Confidence intervals for the variance components in a grapevine experiment

Dário Ferreira[1], João Tiago Mexia[2]

[1]Departamento de Matemática, Universidade da Beira Interior - Covilhă 6200
email:dario@ubi.pt
[2]Departamento de Matemática, F.C.T., Universidade Nova de Lisboa - Quinta da Torre 2825-114

SUMMARY

We study a model with three random effects factors, where the first crosses the second which nests the third, and apply the results to an example in grapevine castes carried out to see if there is genetical homogeneity break-down due to isolation.
Key words: Variance components, random effects factors, cross-nesting, simulation.

1. Introduction

This experiment was set up to see if, in grapevines castes, there is genetical homogeneity break-down due to isolation. This break-down is interesting since grapevines are reproduced vegetatively. Clones will consist of grapevines with a known ancestor and are grouped into castes with an assumed common ancestor. Now a well known Portuguese caste, Touriga Nacional, is grown mainly in two distinct regions. This enabled the use of such a caste in the experiment. Three clones from each of the two regions were randomly selected. The six clones were planted in a homogeneous field, using a grid pattern with one cultivar per column. Three groups of five adjacent rows were randomly selected. The factors to be considered in the experiment were, besides origin, clone and location. The levels of this last factor were the groups of rows. Now the location factor crosses with the origin factor, which nests the clones factor.

Previous analysis of this experiment were carried by *Fonseca et al.* (2003) and *Ferreira et al.* (2004). In the first of these papers all three factors were treated as having random effects. *UMVUE* estimators for the variance components were obtained and the hypothesis of their nullity was tested. While it is clearly acceptable that the location and clone factors have random effects, for the origin factor such an assumption is less forthcoming. Thus it is of interest to point out that (see *Ferreira et al* (2004)) with the origin factor taken as having fixed effects the corresponding test has the same p-value as when, as previously, it is assumed to have random effects.

We now assume the origin factor to have fixed effects and obtain confidence intervals for the variance components, through Monte-Carlo methods based on the orthogonal structure of the model. In the next section we present such a structure, before treating the data.

As is well known (see for instance *Khuri et al* (1997)) there are no unbiased estimators for all variance components given by the difference of two *ANOVA* mean squares. Actually, as we will show for this experiment, variance components may be written as the difference between a positive and a negative part. We will show how to obtain *UMVUE* for both parts, $(\tilde{\sigma}^2)^+$ and $(\tilde{\sigma}^2)^-$, while deriving confidence intervals for them and the corresponding variance components. These *UMVUE* will be useful in validating the use of Monte Carlo methods. To do this we will show that the point $((\tilde{\sigma}^2)^-;(\tilde{\sigma}^2)^+)$ lies along the lines that correspond to the median of the empirical samples for the product, the quotient, sum and difference of both parts.

2. Model

Assigning indexes i, j, k and p to location, origin, clone and replicates, we have, using the notation of *Khuri et al.* (1998), the model

$$y_{i,j,k,p} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_{jk} + (\tau\delta)_{j(ik)} + e_{ijkp}, \tag{1}$$

with $i = 1,\dots,v$ $(=3)$, $j = 1,\dots,b$ $(=2)$, $k = 1,\dots,n$ $(=3)$ and $p = 1,\dots,r$ $(=5)$.

Writing $\theta\chi_n^2$ for the product by θ of a central chi-square with n degrees of freedom, we have $SS(\tau) \sim \gamma_1\chi_{g_1}^2$, $SS(\beta) \sim \gamma_2\chi_{g_2}^2$, $SS(\delta) \sim \gamma_3\chi_{g_3}^2$, $SS(\tau\beta) \sim \gamma_4\chi_{g_4}^2$ and $SS(\tau\delta) \sim \gamma_5\chi_{g_5}^2$, with

$$\begin{cases} \gamma_1 = \sigma^2 + bnr\sigma_\tau^2 + nr\sigma_{\tau\beta}^2 + r\sigma_{\tau\delta}^2 \\ \gamma_2 = \sigma^2 + vnr\sigma_\beta^2 + nr\sigma_{\tau\beta}^2 + vr\sigma_\delta^2 + r\sigma_{\tau\delta}^2 \\ \gamma_3 = \sigma^2 + vr\sigma_\delta^2 + r\sigma_{\tau\delta}^2 \\ \gamma_4 = \sigma^2 + nr\sigma_{\tau\beta}^2 + r\sigma_{\tau\delta}^2 \\ \gamma_5 = \sigma^2 + r\sigma_{\tau\delta}^2 \end{cases}, \tag{2}$$

$$\begin{cases} g_1 = v-1 \\ g_2 = b-1 \\ g_3 = b(n-1) \\ g_4 = (v-1)(b-1)g_4 = (v-1)b(n-1) \end{cases} \tag{3}$$

and $SS_e \sim \sigma^2 \chi^2_{vbn(r-1)}$. Thus

$$\begin{cases} \sigma^2_{\tau\delta} = \frac{1}{r}\left(\gamma_5 - \sigma^2\right) \\ \sigma^2_{\tau\beta} = \frac{1}{nr}\left(\gamma_4 - \gamma_5\right) \\ \sigma^2_{\delta} = \frac{1}{vr}\left(\gamma_3 - \gamma_5\right) \\ \sigma^2_{\beta} = \frac{1}{vnr}\left(\gamma_2 - \gamma_3 - \gamma_4 + \gamma_5\right) \\ \sigma^2_{\tau} = \frac{1}{nbr}\left(\gamma_1 - \gamma_4\right) \end{cases} . \tag{4}$$

The $SS(\tau)$, $SS(\beta)$, $SS(\delta)$, $SS(\tau\beta)$ and $SS(\tau\delta)$ are sufficient statistics from which, see *Fonseca et al.* (2003), *UMVUE* might be derived for the γ, the variance components and their positive and negative parts. These may be obtained directly from expression 4, for instance $\sigma^{2+}_{\beta} = \frac{1}{vnr}\left(\gamma_2 + \gamma_5\right)$

Now if $S \sim \theta\chi^2_n$, when $S = s$, we have the $1-q$ level confidence interval for θ given by

$$\left[\frac{s}{x_{n,1-\frac{q}{2}}};\frac{s}{x_{n,\frac{q}{2}}}\right]. \tag{5}$$

Likewise we can obtain a probability for every closed interval covering θ. Since the family constituted by closed intervals and the void set is closed for intersection, there is (see *William* (1997, pg. 19)) a unique probability measure that takes the same values, for closed intervals, and is associated to θ. The corresponding distribution functions will also be associated to θ. We say that this probability measure is induced by s.

Thus we can use the values s_τ, s_β, $s_{\tau\beta}$, s_δ, $s_{\tau\delta}$ and s_e of the sums of squares to induce probability measures associated to γ_1, γ_2, γ_3, γ_4, γ_5 and σ^2. To obtain independent samples with these probability measures we take sets ($x_{i,1}$, $x_{i,2}$, $x_{i,3}$, $x_{i,4}$, $x_{i,5}$, $x_{i,6}$) of independent central chi-squares with g_1, g_2, g_3, g_4, g_5 and $vbn(r-1)$ degrees of freedom and obtain the : $z_{i,1}=\frac{s_\tau}{x_{i,1}}$, $z_{i,2}=\frac{s_\beta}{x_{i,2}}$, $z_{i,3}=\frac{s_{\tau\beta}}{x_{i,3}}$, $z_{i,4}=\frac{s_\delta}{x_{i,4}}$, $z_{i,5}=\frac{s_{\tau\delta}}{x_{i,5}}$ and $z_{i,6}=\frac{s_e}{x_{i,6}}$, $i=1,\ldots,N$.

Since the variance components are given by measurable functions of γ_1, γ_2, γ_3, γ_4, γ_5 and σ^2 the samples $\{w_{1,j},\ldots,w_{N,j}\}$, $j=1,2,3,4,5$, with

$$\begin{cases} w_{i,1}=\frac{1}{nbr}\left(z_{i,1}-z_{i,4}\right);\ i=1,\ldots,N \\ w_{i,2}=\frac{1}{vnr}\left(z_{i,2}-z_{i,3}-z_{i,4}+z_{i,5}\right);\ i=1,\ldots,N \\ w_{i,3}=\frac{1}{vr}\left(z_{i,3}-z_{i,5}\right);\ i=1,\ldots,N \\ w_{i,4}=\frac{1}{nr}\left(z_{i,4}-z_{i,5}\right);\ i=1,\ldots,N \\ w_{i,5}=\frac{1}{r}\left(z_{i,5}-z_{i,6}\right);\ i=1,\ldots,N \end{cases} \tag{6}$$

will have probability distributions associated with the variance components. Thus we can use the empirical quantiles of these samples to obtain confidence intervals for the variance components.

3. Data treatment

In table 1 we present the yield, in Kg, for the grapevines in the experiment.

From this data we obtained

$$\begin{cases} SS(\tau)=0.6748 \\ SS(\beta)=9.6105 \\ SS(\delta)=6.2590 \\ SS(\tau\beta)=3.4464 \\ SS(\tau\delta)=3.4158 \\ SS_e=26.4333 \end{cases} \tag{7}$$

Table 1. Yields in Kg

Location	Origin 1			Origin 2		
	Clone 1	Clone 2	Clone 2	Clone 1	Clone 2	Clone 3
1	3.00	1.00	1.10	1.75	1.10	1.05
	1.85	1.10	1.50	3.50	1.05	1.25
	0.75	1.00	1.80	2.50	0.50	2.00
	1.35	1.60	1.45	2.00	1.05	1.50
	1.45	1.50	1.25	0.65	1.25	2.10
2	1.80	1.60	0.85	2.00	1.20	1.00
	0.70	1.75	0.65	3.00	1.35	2.70
	2.50	0.50	0.55	2.55	1.20	2.15
	1.70	1.35	0.90	3.00	0.30	2.10
	0.40	1.10	0.90	2.65	2.50	2.70
3	1.05	0.75	0.90	1.60	1.05	1.60
	1.50	0.65	0.90	3.05	1.95	1.10
	1.15	0.90	0.55	0.25	2.00	2.05
	0.85	0.85	0.70	1.66	2.20	1.50
	1.15	1.05	0.35	2.65	2.35	3.00

and the *UMVUE* in (8)

$$\begin{cases} (\tilde{\sigma}^2_\tau)^+ = \frac{1}{30}\tilde{\gamma}_1;\ (\tilde{\sigma}^2_\tau)^- = \frac{1}{30}\tilde{\gamma}_4 \\ (\tilde{\sigma}^2_\beta)^+ = \frac{1}{45}(\tilde{\gamma}_2+\tilde{\gamma}_5);\ (\tilde{\sigma}^2_\beta)^- = \frac{1}{45}(\tilde{\gamma}_3+\tilde{\gamma}_4) \\ (\tilde{\sigma}^2_\delta)^+ = \frac{1}{15}\tilde{\gamma}_3;\ (\tilde{\sigma}^2_\delta)^- = \frac{1}{15}\tilde{\gamma}_5 \\ (\tilde{\sigma}^2_{\tau\beta})^+ = \frac{1}{15}\tilde{\gamma}_4;\ (\tilde{\sigma}^2_{\tau\beta})^- = \frac{1}{15}\tilde{\gamma}_5 \\ (\tilde{\sigma}^2_{\tau\delta})^+ = \frac{1}{5}\tilde{\gamma}_5;\ (\tilde{\sigma}^2_{\tau\delta})^- = \frac{1}{5}\tilde{\sigma}^2 \end{cases} . \quad (8)$$

We now generate the following

$$\begin{cases} w_{i,1}{}^+ = \frac{1}{30}z_{i,1};\ w_{i,1}{}^- = \frac{1}{30}z_{i,4};\ i=1,\ldots,10000 \\ w_{i,2}{}^+ = \frac{1}{45}\left(z_{i,2}+z_{i,5}\right);\ w_{i,2}{}^- = \frac{1}{45}\left(z_{i,3}+z_{i,4}\right);\ i=1,\ldots,10000 \\ w_{i,3}{}^+ = \frac{1}{15}z_{i,3};\ w_{i,3}{}^- = \frac{1}{15}z_{i,5};\ i=1,\ldots,10000 \\ w_{i,4}{}^+ = \frac{1}{15}z_{i,4};\ w_{i,4}{}^- = \frac{1}{15}z_{i,5};\ i=1,\ldots,10000 \\ w_{i,5}{}^+ = \frac{1}{5}z_{i,5}-z_{i,6};\ w_{i,5}{}^- = \frac{1}{5}z_{i,6};\ i=1,\ldots,10000 \end{cases} \quad (9)$$

with $\chi^2_{0.99} < w_{i,j} < \chi^2_{0.01}$, $j = 1,...,4$, thus obtaining examples with the densities induced for the variance components.

These densities are presented in graphs 1 to 5 as well as the corresponding *UMVUE*, which is represented by a red dot. We see that the densities appear to be unimodal with the *UMVUE* near the modes.

These graphs were built from histograms, the breadths of the classes being 0.006599, 0.018545, 0.025544, 0.016655 and 0.018950 for the first, second and third factors and for the interactions 1×2 and 1×3 respectively.

In table 2 we indicate, for the various sets of factors associated to variance components, the corresponding *UMVUE* and central point of the modal class.

Table 2.

Factor	UMVUE	Center of the modal class	p_i	D.F.
1	-0.0462	-0.0171	0.5407	2/2
2	0.1500	0.0170	0.5487	1/8/4/2
3	0.0759	0.0336	0.4368	4/8
1 x 2	0.0864	0.0350	0.4284	2/8
1 x 3	0.0120	0.0065	0.4511	8/72

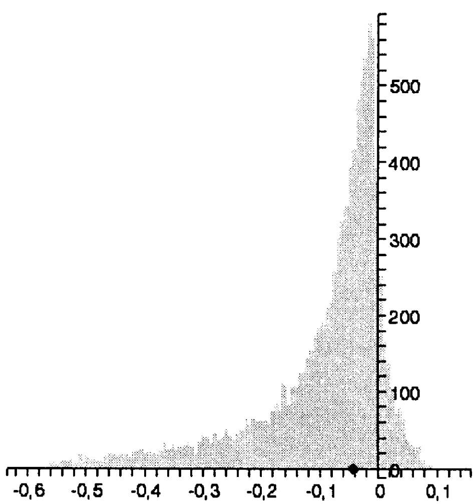

Figure 1. Localization of the *UMVUE*, relative to the empirical density, for the first factor. $D.F. = 2/2$

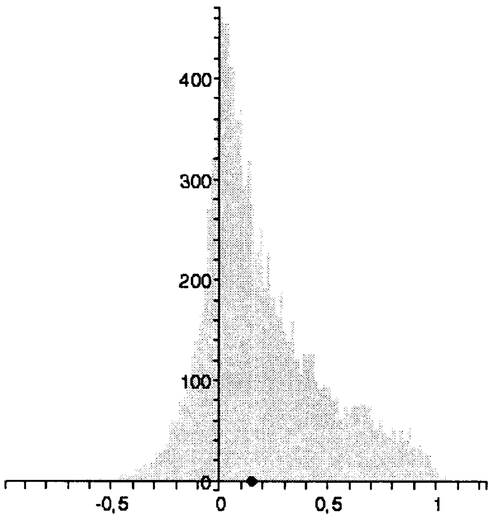

Figure 2. Localization of the *UMVUE*, relative to the empirical density, for the second factor. $D.F. = 1/8/2/4$

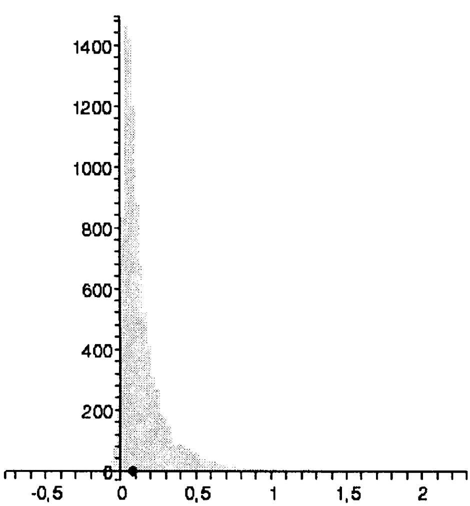

Figure 3. Localization of the *UMVUE*, relative to the empirical density, for the third factor. $D.F. = 4/8$

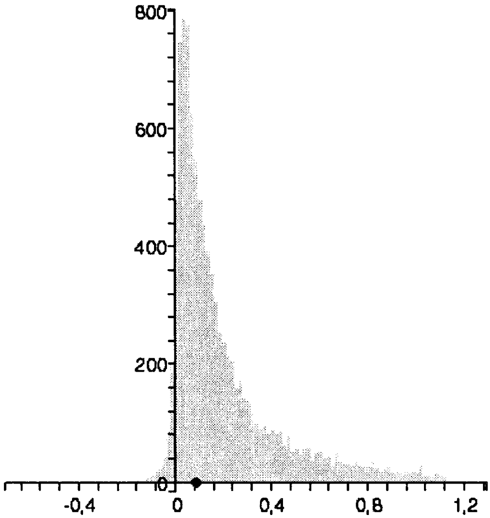

Figure 4. Localization of the *UMVUE*, relative to the empirical density, for the interaction the first and the second factor. *D.F.* = 2/8

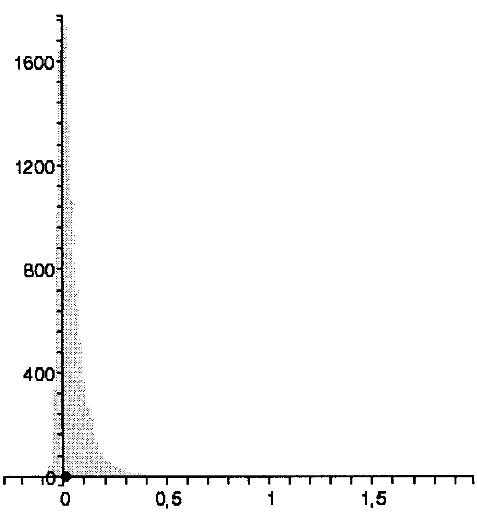

Figure 5. Localization of the *UMVUE*, relative to the empirical density, for the interaction the first and the third factor. *D.F.* = 8/72

As may be seen in expression 4, the variance components have a positive and a negative part. For instance, $(\sigma_\beta^2)^+ = \frac{1}{vnr}(\gamma_2 + \gamma_5)$ and $(\sigma_\beta^2)^- = \frac{1}{vnr}(\gamma_3 + \gamma_4)$. In graphs 6 to 10 we present empirical joint densities induced for both parts of the variance components.

The coordinates in the xoy plane of the point located above each joint density are the $UMVUE$ estimators of both parts. It is clearly seen that the joint densities are unimodal and that the mode and the point estimators are close neighbors for all variance components.

These empirical densities were obtained from pairs (X_i^-, X_i^+), $i = 1,\ldots,$ of generated values for the positive and negative parts of the variance components. This enabled us to pursue the study of the location of the $UMVUE$ in relation to the bidimensional empirical densities. Let u_q $[v_q]$ be the empirical quantile for probability q of the quotient $\frac{X_i^+}{X_i^-}$ [product $X_i^+ X_i^-$]. In graphs 12 to 16 the curves $y = u_q x$ and $y = \frac{v_q}{x}$ are presented for the various variance components and for the values $q = 0.05$, 0.1, 0.3, 0.5, 0.7, 0.9 and 0.95. The coordinates of the red dot are the $UMVUE$ estimators for the positive and negative parts.

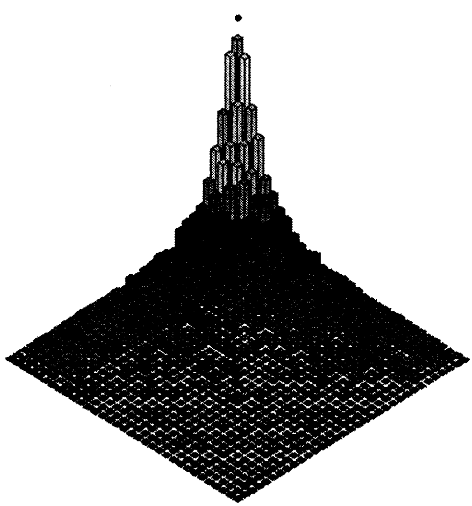

Figure 6. Localization of the $UMVUE$, relative to the bidimensional density, for the first factor. $D.F. = 2/2$

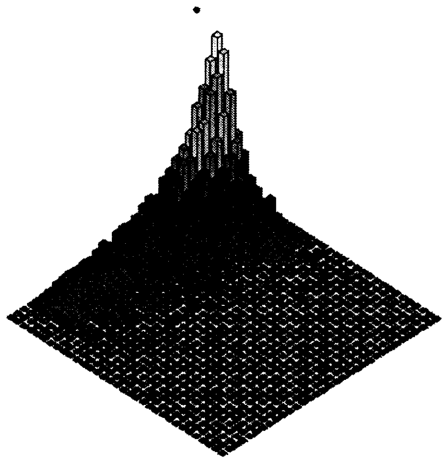

Figure 7. Localization of the *UMVUE*, relative to the bidimensional density, for the second factor. $D.F. = 1/8/2/4$

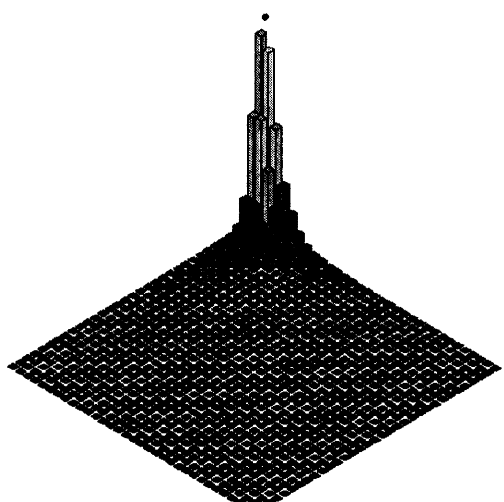

Figure 8. Localization of the *UMVUE*, relative to the bidimensional density, for the third factor. $D.F. = 4/8$

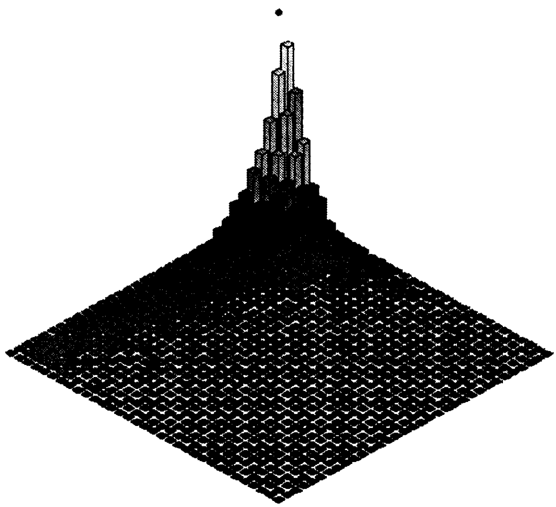

Figure 9. Localization of the $UMVUE$, relative to the bidimensional density, for interaction between the first and the second factor. $D.F. = 2/8$

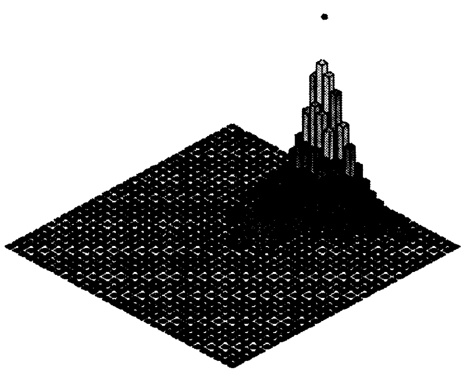

Figure 10. Localization of the $UMVUE$, relative to the bidimensional density, for interaction between the first and the third factor. $D.F. = 8/72$

A legend for the diagrams is given on Figure 11.

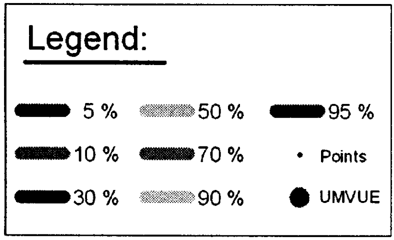

Figure 11. Legend

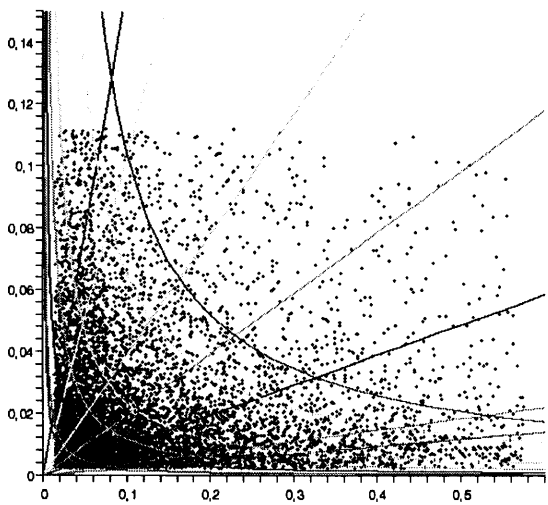

Figure 12. First factor (j=1), $D.F. = 2/2$

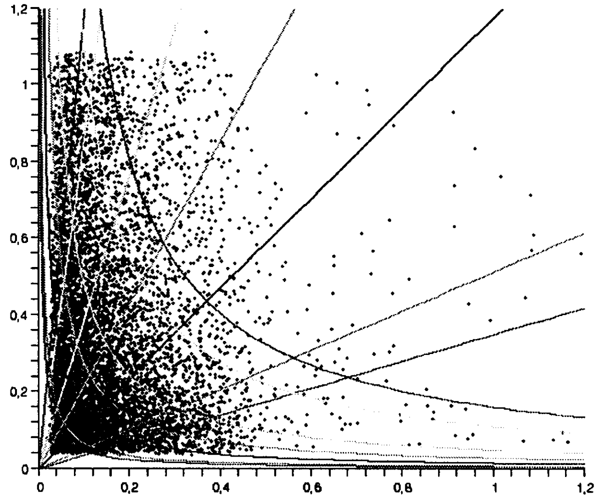

Figure 13. Second factor (j=2), $D.F. = 1/8/2/4$

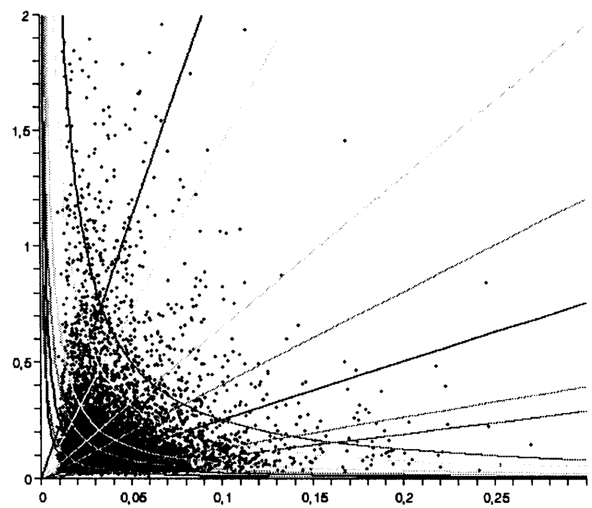

Figure 14. Third factor (j=3), $D.F. = 4/8$

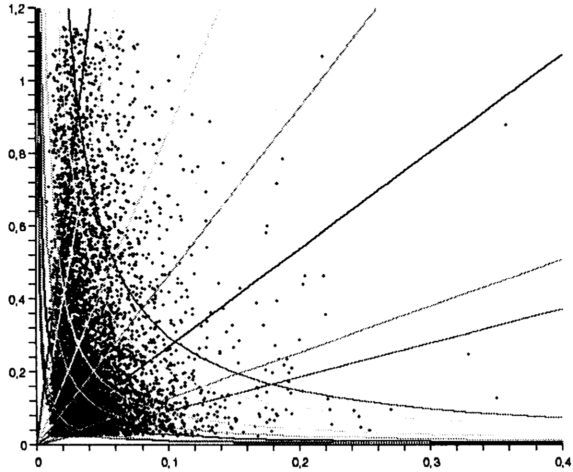

Figure 15. Interaction of the first and the second factor (j=4), $D.F. = 2/8$

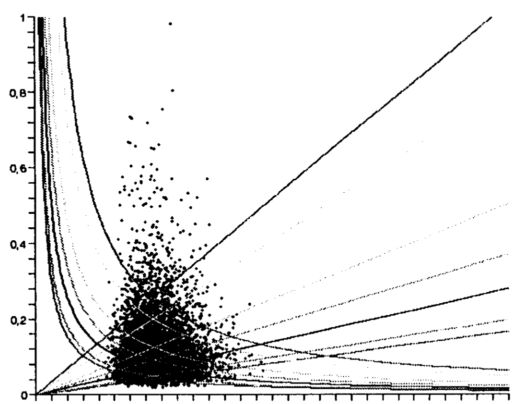

Figure 16. Interaction of the first and the third factor (j=5), $D.F. = 8/72$

In Table 3 we present, for the positive and negative parts of the variance components, the *UMVUE* and the empirical quantiles $v_{q,i,j}$ and $u_{q,i,j}$ for the product and quotient.

Table 3.

Factor	$\sigma^2_{c,i,j}$	Quantlle	5%	10%	30%	50%	70%	90%	95%
1	(0.05744;0.01125)	$v_{q,1,0}$	0.00018	0.00024	0.00060	0.00117	0.00241	0.00637	0.01041
		$u_{q,1,0}$	0.0221	0.0364	0.0967	0.1966	0.3978	1.0757	1.7053
2	(0.07307;0.22308)	$v_{q,0,1}$	0.00488	0.00686	0.01488	0.02693	0.04949	0.10790	0.15365
		$u_{q,0,1}$	0.3564	0.5071	1.1637	2.1107	3.8501	8.4411	11.6484
3	(0.02847;0.10432)	$v_{q,0,2}$	0.00110	0.00143	0.00261	0.00417	0.00695	0.01524	0.02330
		$u_{q,0,2}$	1.0247	1.3591	2.5612	4.0168	6.4727	14.0548	21.6227
1x2	(0.02847;0.11488)	$v_{q,1,1}$	0.00099	0.00135	0.00278	0.00488	0.00890	0.02115	0.03030
		$u_{q,1,1}$	0.9012	1.2641	2.6619	4.6808	8.5597	19.6296	28.7576
1x3	(0.07343;0.08540)	$v_{q,1,2}$	0.00313	0.00368	0.00527	0.00696	0.00939	0.01502	0.01941
		$u_{q,1,2}$	0.5520	0.6648	0.9477	1.2665	1.7171	2.6976	3.4988

Now let w_q $[z_q]$ be the probability q empirical quantiles corresponding to the sum [difference] of the positive and negative parts. In graphs 17 to 21 the curves $y = w_q - x$ and $y = u_q + x$ are presented for the various variance components and for the values $q = 0.05$, 0.1, 0.3, 0.5, 0.7, 0.9 and 0.95.

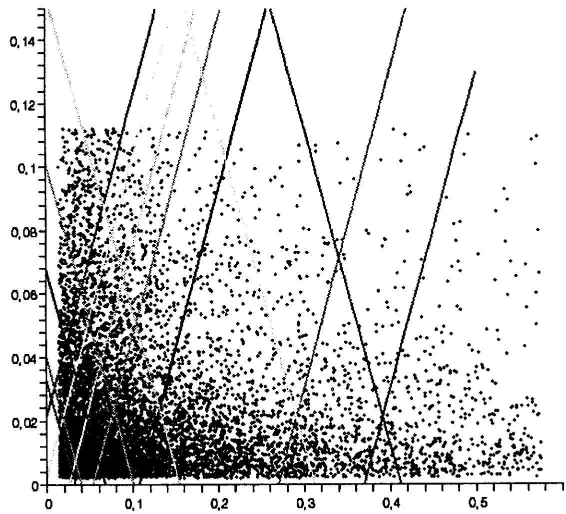

Figure 17. First factor (j=1), $D.F. = 2/2$

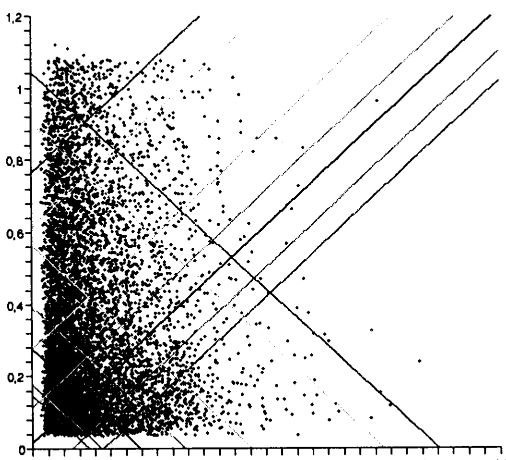

Figure 18. Second factor (j=2), $D.F. = 1/8/2/4$

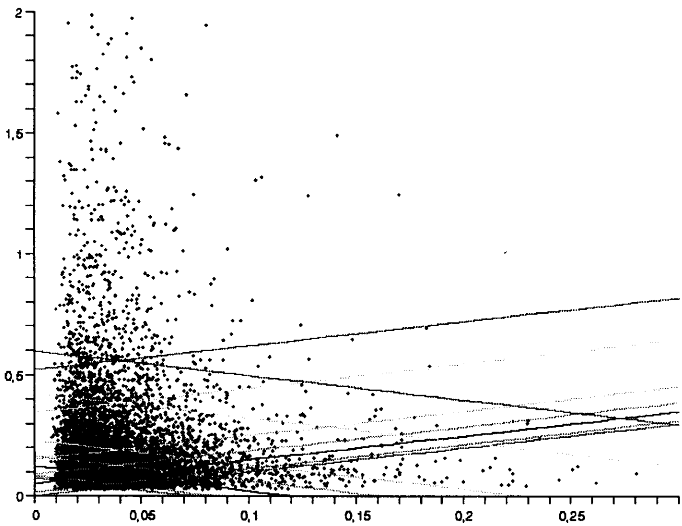

Figure 19. Third factor (j=3), $D.F. = 4/8$

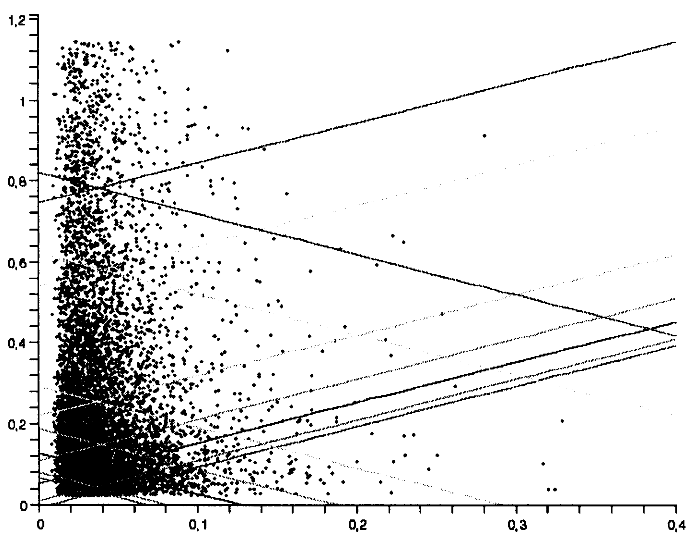

Figure 20. Interaction of the first and the second factor (j=4), $D.F. = 2/8$

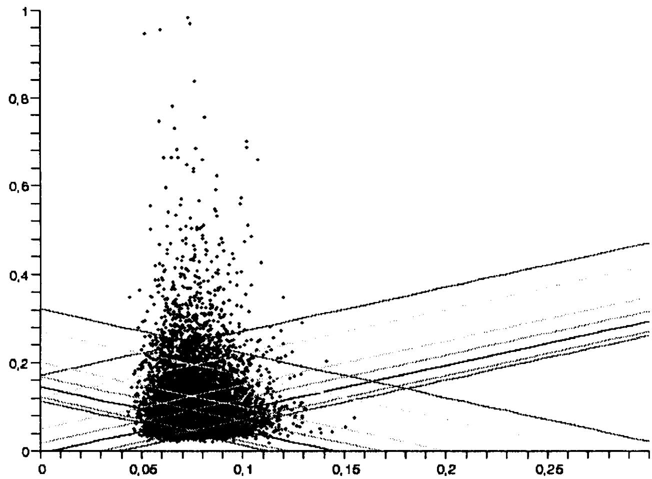

Figure 21. Interaction of the first and the third factor (j=5), $D.F. = 8/72$

In table 4 we present, for the positive and negative parts of the variance components, the *UMVUE* and the empirical quantiles $w_{q,i,j}$ and $z_{q,i,j}$, for the sum and difference. Thus empirical quantiles for the variance components are presented in this table as $z_{q,i,j}$.

Table 4.

Factor	$\bar{\sigma}^2_{c,i,j}$	Quancue	5%	10%	30%	50%	70%	90%	95%
1	(0.05744;0.01125)	$w_{q,1,0}$	0.0326	0.0400	0.0678	0.1011	0.1549	0.3120	0.4125
		$z_{q,1,0}$	-0.3698	-0.2696	-0.1073	-0.0535	-0.0247	0.0026	0.0216
2	(0.07307;0.22306)	$w_{q,0,1}$	0.1470	0.1779	0.2770	0.3941	0.5658	0.9022	1.0437
		$z_{q,0,1}$	-0.1792	-0.0981	0.0181	0.1142	0.2753	0.6147	0.7656
3	(0.02847;0.10432)	$w_{q,0,2}$	0.0755	0.0873	0.1243	0.1659	0.2316	0.4257	0.5998
		$z_{q,0,2}$	0.0016	0.0155	0.0515	0.0896	0.1534	0.3471	0.5219
1 x 2	(0.02847:0.11488)	$w_{q,1,1}$	0.0693	0.0818	0.1287	0.1886	0.2927	0.6216	0.8209
		$z_{q,1,1}$	-0.0053	0.0109	0.0542	0.1116	0.2204	0.5407	0.7466
1 x 3	(0.07343:0.05640)	$w_{q,1,2}$	0.1151	0.1235	0.1469	0.1698	0.2016	0.2732	0.3228
		$z_{q,1,2}$	-0.0371	-0.0270	-0.0040	0.0194	0.0515	0.1215	0.1731

We point out that, in all these figures, the red dot lies near the intersection of the quantile line for $q = 0.5$. This gives us an additional validation of the simulation technique used.

In Table 5 we can see that the point whose coordinates are the $UMVUE$ is always covered by the confidence limits, for $\alpha = 10\%$.

Table 5.

Factor	UMVUE	$z_{0.05,i,j}$	$z_{0.95,i,j}$	D.F.
1	-0.0462	-0.3698	0.0216	2/2
2	0.1500	-0.1792	0.7656	1/8/4/2
3	0.0759	0.0016	0.5219	6/8
1 x 2	0.0864	-0.0053	0.7466	2/8
1 x 3	0.0120	-0.0371	0.1215	8/72

As a parting remark, we point out that through the validation we have unified two distinct developments emerging from the use of sufficient complete statistics.

REFERENCES

Ferreira D., Mexia J.T., Saraiva S. (2005). *Stable hypothesis for mixed models with balanced cross-nesting*. Accepted by Discussiones Mathematicae - Probability and Statistics.

Fonseca M., Mexia J.T., Zmyślony R. (2003). *Estimating and Testing of Variance Components: An Application to a Grapevine Experiment*. Biometrical Letters, 40,(1), 1–7.

Khuri A.I., Matthew T., Sinha B.K., (1998). *Statistical Tests for Mixed Linear Models*. John Wiley & Sons, New York.

Williams (1997). *Probability with martingales*. Cambridge University Press.