SISVAR: A COMPUTER ANALYSIS SYSTEM TO FIXED EFFECTS SPLIT PLOT TYPE DESIGNS

Daniel Furtado FERREIRA[1]

■ ABSTRACT: This paper presents a special capability of Sisvar to deal with fixed-effect models with several restrictions in the randomization procedure. These restrictions lead to models with fixed treatment effects, but with several specifications of the error structure. One way do deal with models of this kind is to perform a mixed model analysis, considering only the error effects in the model as random effects and with different covariance structures for the error terms. Another way is to perform an analysis of variance with several errors. These kinds of analysis, for balanced and orthogonal designs, can be done by using Sisvar. The software leads an exact F test for the fixed effects and allows the user to apply multiple comparison procedures or regression analysis for the factor levels of fixed effects, and also in crossed and nested classifications. Sisvar is an interesting statistical computer system for using in balanced and orthogonal agricultural and industrial applications.

■ KEYWORDS: Mixed model; balanced data; agricultural experiments.

1 Introduction

Sisvar was first released in 1996, but its development began in 1994. It was compiled by the Borland Turbo Pascal 3.0 and implemented in the Pascal programming language. The database needed used unique features of the language. A notepad type editor was used by Sisvar to report the results. Sisvar had many capabilities as the Anova, multiple comparison procedures and regression analysis for quantitative factor effects. Moreover, interaction Anova factor effects could be sliced. The first objective of this software was to provide a tool that can be used directly in the design and analysis of experiments course of the Federal University of Lavras (UFLA). It was expected that there would be an improvement in learning to students whose major was not statistics and teaching processes.

[1]Universidade Federal de Lavras – UFLA, Departamento de Estatística, CEP: 37200-900, Lavras, MG, Brazil. E-mail: *danielff@ufla.br* - Bolsista do CNPq

The second objective was to initiate the development of a genuinely Brazilian free software that meets most researcher's demands and peculiarities in the country. There was a fear that non-free programs become increasingly inaccessible to Brazilian researchers, teachers, and students. There were some Brazilian programs for statistical analysis at the time as Sanest, SOC, and Saeg. Some limitations in these programs were the reasons for the Sisvar project, such that it was designed to overcome them. The third goal was to introduce a computer statistical analysis software for the Brazilian scientific community that would enable that research results could be analyzed efficiently and reliably. As Sisvar has being developed by only one researcher, it took approximately two and a half years to have the first version (FERREIRA, 2011, 2014).

Later, in 1997, a Borland Delphi 1.0 Sisvar version for Windows was released. In this version and its updates, several analysis modules were incorporated and the bugs were corrected as they arose. At that time the Sisvar interface was already popular among practical researchers and allowed users less familiar with computers to have a great facility to perform their statistical analysis. In this new versions for Windows, several statistical analysis modules were implemented such as: descriptive statistics, hypothesis testing for various parameters, interval estimation for several population parameters, normality tests, kernel density estimation, simple and multiple linear regression adjustments, methods of model selection as stepwise, backward and forward and more recently, multiple comparisons using bootstrap. The currently Sisvar release is 5.7.

The main analysis module is the analysis of variance (Anova) of linear statistical models. Although very powerful, this option can only be used for balanced and orthogonal data in Sisvar, except for the case of one-way Anova design with one factor. For this specific model, when there are missing data, the user should simply omit the line corresponding to the missing values in the file. Among the advantages that Sisvar has over its competitor, statistical analysis systems is the ability to slice the interaction and nested effects among fixed factors of linear models. Besides the analysis of variance, there is the possibility to apply multiple comparison procedures and contrast of the sliced means of one factor into settled levels of the other including the Scott-Knott test, absent in most of the competitor statistical programs. For quantitative factor effects, the sliced crossover or hierarchical effects could be done by regression analysis. Several other feature of Sisvar could be seen in FERREIRA (2011) or in FERREIRA (2014).

This paper presents a special capability of Sisvar to deal with fixed effects models with several restrictions in the randomization procedure. These restrictions lead models with fixed treatment effects, but with several specifications of error structure. One way do deal with models of this kind is to perform a mixed model analysis, considering only the error effects in the model as random effects and with different covariance structures for the error terms. Another way is to perform an analysis of variance with several error effects. These kinds of analysis, for balanced and orthogonal designs, can be done by using Sisvar. The software leads an exact F test for the fixed effects and allows the user to apply multiple comparison procedures or regression analysis for the factor levels of fixed effect, and also in crossed and nested classifications.

2 Split-plot designs type

There exist many situations where for a factorial experiment different types of experimental units are being used and where the levels of some factors are applied sequentially, necessitating separate randomization procedures (HINKELMANN; KEMPTHORNE, 2008). The simplest case consists of two factors where the experimental units of one size are randomized for the levels of one of the two factors. Those experimental units are then subdivided into smaller experimental units to which the levels of the second factor are randomized. The precisions for the comparison among levels of each factor are different. Also, the fact that each factor is associated with different types of experimental units leads to different experimental error variances associated with these comparisons. This is the main reason that makes this kind of experiment different from the factorial experiments.

The above-described experiment is a simple split-plot design (SPD). There are two kinds of these experiments. The first case considers a block design by superimposing one randomized complete block design (RCBD) on top of another RCBD. The first RCDB involves the whole-plot factor and the second involves the split-plots and the split-plot factor. In the second case, the split-plot design considers a completely randomized design (CRD) to the first factor and a split-plot to the second. This second experiment is called between-and-within-subjects design (HINKELMANN; KEMPTHORNE, 2008). Suppose in each case the factor A is the whole-plot factor with a levels and B is the split-plot factor with b. In the first case, the analysis of variance (Anova) table is presented in Table 1, considering r replicates. This kind of design is referred by SPD(RCBD, RCBD).

Table 1 - Simplified Anova scheme for the split-plot design considering randomized complete block design, SPD(RCBD, RCBD).

Source of Variation	d.f.
Replicates	$r-1$
A	$a-1$
Error (A)	$(r-1)(a-1)$
B	$b-1$
A $\times$ B	$(a-1)(b-1)$
Error (B)	$(r-1)a(b-1)$
Total	$rab-1$

In the second case, the analysis of variance (Anova) table is presented in Table 2, also considering r replicates in a completely randomized design for the whole-plots factor. This kind of design is referred by SPD(CRD, RCBD).

Table 2 - Simplified Anova scheme for the split-plot design considering complete randomized design, SPD(CRD, RCBD).

Source of Variation	d.f.
A	$a - 1$
Error (A)	$a(r - 1)$
B	$b - 1$
A $\times$ B	$(a - 1)(b - 1)$
Error (B)	$a(r - 1)(b - 1)$
Total	$rab - 1$

When for a A-factor the observations are made along specified times (hours, days, months, etc.) the design for such an experiment is of the form of split-plot designs (SPD), such as SPD(RCBD, RCBD) or SPD(CRD, RCBD). They are also referred erroneously to a split-plot design in time (STEEL; TORRIE, 1996; GOMES, 2009). There are many issues with this viewpoint. First, the B-factor levels (time) are not randomized. Secondly, there is a covariance structure for the observations and hence for the errors besides that ordinarily induced by the randomization procedure. This may invalidate the Anova presented above. In these cases, only if the covariance structure satisfies the Huynh-Feldt conditions (sphericity) (HUYNH; FELDT, 1970) the F tests for the factors B and $A \times B$ will be valid. There are several other types of SPD, as can be seen in Hinkelmann and Kempthorne (2008). This kind of design is also obtained when the time observations were replaced by space observations, as, for instance, different depths of soil in a compaction study.

Sisvar can deal with those experiments quickly. An easy approach for the SPD(RCBD, RCBD), when the B-factor is the time or the depth of soil can be taken by considering an Anova model with 3 error terms as shown in Table 3. These types of analyses have an exact F test when Huynh-Feldt conditions (sphericity or compound symmetry) (HUYNH; FELDT, 1970) are satisfied. Only if the covariance structure satisfies the Huynh-Feldt conditions (HUYNH; FELDT, 1970) of MS(T)/MS(Error(B)) and MS(A $\times$ T) /MS (Error (B)) have F-distributions, where MS stands by mean square. However, it is unlikely that the compound symmetry (CS) is appropriate for repeated measures data, but if this structure holds then the analysis is equivalent to the Anova given in Table 3. This is the reason why the CS structure is appealing and frequently used. We caution the user to be very careful with its use when the CS structure is not appropriate. Therefore, the CS is not the cause of the three error model, but it reflects in the F-test quality.

The additional error in this model has the justification for its inclusion in the model due to the lack of randomization of the time factor along the blocks or levels of the factor under test (treatments) (A-factor). Compound symmetry (CS) or Huynh-Feldt conditions are necessary assumptions, regardless of this additional error, for the validity of the F tests in Anova. One way do deal with models of this kind is to perform a mixed model analysis, considering only the error effects in the model as random effects and with different covariance structures for the error terms. The mixed model analysis can deal with several other error covariance structures, but can not be done by Sisvar.

Experimenters and statisticians sometimes mistake a split block design with a split-plot design. This is especially true when one of the factors involves periods, say factor B, or other forms of repeated measurements through time, which they designate as split-plot treatments. Whenever the times are calendar or clock times, the periods are crossed over the other set of treatments, say A. This means that they are in a split block arrangement. Such experiments could be considered to be in a repeated measurement category but in many instances, each of the time periods and especially the interactions of factor A with time periods B are of interest to an experimenter, for example, the various pickings of a tomato or bean crop or the various cuttings of a hay crop.

Table 3 - Simplified Anova scheme for the split-plot design considering repeated measure (T-factor with t levels) in the SPD(RCBD, RCBD).

Source of Variation	d.f.
Replicates	$r-1$
A	$a-1$
Error (A)	$(r-1)(a-1)$
T	$t-1$
Error (B)	$(t-1)(r-1)$
A $\times$ T	$(a-1)(t-1)$
Error (C)	$(r-1)(a-1)(t-1)$
Total	$rat-1$

3 Multiple comparisons procedures and regression in Anova

For the split-plot type designs, when interaction effects are significantly $(P \leq 0.05)$ different of 0, the researcher has an interest in performing an analysis to achieve the best level of one factor within specific levels of the other factors. However, as there are several errors in the Anova model, comparisons between means involve combined variances of the various error mean squares. For example, in the Anova showed in Table 3, the variance of the difference between two means of levels of A-factor to a fixed level of the time is

$$\text{Variance}(\bar{Y}_{ik.} - \bar{Y}_{i'k.}) = \frac{2}{r}\left[\frac{MS(Error(A)) + (t-1)MS(Error(C))}{t}\right].$$

In the same way, the variance of the difference between two means of levels of T-factor to a fixed level of the A-factor is

$$\text{Variance}(\bar{Y}_{ik.} - \bar{Y}_{ik'.}) = \frac{2}{r}\left[\frac{MS(Error(B)) + (a-1)MS(Error(C))}{a}\right].$$

This is what Sisvar called “Complex Variance” using the usual notation of specialized literature. Someone can deal with this situation in Sisvar performing all F tests and all multiple comparison procedures immediately and easily. Several multiple comparison procedures are available in Sisvar. Among them there is the Scott-Knott test to which the results are some grouping of the treatment means where there are no ambiguous results as in the traditional multiple comparison procedures as Tukey and SNK. Also, if the treatment levels are continuous variables (fixed effects), the appropriate analysis is to fit a linear regression model. The Sisvar can fit several linear models, where the effects to be included in the model are chosen by the users.

4 Conclusions

The main advantage to use Sisvar is its capability to deal with split-plot type of design, including split-block and split-plot with repeated measures employing the simple Anova method. Also, it allows multiple comparisons analysis and regression on the interaction effects present in the model easily.

Acknowledgments

I would like to thank reviewer and editors for their comments and suggestions.

FERREIRA, D. F. Sisvar: Um sistema de análise estatística para delineamentos de efeitos fixos do tipo parcelas subdivididas. *Rev. Bras. Biom.,* Lavras, v.37, n.4, p.529-535, 2019.

- *RESUMO: Este trabalho apresenta uma capacidade especial do Sisvar em lidar com modelos fixos com severas restrições no procedimento de aleatorização. Estas restrições conduzem a modelos com efeitos fixos de tratamentos, mas com várias especificações da estrutura de erros. Uma maneira de lidar com modelos deste tipo é aplicar uma análise de modelos mistos, considerando apenas os efeitos dos erros no modelo com aleatórios e com diferentes estruturas de covariâncias para o efeito dos erros do modelo. Outra forma é realizar uma análise de variância com vários erros no modelo. Este tipo de análise, quando os dados são balanceados e ortogonais, pode ser feita usando-se o Sisvar. O programa produz um teste F exato para os efeitos fixos e permite ao usuário aplicar procedimentos de comparações múltiplas ou análise de regressão para os níveis dos efeitos fixos do modelo, em estruturas de classificação cruzadas ou aninhadas. O Sisvar é um sistema computacional de análises estatísticas interessante para ser usado em conjuntos de dados balanceados e ortogonais das aplicações agrícolas ou industriais.*

- *PALAVRAS-CHAVE: Modelos mistos; dados balanceados; experimentos da agricultura.*

References

FERREIRA, D. F. Sisvar: a computer statistical analysis system. *Ciência e Agrotecnologia*, v. 35, n. 6, p. 1582–1589, 2011.

FERREIRA, D. F. Sisvar: a guide for its bootstrap procedures in multiple comparisons. *Ciência e Agrotecnologia*, v. 38, n. 2, p. 1582–1589, 2014.

GOMES, F. P. *Curso de estatística experimental*. 15.ed. Piracicaba: Fealq, 2009. 451 p.

HINKELMANN, K.; KEMPTHORNE, O. *Design and analysis of experiments - Introduction to experimental design*. 2.ed. New York: John Wiley, 2008. v. 1. 631 p.

HUYNH, H.; FELDT, L. S. Condition under which mean square rations in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association*, v. 65, n. 332, p. 1582–1589, 1970.

STEEL, R. G. D.; TORRIE, J. H. *Principles and Procedures of Statistics: a biometrical approach*. 3.ed. New York: McGraw-Hill Book Company, 1996. 688 p.

Received on 12.08.2019.

Approved after revised on 04.10.2019.