'JY.

$$= [105\ 104\ 105\ \dots\ 87] \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \dots \\ 87 \\ 84 \\ 87 \end{bmatrix} - (\frac{1}{4(4)(3)})[105\ 104\ 105\ \dots\ 87] \begin{bmatrix} 1\ 1\ 1 & \cdots & 1\ 1\ 1 \\ \vdots & \ddots & \vdots \\ 1\ 1\ 1 & \cdots & 1\ 1\ 1 \end{bmatrix} \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \dots \\ 87 \\ 84 \\ 87 \end{bmatrix} =$$

= 432 483 - 429 976.02= 2506.98

SS Treat = $(\frac{1}{hs})$ (Y'..k.)(Y..k.) - $(\frac{1}{hts})$ Y'JY

$$= (\frac{1}{4(3)}) \,[1261 \; 1119 \; 1141 \; 1022] \begin{bmatrix} 1261 \\ 1119 \\ 1141 \\ 1022 \end{bmatrix} - (\frac{1}{4(4)(3)})[105 \; 104 \; 105 \; \ldots \; 87] \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ \vdots & & \ddots & & & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \ldots \\ 87 \\ 84 \\ 87 \end{bmatrix} =$$

$$= 432\ 387.25 - 429\ 976.02 = 2411.23$$

Totals were captured in the first matrix, in the order first, second, third and fourth treatment. The same order should be maintained for rows and columns. The common expression, subtracted in the two previous calculations, corresponds to the correction factor. In this section, as suggested in Jasso *et al*. (2022); Pérez *et al*. (2022), the use of a matrix calculator, freely available on its website: (https://matrixcalc.org), will also be very useful.

Database and statistical analysis

A table is made in InfoStat and InfoGen, vertically labeling with H, C, T, S, agh, to identify rows, columns, treatments, subsample and response variable, respectively. For each combination of row, column and treatment, the values of samples 1, 2, 3 are captured cyclically. After verifying the database and making its backup, in the main menu choose: 'estadísticas/análisis de la varianza'. With forward arrow, 'agh' will be sent to dependent variables and 'H, C, T and S' will be transferred to classification variables, choose 'aceptar'. 'H, C, T and H*C*T' will be captured vertically in 'especificación de los términos del modelo', in the first three, after the backslash (\), H*C*T must be written, which is the experimental error. In classification variables, 'H, C, T and S' should be displayed.

Results

With the previous procedure, the analysis of variance will be generated for an LSD with subsampling. In Figures 1, 2, 3, 4 showing the previous procedure.
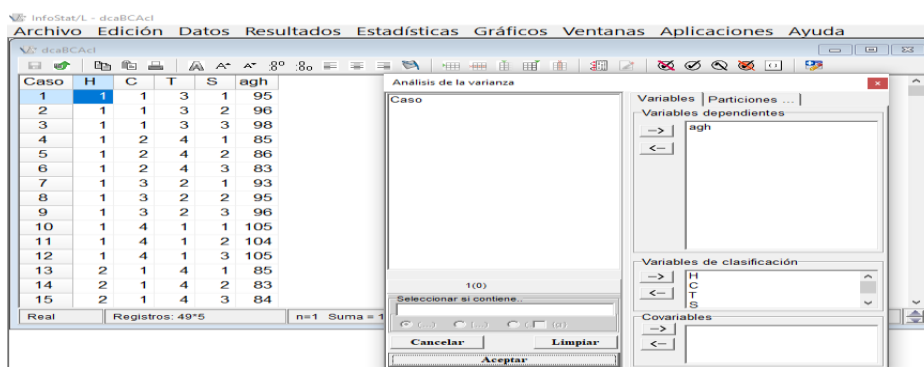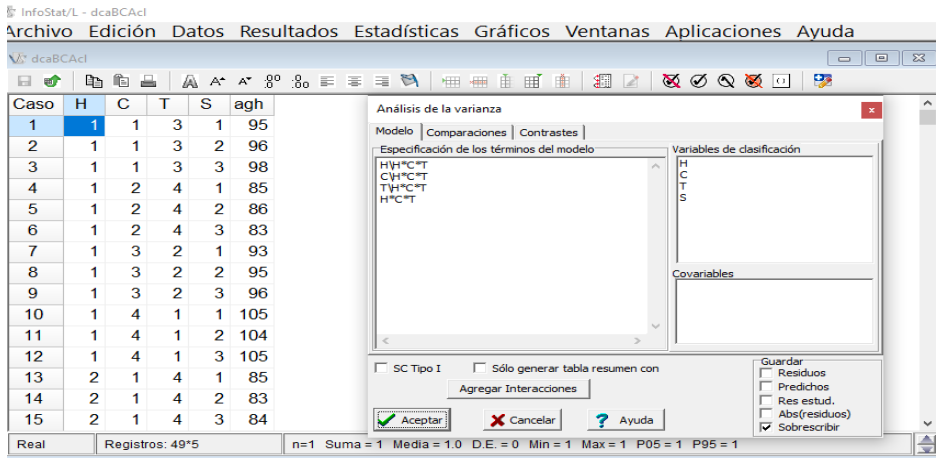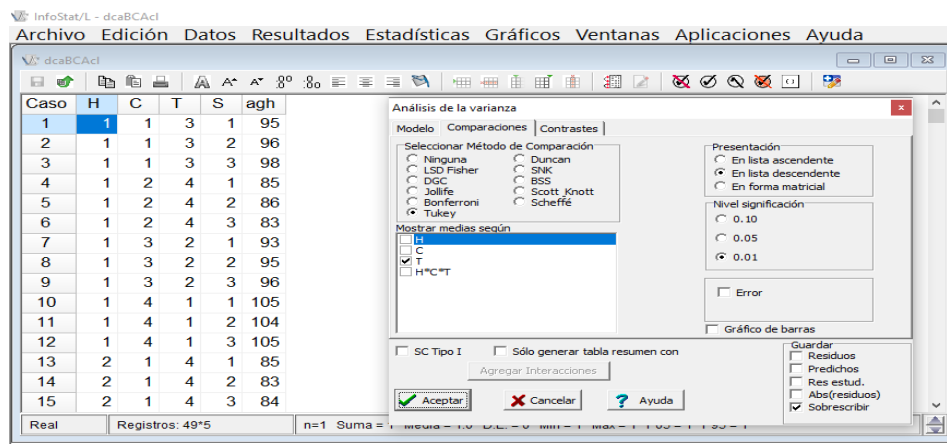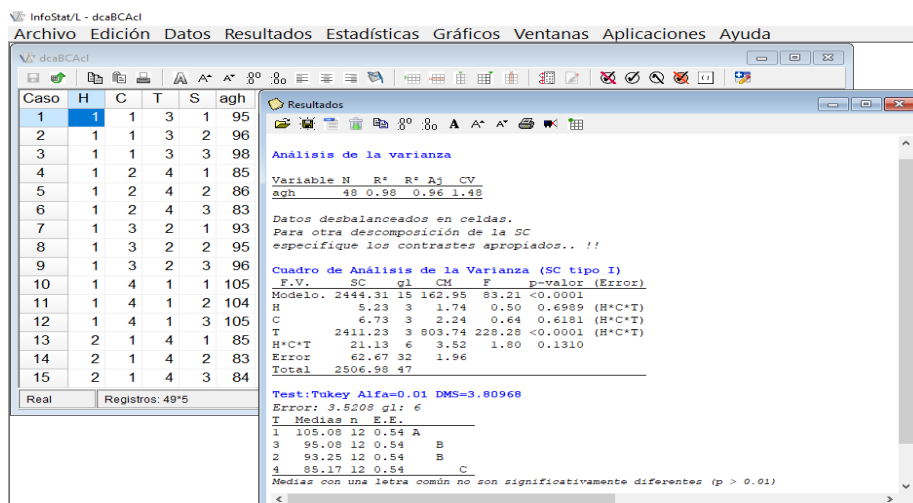

Figure 1.

Figure 2.

Figure 3.

Figure 4.

To generate the Anova for a RCBD with subsampling, 'agh' will be sent to dependent variables and 'H, T, S' must be displayed in classification variables; choose 'aceptar'. In 'especificación de los términos del modelo', 'H, T and H∗T' should be captured vertically, in the first two, after the sign \, H∗T should be written, which is the experimental error. In classification variables, the software will display 'H, T and S'. The procedure is shown in Figures 5, 6, 7, 8.
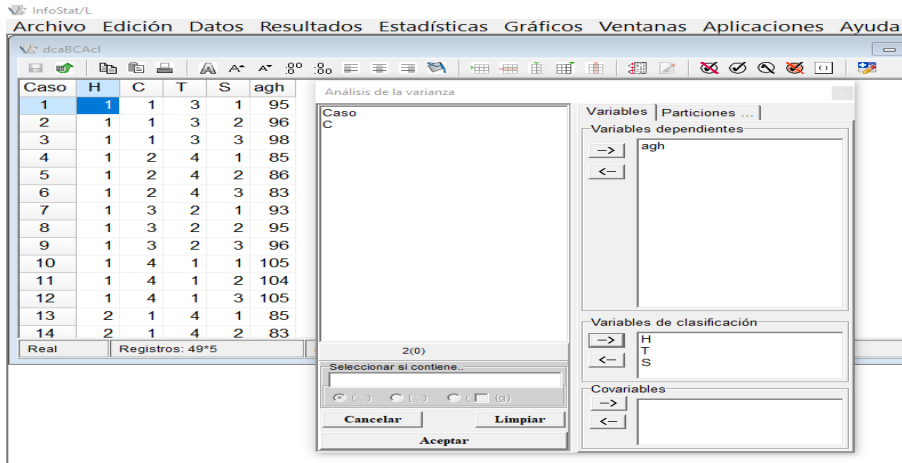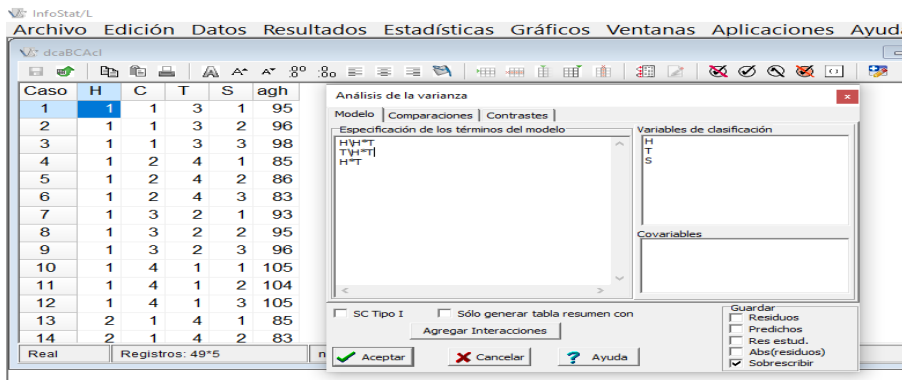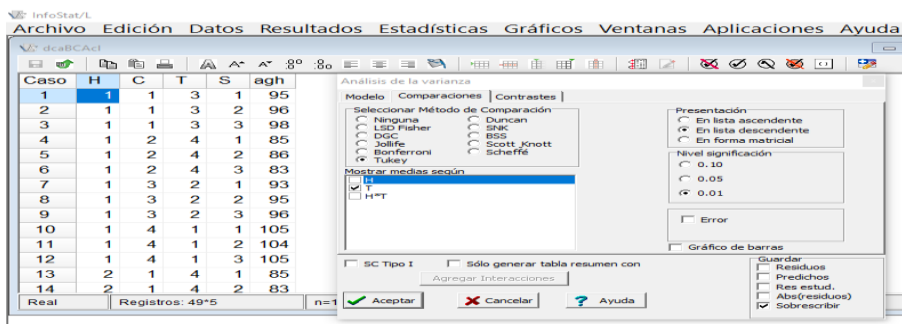

Figure 5.
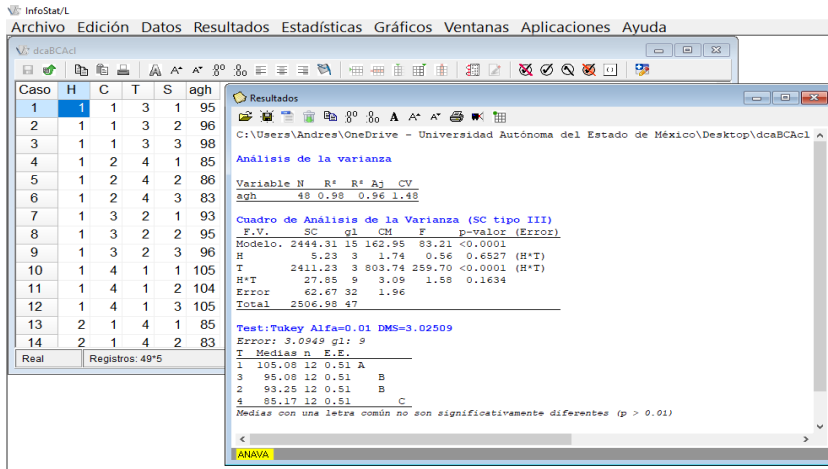

Figure 6.

Figure 7.

Figure 8.

To obtain the Anova for a CRD with subsampling, 'agh' will be sent to dependent variables and 'H, C, T' must be displayed in classification variables. When choosing 'aceptar', the following will be displayed vertically in 'especificación de los términos del modelo': 'T\H∗C∗T' and 'H∗C∗T'. The classification variables will be 'H, C, T'. Choose 'aceptar'. In the three experimental designs, the software will calculate by default the residual of the model or sampling error (Figures 9, 10, 11, 12).
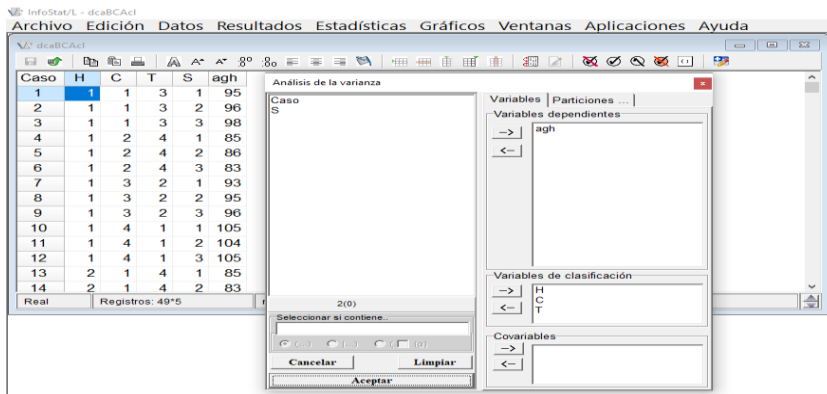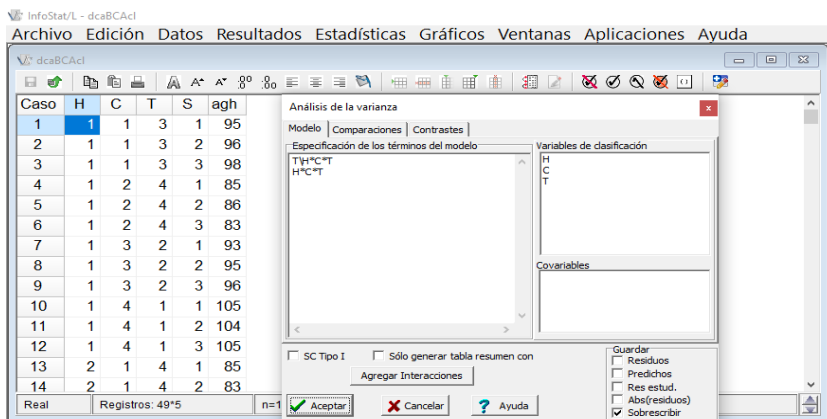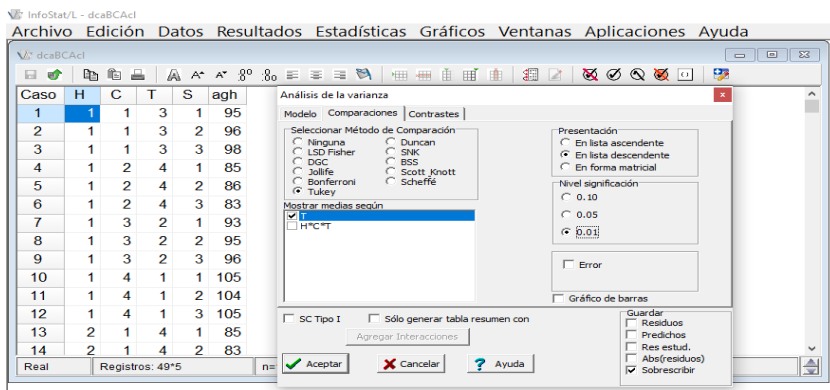

Figure 9.

Figure 10.

Figure 11.
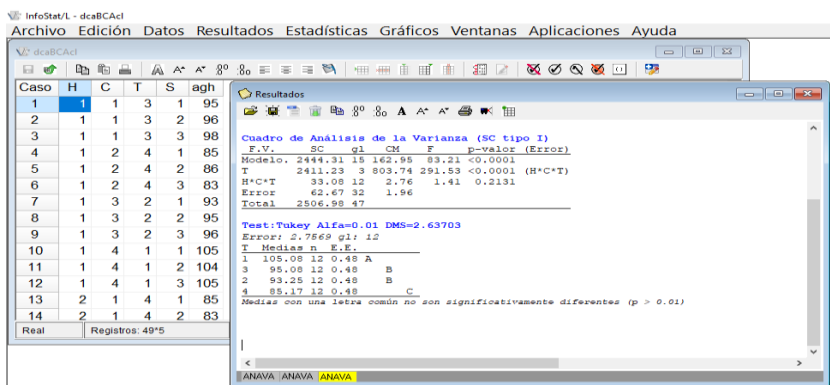
Figure 12.

Validation of the results with SAS

From the editor program of this software, the following is captured: Title 1 'subsampling within experimental units in three designs'; Title 2 'Anova and comparison of treatment means'; data agh22; input H C T S Z; cards;

1 1 3 1 95

1 1 3 2 96

1 1 3 3 98

 ...

4 4 4 3 87

;

Proc GLM; Class H C T; Model Z = H C T H∗C∗T/ss3; Test h= H C T e= H∗C∗T; Means T/Tukey lines alpha=0.01 e= H∗C∗T; Run; Quit;

Proc GLM; Class H T; Model Z = H T H∗T/ss3; Test h=H T e= H∗T; Means T/Tukey lines alpha= 0.01 e= H∗T; Run; Quit;

Proc GLM; Class H C T; Model Z = T H∗C∗T/ss3; Test h= T e= H∗C∗T; Means T/Tukey lines alpha= 0.01 e= H∗C∗T; Run; Quit;

Discussion

The analysis of variance (Anova) and the comparison of means of treatments are two very useful methodologies in the design and analysis of experiments; the first always conditions the use of the second. The Anova allows testing statistical hypotheses in relation to the components of fixed, random or mixed nature that make up the mathematical models that are frequently used in agricultural and forestry sciences, among others; the total variability that is measured in each of the variables of interest is divided into effects and variances for each of its components (Martínez, 1988; Sahagún, 1998; Piepho *et al*., 2003; Restrepo, 2007a, 2007b).

The multiple comparisons of means of treatments, the contrast between a control *versus* the remaining t-1 treatments, the subdivision of the variability contained in the structure of treatments into orthogonal contrasts or polynomials, the application of univariate or multivariate techniques, such as regression or principal component analyses, are only justifiable if, by means of the Anova, the null hypotheses that were raised *a priori* or *a posteriori* are rejected (Sahagún, 1991; Di Rienzo *et al*., 2008; Balzarini *et al*., 2008; Balzarini *et al*., 2016).

Subsampling in experimental designs is also based on the application of an Anova (Gomez and Gomez, 1984; Martínez, 1988; Zamudio and Alvarado, 1996; Hansen *et al*., 2006). To simplify the calculations, the joint error that is considered in the present study, which is the residual in the three statistical models described, is divided into experimental error and sampling error, in the experimental designs CRD, RCBD and LSD, it leads to tests of hypotheses associated with effects and variances that involve both types of error, with and without subsampling.

In the present study, the procedures outlined for InfoStat and InfoGen, as well as those that allow the validation of manual calculations and outputs generated with both software or with SAS, are easy to implement on their platforms and are reliable to individually analyze each of the three experimental designs or to analyze them in a single run, both conventionally and by means of subsampling. Balzarini *et al*. (2008); Di Rienzo *et al*. (2008); Balzarini *et al*. (2016) describe these procedures to obtain an Anova, a comparison of means of treatments, or to apply orthogonal contrasts when using InfoStat and InfoGen, but they are not if subsampling is considered within the experimental unit or plot. It is in this context that SAS is used to validate the results generated by both statistical packages.

Gómez and Gómez (1984) showed the calculations for subsampling in a RCBD, based on the mean squares of the Anova, they estimated the variances of the experimental error and the sampling error and calculated the coefficients of variation for the response variable using both variances. Freund and Wilson (1993) also showed an output generated with SAS for the case of a RCBD.

The codes and procedures that are used in the present study correctly validated the results they showed in relation to the Anova, although in the SAS output presented by Freund and Wilson (1993), the F tests for repetitions, treatments and experimental error were made using the mean square of the sampling error, so, according to Martínez (1988); Zamudio and Alvarado (1996); Sahagún (1998), these would not be correct for the case of the first two sources of variation.

Authors such as Zamudio and Alvarado (1996) masterfully developed matrix theory and applied SAS to analyze three databases to generate, independently, an Anova for CRD, RCBD and LSD. They focused their attention on the partition of the total variation that was recorded in the response variable into two components: that of the experimental unit (UA) and that of the sampling error, for the CRD, in the EU they included treatments and EE, for the RCBD, in addition to the previous two, they also included repetitions, for the LSD, the UA was defined by rows, columns, treatments and EE.

The code for SAS that they implemented in their scientific contribution generates a partition of the effects associated with the UA and the estimation of the sampling error. With the procedures that are considered in the present study, Infostat, InfoGen and SAS could validate the previous results and additionally, the values for the joint error, formed by the experimental error and the sampling error, would be generated.

Studies such as those of Zamudio and Alvarado (1996) mention the use of the quotient that originates between the mean squares of the experimental error on that of the sampling error as a test of relevant hypothesis in the designs CRD, RCBD and LSD, Martínez (1988) also performed it for the CRD, and Freund and Wilson (1993) performed this procedure in a RCBD. On the other hand, Gomez and Gomez (1988) only used both errors to estimate their corresponding variances.

In the present study, the outputs show how statistical significance was obtained in the different sources of variation of the Anova in the three experimental reference designs, as suggested by Zamudio and Alvarado (1996); Martínez (1988); likewise, the classification of the means of treatments using the honest minimum significant difference (HMSD) test, also called Tukey's test, is shown. The Anova indicates the appropriate error terms to test the statistical hypotheses of interest to users when applying InfoStat and InfoGen (see the corresponding images), if in the procedure these are suppressed in the RCBD design, the result generated by SAS presented by Freund and Wilson (1993) will be obtained.

As can be seen in the images shown in the present research, it was considered that the corn varieties that were evaluated based on pollen dehiscence are considered as a fixed effect factor. In this context, the denominator that originates its value of F in treatments, rows or columns is always the mean square of the experimental error, as suggested by Martínez (1988); Zamudio and Alvarado (1996); Sahagún (1998); Restrepo (2007a, 2007b).

Without a doubt, SAS is the most versatile and fastest statistical package that currently exists for the design and analysis of experiments, especially for those whose treatment structure is more complex, such as those analyzed in series of experiments in time and space, in arrangements of divided, subdivided plots or divided blocks, in 2^n, 3^n, or 4^n trials, as well as in different types of lattices (Martínez, 1988; Sánchez, 1995; SAS, 1998; González *et al*., 2019), but its commercial license is more expensive than that of InfoStat and InfoGen. Alternatively, the user could download academic test versions of these three packages free of charge through the internet, but it is easier and faster to do so for InfoStat and InfoGen and these last two could enhance their usefulness using R-Software, which is also freely available through the internet.

Conclusions

The manual calculations presented in this study were correctly validated by the academic test versions of the three statistical packages; InfoStat and InfoGen should be preferred over SAS because both can be downloaded quickly and easily from their websites, but SAS surpasses both by generating in less time, with less effort and in a single procedure, an analysis of variance with subsampling in the completely randomized, randomized complete blocks and Latin square experimental designs, as well as in the application of a comparison of means of treatments with the Tukey test. Additionally, InfoStat and InfoGen commercial licenses are cheaper.

Cited literature

Balzarini, M. G.; González, L.; Tablada, M.; Casanoves, F.; Rienzo, J. A. y Robledo, C. W. 2008. Manual del usuario de infoStat. Ed. Brujas, Córdoba, Argentina. 348 p.

Balzarini, M. G. y Di Rienzo, J. A. 2016. InfoGen. FCA. Universidad Nacional de Córdoba, Argentina. http://www.info-Gen.Com.mx.

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

Freund, R. J. and Wilson, W. J. 1993. Statistical methods. First Ed. Academic Press, Inc. San Diego, CA. USA. 440-452 pp.

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

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

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

González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R., F.; Franco, M.; J. R. P.; Padilla, L. A. 2019. InfoStat, InfoGen y SAS para contrastes mutuamente ortogonales en experimentos en bloques completos al azar en parcelas subdivididas. Rev. Mex. Cienc. Agríc. 10(6):1417-1431.

Hansen, M. J.; Beard, J. T. D. and Hayes, D. B. 2006. Sampling and experimental design. Chapter 3. *In*: analysis and interpretation of freshwater fisheries data. Guy, Ch. S. and. Brown, M. L. Eds. American Fisheries Society. 30-45 pp. Doi: https://doi.org/10.47866/ 9781888569773.

Jasso, B. G.; González, H. A.; Pérez, L. D. J.; Franco, M. J. R. P.; Rubí, A. M. y Mejía, C. J. 2022. Uso de Opstat para validar resultados en un dialélico parcial con ocho líneas de maíz evaluadas en un ambiente. Rev. Mex. Cienc. Agríc. 13(1):41-52.

Martínez, G. A. 1988. Diseños experimentales. Métodos y elementos de teoría. Editorial Trillas, 1$^{ra.}$ Ed. México, DF. 756 p.

Pérez, L. D.; Jasso, B. G.; Saavedra, G. C.; Franco, M. J. R. P.; Ramírez, D. J. F. y González, H. A. 2022. Uso de artificios en Opstat para analizar series de experimentos en dialélico parcial. Rev. Mex. Cienc. Agríc. 13(2):273-287.

Piepho, H. P.; Büchse, A. and Emrich, K. 2003. A Hitchhiker's guide to mixed models for randomized experiments. J. Agron. Crop Sci. 189(3):310-322.

Restrepo, L. F. 2007. Diagramas de estructuras en el análisis de varianza. Rev. Colomb. Cienc. Pec. 20(2):202-208.

Restrepo, B. L. F. 2007 b. La esperanza del cuadrado medio. Rev. Colomb. Cienc. Pec. 20(2):193-201.

Sahagún, C. J. 1991. Utilidad del análisis de varianza en el estudio de la interacción entre genotipos y ambientes. Xilonen. 1(1):21-32.

Sahagún, C. J. 1998. Construcción y análisis de los modelos fijos aleatorios y mixtos. Universidad Autónoma Chapingo (UACH)-Departamento de Fitotecnia-Programa Nacional de Investigación en Olericultura. Boletín técnico núm. 3. 64 p.

Sánchez, G. J. J. 1995. El análisis Biplot en clasificación. Rev. Fitotec. Mex. 18(2):188-203.

Statistical Analysis System (SAS). 1998. SAS/STAT Users Guide. Release 6.03 SAS Institute. Carry North Caroline, USA.

Zamudio, S. F. J. y Alvarado, S. A. A. 1996. Análisis de diseños experimentales con igual número de submuestras. 1ra. Ed. Universidad Autónoma Chapingo (UACH)-División de Ciencias Forestales. México, DF. 85 p.