Discussiones Mathematicae
Probability and Statistics 35 (2015) 45–56
doi:10.7151/dmps.1175

CONFIDENCE INTERVALS FOR LARGE NON-CENTRALITY PARAMETERS

Sonia Inacio

FCT New University of Lisbon
Department of Mathematics, Portugal

e-mail: sonia.maria.inacio@gmail.com

Manuela M. Oliveira

University of Évora, Department of Mathematics
Center for Research on Mathematics and its Applications, Portugal

e-mail: mmo@uevora.pt

and

João T. Mexia

FCT New University of Lisbon
Department of Mathematics, Portugal

e-mail: jmt@fct.unl.pt

Abstract

We use asymptotic linearity to derive confidence intervals for large non-centrality parameters. These results enable us to measure relevance of effects and interactions in multifactors models when we get highly statistically significant the values of F tests statistics. We show how to use our approach by considering two sets of data as application examples.

Keywords: asymptotic linearity, non-centrality parameters, highly significant F tests, measure relevance.

2010 Mathematics Subject Classification: 62F03.

1. Introduction

Applying ANOVA for multifactors models we may have several highly significant F tests. The corresponding statistics will have, quite possibly, F distributions with large non-centrality parameters. The larger one such parameter the more relevant are the corresponding effects or interactions. We were then led to derive confidence intervals for large non-centrality parameters.

Let $U \sim \sigma^2\chi^2_{k,\delta}$ represent the product by σ^2 of a chi-square with k degrees of freedom and non-centrality parameter δ. If U is independent of $V \sim \sigma^2\chi^2_g$ we can use the fact that

$$\mathfrak{F} = \frac{g}{k}\frac{U}{V}$$

has F distribution with k and g degree of freedom and non-centrality parameter δ, $F(f \mid k, g, \delta)$, which is a decreasing function of δ to, given the value f taken by $\mathfrak{F}$, obtain confidence intervals for δ when $\mathfrak{F}$ is not too large. Thus solving, in order to δ, the equation

$$F(\mathfrak{F}|k, g, \delta) = 1 - p$$

we get the $p-th$ quantile δ_p for δ and it is straightforward to build the confidence intervals. But when $\mathfrak{F}$ is large we will have for δ, that are not extremely large,

$$F(\mathfrak{F}|k, g, \delta) \approx 1$$

and so we cannot apply this approach to obtain stochastic bounds for large non centrality parameters. Now we can always assume that

$$U = \sum_{j=1}^{k} Y_j^2$$

with $Y_1, \dots, Y_k$ normal, independent with mean values $\mu_1, \dots, \mu_k$ and variance σ^2. Namely we will have $\delta = \frac{1}{\sigma^2}\|\boldsymbol{\mu}\|^2$. In the next section we use assymptotic linearity, see Mexia & Oliveira (2011), to show that the limit distribution of

$$t = \frac{\|\underline{Y}\| - \|\underline{\mu}\|}{2\sqrt{V}g},$$

when $\|\underline{\mu}\| \to \infty$, is indeed a t distribution with g degrees of freedom. We will use that distribution to obtain the bounds for δ. We will also obtain confidence intervals for $\|\boldsymbol{\mu}\|^2$. In Section 4 we obtain, through duality, tests for hypothesis on δ. In Section 5 we will apply our approach to STATIS methodology and to age of death cancer patients.

2. LIMIT DISTRIBUTION

We start by introducing assymptotic linearity. Given $\boldsymbol{x} \varepsilon \boldsymbol{R}^k$ let the function $g(\boldsymbol{x})$ have gradient $\underline{\boldsymbol{g}}(\boldsymbol{x})$ and hessian matrix $\underline{\underline{\boldsymbol{g}}}(\boldsymbol{x})$ with spectral radius $r(\boldsymbol{x})$. Putting

$$P_d(\boldsymbol{x}) = Sup\{r(\boldsymbol{x}+\boldsymbol{\mu}); \|\boldsymbol{\mu}\| \leq d\},$$

$g(\cdot)$ will enjoy assymptotic linearity if $\forall d > 0$

$$u_d(h) = Max\left\{\frac{P_d(\boldsymbol{x})}{\|\underline{\boldsymbol{g}}(\boldsymbol{x})\|}, \|\boldsymbol{x}\| > h\right\} \xrightarrow[h\to\infty]{} 0.$$

Then, whatever the distribution F_e of e, with

$$Z = \frac{g(\underline{\mu}+\underline{e}) - g(\underline{\mu})}{\|\underline{g}(\underline{\mu})\|}$$

and

$$Z^{\circ} = \frac{\underline{g}(\underline{\mu})'\underline{e}}{\|\underline{g}(\underline{\mu})\|},$$

we have

$$Sup\{\|F_Z(z) - F_{Z^{\circ}}(z)\|\} \xrightarrow[\|\boldsymbol{\mu}\|\to\infty]{} 0,$$

(e.g. Mexia & Oliveira (2010)). Thus if, Z° has limit distribution F°, when $\|\boldsymbol{\mu}\| \to \infty$, Z will have the same limit distribution. Namely, for

$$g(\underline{x}) = \|\underline{\boldsymbol{x}}\|^2,$$

we have

$$\underline{g}(\underline{x}) = 2\underline{x}$$

$$\underline{\underline{g}}(\underline{x}) = 2I_k$$

so, clearly, $g(\underline{x})$ enjoys assymptotical linearity. Moreover, if $U \sim \sigma^2\chi^2_{k,\delta}$ we can always assume that

$$U = g(\underline{\mu}+\underline{e}),$$

where $\underline{e}$ is normal with null mean vector and variance covariance matrix $\sigma^2 I_k, \underline{e} \sim \mathbf{N}(\underline{0}, \sigma^2 I_k)$. Then,

$$\delta = \frac{\|\underline{\boldsymbol{\mu}}\|^2}{\sigma^2},$$

and

$$Z^\circ = \frac{\underline{\mu}^t \underline{e}}{\|\underline{\boldsymbol{\mu}}\|^2} \sim \mathbf{N}(0, \sigma^2).$$

So, F_{Z° will not depend on $\underline{\mu}$, and the limit distribution of

$$Z = \frac{\|\underline{\mu} + \underline{e}\|^2 - \|\underline{\mu}\|^2}{2\|\underline{\mu}\|},$$

when $\|\underline{\mu}\| \to \infty$, will also be $\mathbf{N}(0, \sigma^2)$. With z_q the well known bounds for tails with probability $q/2$ for $\mathbf{N}(0,1)$, and taking $\underline{Y} = \underline{\mu} + \underline{e}$, we will have

$$Pr\left(-z_q\sigma \le \frac{\|\underline{Y}\|^2 - \|\underline{\mu}\|^2}{2\|\underline{\mu}\|} \le z_q\sigma\right) = 1 - q,$$

which is equivalent to

$$(1) \quad \begin{aligned} Pr\Bigg(& \|\underline{Y}\|^2 + 2z_q^2\sigma^2 - \sqrt{4z_q^2\sigma^2\|\underline{Y}\|^2 + 4z_q^2\sigma^4} \le \|\underline{\mu}\|^2 \le \|\underline{Y}\|^2 + 2z_q^2\sigma^2 \\ & + \sqrt{4z_q^2\sigma^2\|\underline{Y}\|^2 + 4z_q^2\sigma^4}\Bigg) = 1 - q \end{aligned}$$

and to

$$\begin{aligned} Pr\Bigg(& 1 + 2z_q^2\frac{\sigma^2}{\|\underline{Y}\|^2} - 2z_q\frac{\sigma}{\|\underline{Y}\|}\sqrt{1 + z_q^2\frac{\sigma^2}{\|\underline{Y}\|^2}} \le \frac{\|\underline{\mu}\|^2}{\|\underline{Y}\|^2} \le 1 + 2z_q^2\frac{\sigma^2}{\|\underline{Y}\|^2} \\ & + 2z_q\frac{\sigma}{\|\underline{Y}\|}\sqrt{1 + z_q^2\frac{\sigma^2}{\|\underline{Y}\|^2}}\Bigg) = 1 - q. \end{aligned}$$

In many instances $\frac{\sigma}{\|\underline{Y}\|}$ is quite small so the relative error ε_r of $\|\underline{Y}\|^2$ as estimator of $\|\underline{\mu}\|^2$ is also quite small. However the confidence interval for $\|\underline{\mu}\|^2$ given by (1) has the draw-back of it's bounds depending on σ^2. Let us now assume $\underline{Y}$ to be independent from $V \sim \sigma^2\chi_g^2$. Now, with $\xrightarrow{p}$ indicating stochastic convergence, we have $\frac{\|\underline{Y}\|}{\|\underline{\mu}\|} \xrightarrow[\|\underline{\boldsymbol{\mu}}\|\to\infty]{p} 1$ so, since the limit distribution of

$$Z = \frac{\|\underline{Y}\|^2}{2\|\underline{\mu}\|^2} - \frac{\|\underline{\mu}\|}{2}$$

is $\mathbf{N}(0,\sigma^2)$ the corresponding limit distribution of $\frac{1}{2}(\|\underline{Y}\| - \|\underline{\mu}\|)$ will also be $\mathbf{N}(0,\sigma^2)$ and the limit distribution of

$$t = \frac{\|\underline{Y}\| - \|\underline{\mu}\|}{2\sqrt{V}g},$$

will be the t distribution with g degrees of freedom (see Ferreira *et al.* (2013) and Mexia (1992)). For the tails, with probability $\frac{q}{2}$, of this distribution we have the bounds $t_{g,q}$. Now we have

$$Pr\left(-2t_{g,q}\sqrt{\frac{V}{g}} \leq \|\underline{Y}\| - \|\underline{\mu}\| \leq 2t_{g,q}\sqrt{\frac{V}{g}}\right) = 1-q$$

which is equivalent to

$$Pr\left((\|\underline{Y}\| - 2t_{g,q}\sqrt{\frac{V}{g}})^2 \leq \|\underline{\mu}\|^2 \leq (\|\underline{Y}\| + 2t_{g,q}\sqrt{\frac{V}{g}})^2\right) = 1-q.$$

These bounds indicate that, when $\frac{V}{\|\underline{Y}\|^2}$ is small, ε_r will be small. When we have small relative errors for estimating $\|\underline{\mu}_i\|^2, i = 1,\dots,h$ by the $\|\underline{Y}_i\|^2, i = 1,\dots,h$ we will also have small relative errors for estimating the

$$\tau_i = \frac{\|\underline{\mu}_i\|^2}{\sum_{l\neq i}\|\underline{\mu}_l\|^2},\ i = 1,\dots,h$$

by the

$$\tilde{\tau}_i = \frac{\|\underline{Y}_i\|^2}{\sum\limits_{l\neq i}\|\underline{Y}_l\|^2}, i = 1,\dots,h.$$

With $x_{g,p}$ the $p-th$ quantile for χ^2_g, we have for σ^2 and $\frac{1}{\sigma^2}$, the $1-q'$ level confidence intervals

$$\left[\frac{V}{x_{g,q'/2}}; \frac{V}{x_{g,1-q'/2}}\right]$$

and

$$\left[\frac{x_{g,q'/2}}{V}; \frac{x_{g,1-q'/2}}{V}\right].$$

Moreover, when

$$\left(\|\underline{Y}\| - 2t_{g,q}\sqrt{\frac{V}{g}}\right)^2 \leq \|\underline{\mu}\|^2 \leq \left(\|\underline{Y}\| + 2t_{g,q}\sqrt{\frac{V}{g}}\right)^2$$

and

$$\frac{x_{g,q'/2}}{V} \leq \frac{1}{\sigma^2} \leq \frac{x_{g,1-q'/2}}{V}$$

we have, since $\delta = \frac{\|\underline{\mu}\|^2}{\sigma^2}$,

$$\left(\|\underline{Y}\| - 2t_{g,q}\sqrt{\frac{V}{g}}\right)^2 \frac{x_{g,q'/2}}{V} \leq \delta \leq \left(\|\underline{Y}\| + 2t_{g,q}\sqrt{\frac{V}{g}}\right)^2 \frac{x_{g,1-q'/2}}{V}.$$

Thus, according to the BOOLE inequalities we have

$$Pr\left[\left(\|\underline{Y}\| - 2t_{g,q}\sqrt{\frac{V}{g}}\right)^2 \frac{x_{g,q'/2}}{V} \leq \delta \leq \left(\|\underline{Y}\| + 2t_{g,q}\sqrt{\frac{V}{g}}\right)^2 \frac{x_{g,1-q'/2}}{V}\right] = 1-q-q'.$$

Usually we can take q much smaller than q' since the relative precision we have for estimating $\|\underline{\mu}\|^2$ by $\|\underline{Y}\|^2$ is much higher than that of estimating σ^2 by $\frac{V}{g}$.

3. Tests for hypothesis on δ

Given ω a sub-space with dimension p of Ω which has dimension m,

$$\overline{\omega} = \omega^{\perp} \cap \Omega$$

will have dimension $m-p$. Moreover given $\underline{Y} \sim \mathbf{N}(\mu, \sigma^2 I_n)$ the F test statistic for testing

$$H_0 : \underline{\mu} \in \omega$$

assuming $\mu \in \Omega$ will have F distribution with $m-p$ and $n-m$ degrees of freedom and non centrality parameter

$$\delta = \frac{\|\underline{\mu_{\overline{\omega}}}\|^2}{\sigma^2}$$

with $\underline{v}_{\nabla}$ the orthogonal projection of $\underline{v}$ on ∇. Now H_0 holds if and only if $\delta = 0$. Moreover δ may be used to measure the "distance" between the

$$H(\underline{b}) : \underline{\mu} - \underline{b} \in \omega$$

and

$$H_0 = H(\underline{0}).$$

Let the row vectors of A constitute an orthonormal basis for $\overline{\omega}$. We then can obtain the components $Y_1^\circ \dots Y_{m-p}^\circ$ of $A\underline{Y}$. Since δ is the non-centrality parameter for this sample it is quite straightfoward to apply our results. Actually this approach may be useful when we have an orthogonal partition

$$\Omega = \bigoplus_{j=1}^{l} \overline{\omega}_j$$

and, assuming that $\underline{\mu} \in \Omega$, we want to test

$$H_{0,j} : \underline{\mu} \in \omega_j = \overline{\omega}_j^{\perp} \cap \Omega, j = 1, \dots, l.$$

Then it may happen that several hypothesis are rejected which may render the results difficult to analyse. In such cases we may relied on the corresponding non centrality parameters to see which rejections were really significant. We may then use the relevance measures

$$\lambda_j = \frac{\delta_j}{\sum_{\mathbf{j}'\varepsilon\mathcal{C}} \delta_{\mathbf{j}'}}, j = 1, \dots, l$$

associated to the different hypothesis established then by the

$$\widetilde{\lambda}_j = \frac{\widetilde{\delta}_j}{\sum_{\mathbf{j}'\varepsilon\mathcal{C}} \widetilde{\delta}_{\mathbf{j}'}}, j = 1, \dots, l$$

with $\mathcal{C}$ the set of indexes for which we have significant F test.

If the F statistics $\mathbf{F}_j, j = 1, \dots, l$, had g_j and g degrees of freedom we have, see Mexia 1992, the estimator

$$\widetilde{\delta}_j = (g-2) \times \frac{2}{g_j - 1} F - g_j \approx (g-2) \times \overline{F} - g_j.$$

4. An application

4.1. A first application

The data considered in this application is for the results of local elections in six districts of Portugal corresponding to the years 1976 to 1997.

These data was treated using an ANOVA based approach for STATIS Methodology see Oliveira & Mexia (2007) and applied to six districts, Braga, Vila Real, Guarda, Leiria, Setúbal and Beja. Grouped into the following factors and levels:

– Factor 1: Longitude - with two levels: Coast and Interior;

– Factor 2: Administrative importance - with two levels: district capital and other non-capital municipalities;

– Factor 3: Latitude - with three levels: North, Central and South.

The action of these factors on the evolution of the elections results was one of the points considered. From the analysis of effects on three factors we obtained the F tests results. The significant results are marked with $*$.

The distribution of $\overline{F}_i\{1\}$ e $\overline{F}_i\{2\}$ is $\overline{\mathbf{F}}(z|1,252)$ and of $\overline{F}_i\{3\}$ is $\overline{\mathbf{F}}(z|2,252)$.

Table 1. F tests for the factors effects.

Studies	$\overline{F}_i\{1\}$	$\overline{F}_i\{2\}$	$\overline{F}_i\{3\}$
1976	$244,33*$	$447,94*$	$1839,41*$
1979	$35,94*$	$198,47*$	$1217,37*$
1982	$245,74*$	$205,1*$	$1498,04*$
1985	$47,52*$	$180,02*$	$1209,28*$
1989	$47,23*$	$36,73*$	$983,33*$
1993	$9,79*$	$58,96*$	$703,78*$
1997	0,19	$19,95*$	$722,48*$

In the analysis of effect for the first factor in the year of 1997 there is no significant difference between the municipalities of the coast and interior. For the remaining factors the effects are significant in all years. Being higher for the third factor (North, Central and South).

In the analysis of interactions between factors, the distribution of $\overline{F}_i\{1,2\}$ is $\overline{\mathbf{F}}(z|1,252)$ and of $\overline{F}_i\{1,3\}, \overline{F}_i\{2,3\}, \overline{F}_i\{1,2,3\}$ is $\overline{\mathbf{F}}(z|2,252)$. From the F tests for the interactions we found that:

– Between the first and the second factor there was no interaction in the year 1989, and in the years 1979 and 1982 there was a higher interaction;

– Between the first and the third factor there was interaction for all years;

– Between the second and third factor there was no interaction from the year 1989 and for the year 1982 there was a higher interaction;

– Among the three factors there was interaction in every year, emphasizing a higher interaction in year 1976.

Table 2. F tests for the factors interactions.

Studies	$\overline{F}_i\{1,2\}$	$\overline{F}_i\{1,3\}$	$\overline{F}_i\{2,3\}$	$\overline{F}_i\{1,2,3\}$
1976	$35,63*$	$307,27*$	$31,44*$	$70,35*$
1979	$218,35*$	$209,9*$	$95,54*$	$41,15*$
1982	$118,98*$	$166,55*$	$118,18*$	$42,95*$
1985	$49,4*$	$80,71*$	$65,8*$	$8,97*$
1989	0,25	$38,26*$	0,12	$47,44*$
1993	$20,45*$	$9,45*$	0,58	$27,31*$
1997	$11,23*$	$39,3*$	2,22	$19,7*$

Since there are a large number of highly significant effects and interactions, we refine the analysis by the use of non-centrality parameter whenever $g_j > 1$. Which led us to obtained the estimates for the non-centrality parameters.

For the $\overline{F}_i\{1\}, \overline{F}_i\{2\}$ and $\overline{F}_i\{1,2\}$ we could not estimate the non-centrality parameter.

In the table below *n.s.* means *non significant*.

Table 3. Estimators for non-centrality parameter δ.

Studies	$\widetilde{\delta}\{3\}$	$\widetilde{\delta}\{1,3\}$	$\widetilde{\delta}\{2,3\}$	$\widetilde{\delta}\{1,2,3\}$
1976	3647,63	607,65	60,37	137,59
1979	2413,42	414,46	187,56	79,65
1982	2970,3	328,46	232,49	83,23
1985	2397,37	158,13	128,56	15,8
1989	1949,05	73,91	n.s.	92,12
1993	1394,39	16,74	n.s.	52,18
1997	1431,49	75,98	n.s.	37,09

We may use these estimators to evaluate the relevance measures

$$r(\mathcal{C}) = \frac{\tilde{\delta}(\mathcal{C})}{\sum_{\{3\}\subseteq\mathcal{C}'} \tilde{\delta}(\mathcal{C}')},$$

with $\{3\} \subseteq C$ assuming that $\tilde{\delta}(\emptyset) = 0$.

The results show the strong predominance of the third factor, with levels North, Center and South, in all studies.

Table 4. Relevance estimators.

Studies	$\{3\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
1976	81,9%	13,6%	1,4%	3,1%
1979	78%	13,4%	6,1%	2,6%
1982	82,2%	9,1%	6,4%	2,3%
1985	88,8%	5,9%	4,8%	0,6%
1989	92,2%	3,5%	-	4,4%
1993	95,3%	1,1%	-	3,6%
1997	92,7%	4,9%	-	2,4%

We also use the asymptotic linearity for confidence intervals for the parameters of non centrality.

Table 5. Confidence Intervals (95%) for δ - Effects and Interactions.

Studies	C.I. $\widetilde{\delta}\{3\}$	C.I. $\widetilde{\delta}\{1,3\}$	C.I. $\widetilde{\delta}\{2,3\}$	C.I. $\widetilde{\delta}\{1,2,3\}$
1976	$[3051,89;4366,39]*$	$[506,75;733,75]*$	$[50,73;76,68]*$	$[114,74;169,85]*$
1979	$[2017,9;2892,53]*$	$[345,43;502,29]*$	$[156,29;229,97]*$	$[66,68;99,99]*$
1982	$[2484,38;3557,62]*$	$[273,68;399,17]*$	$[193,7;283,97]*$	$[69,64;104,31]*$
1985	$[2004,45;2873,35]*$	$[131,82;194,58]*$	$[107,24;158,98]*$	$[14,07;22,48]*$
1989	$[1629,01;2337,76]*$	$[61,92;93,05]*$	-	$[77,01;115,05]*$
1993	$[1164,7;1674,89]*$	$[14,84;23,64]*$	-	$[43,97;66,77]*$
1997	$[1195,75;1719,23]*$	$[63,63;95,56]*$	-	$[31,52;48,45]*$

Duality enables us to test hypothesis of nullity for non-centrality parameter. When the $1-p$ level intervals does not contain the origin we may reject at the p level the nullity hypothesis.

4.2. A second application

We now consider an application to age of death of cancer patients. In this application the factors considered were:

– Factor 1: Pathology with three levels: leukaemia, thyroid cancer and pancreas cancer;

– Factor 2: Gender.

Thus besides the two factors we had to consider the interaction. The $\overline{F}$ test statistics were obtained in Nunes et al (2014).

The distribution of $\overline{F}\{1\}$ and $\overline{F}\{1{\times}2\}$ is $\mathbf{F}(z|2,311)$ and of $\overline{F}\{2\}$ is $\mathbf{F}(z|1,311)$. We then can estimates the non centrality parameters, whenever $g_j > 1$, corre-

Table 6. $\overline{F}$ tests for the factors effects and interaction.

$\overline{F}\{1\}$	$\overline{F}\{2\}$	$\overline{F}\{1\times 2\}$
$0,154*$	$0,0006$	$1,524*$

sponding to the first factor and the interaction obtaining the results in the table below.

Table 7. Estimators for non-centrality parameter δ.

$\widetilde{\delta}\{1\}$	$\widetilde{\delta}\{1,2\}$
45,186	468,978

For the second factor we could not estimate the non-centrality parameter. However since the corresponding F tests were not significant we did not have to consider them.

We may get for the relative relevance, the estimators

$$\widetilde{\nu}_1 = 100 \times \frac{\widetilde{\delta}_1}{\widetilde{\delta}_1 + \widetilde{\delta}_{1\times 2}} = 8,7\%$$

$$\widetilde{\nu}_{1\times 2} = 100 \times \frac{\widetilde{\delta}_{1\times 2}}{\widetilde{\delta}_1 + \widetilde{\delta}_{1\times 2}} = 91,3\%$$

This example is interesting in which it shows that for two hypothesis with highly significant statistics the relevance of one of them was much higher than that of the other.

To obtain the confidence intervals for the parameters of non centrality we use the asymptotic linearity and then through duality we tested for nullity.

Table 8. Confidence Intervals (95%) for δ.

C.I. $\widetilde{\delta}\{1\}$	C.I. $\widetilde{\delta}\{1,2\}$
$[39,22;53,74]*$	$[397,83;545,04]*$

4.3. Final remarks

The concept of significance is of utmost importance in statistical inference. Nevertheless it may be completed using the concept of relevance measure in situations in which we have several highly significant tests. As the first application using STATIS methodology clearly shows the relevance may differ largely between highly significant results. Thus, some of these may be responsible for a large portion of the differences observed being the relevant ones.

References

[1] D. Ferreira, S.S. Ferreira, C. Nunes and S. Inácio, *Inducing pivot variables and non-centrality parameters in elliptical distributions*, AIP Conf. Proc. **1558** (2013) 833.

[2] J.T. Mexia, Assymptotic Chi-squared Tests, Design and Log-Linear Models (Trabalhos de Investigação, 1. Departamento de Matemática, Faculdade de Ciȩncias e Tecnologia, Universidade Nova de Lisboa, 1992).

[3] J.T. Mexia and M.M. Oliveira, *Asymptotic linearity and limit distributions, approximations*, Journal of Statistical Planning and Inference **140** (2011) 353–357.

[4] M.M. Oliveira and J.T. Mexia, *ANOVA like analysis of matched series of studies with a common structure*, Journal of Statistical Planning and Inference **137** (2007) 1862–1870.

[5] C. Nunes, S.S. Ferreira and J.T. Mexia, *Fixed effects NOVA: an extension to samples with random size*, Journal of Statistical Computation and Simulation **84** (2014) 2316–2328.

[6] H. Scheffé, The Analysis of Variance (New York-John Wiley & Sons, 1959).

Received 15 March 2015