F tests with random samples size. Theory and applications

Elsa E. Moreira[a,*], João T. Mexia[a], Christoph E. Minder[b]

[a] *CMA - Center of Mathematics and Applications, Faculty of Sciences and Technology, Nova University of Lisbon, Campus Caparica, 2829-516 Caparica, Portugal*
[b] *Horten Zentum, Medical Faculty, University of Zurich*

Abstract

Given a time span for collecting observations in a comparison study, it is advisable to consider the samples size of the ANOVA levels as random variables. More powerful F tests with F distribution conditional to $N = n$, leading to lower critical values, are developed. The approach is used to obtain the minimum duration for the data collection that ensures a pre-fixed test power.

Keywords: ANOVA, F distribution, Pathologies comparison, Poisson distribution, Power analysis

1. Introduction

The theoretical developments presented in this paper were motivated by a real case situation in the field of medicine. Consider the numbers of patients that arrive to an hospital with different pathologies during a given time span. These numbers cannot be known in advance because, if we decide to do the counting in a different time period of the same length, the number of patients obtained with those pathologies will be different from the first counting. So, if, due to limitations in the budget, we have to conduct just one study to compare the pathologies, it is more correct to consider the sample dimensions as realizations of random variables. The data will be collected from the patients with each pathology as soon as they present themselves.

*Corresponding author
Email address: efnm@fct.unl.pt (Elsa E. Moreira)

This situation arises mostly because there is a given time span for collecting the observations. After the elapsed time interval, the F test statistics can be obtained to test hypotheses about the different means values. In what follows, it is assumed that the sample dimensions $n_1, ..., n_k$ for the k pathologies are realizations of independent Poisson variables with parameters $\lambda_1, ..., \lambda_k$ and the observations in these sample are normal, independent with variance σ^2. As a result, the F statistic to test the null hypotheses $H_0 : \mu_1 = ... = \mu_k$ will have conditional F distribution on the number of observations as will be seen in the next section.

The approach presented is also useful while planning studies, before the data are in. The minimum data collection duration to ensure, with a given probability, reliable results will be obtained. Thus, when all k pathologies are covered, the conditional distribution of the F test statistic has $k-1$ and $n-k$ degrees of freedom, with n the sum of all $n_i, i = 1, ..., k$ and non-centrality parameter δ, which is null when the mean values $\mu_1, ..., \mu_k$ for different pathologies are equal. Thus, δ will measure the "distance" of the alternatives from H_0. So, we may look for the minimum duration that ensures, with a given probability, that all pathologies have at least a minimum number of observations that allow the conditional power of the α level test for a given δ to be sufficient high. Therefore, in sub-section 3.1, we see how to determine these minimum samples sizes $\dot{n}_i, i = 1, ..., k$ to have a sufficient powerful F test, while in sub-section 3.2, the minimum time span for, with a probability p, getting these $\dot{n}_1, ..., \dot{n}_k$ is obtained. Finally, an application with simulated data is presented in section 5 to illustrate the methodology.

2. Conditional F distributions

Let the components of the vector $\mathbf{n} = (n_1, ..., n_k)$ of the samples sizes be realizations of the components of the vector $\mathbf{N} = (N_1, ..., N_k)$. The components of $\mathbf{N}$ are independent Poisson variables, with vector of parameters $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_k)$. Assuming that, when $\mathbf{n} > \mathbf{0}$, the samples $x_{i,1}, ..., x_{i,n_i}$, $i = 1, ..., k$ are normal with mean values $\mu_1, ..., \mu_k$ and variance σ^2, for testing $H_0 : \mu_1 = ... = \mu_k$ the $\mathcal{F}$ test statistic

$$\mathcal{F} = \frac{n-k}{k-1} \frac{\sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{n}}{\sum_{i=1}^{k} S_i}$$

is used, where $T_i = \sum_{j=1}^{n_i} x_{i,j}$, $S_i = \sum_{j=1}^{n_i} x_{i,j}^2 - T_i^2/n_i$, $T = \sum_{i=1}^{k} T_i$ and $n = \sum_{i=1}^{k} n_i$.

Due to our assumption for N, the $\mathcal{F}$ statistics will have F distribution conditional to $\mathbf{N} = \mathbf{n}$, with $k-1$ and $n-k$ degrees of freedom and non-centrality parameter

$$\delta = \frac{1}{\sigma^2}\sum_{i=1}^{k} n_i(\mu_i - \mu.)^2$$

where

$$\mu. = \frac{1}{n}\sum_{i=1}^{k} n_i\mu_i$$

is the general mean value (Hocking, 2003; Mexia, 1990).

In a previous paper (Mexia and Moreira, 2010) the unconditional distribution for the $\mathcal{F}$ statistics was derived. This distribution is given by the following series

$$\dot{F}(z) = \sum_{\mathbf{n}>\mathbf{0}} q(\mathbf{n})F(z|k-1, n-k, \delta(\mathbf{n})) \tag{1}$$

whose terms corresponds to all the vectors $\mathbf{n} = (n_1, ..., n_k)$ with $n_i > 0, i = 1, ..., k$ and where

$$q(\mathbf{n}) = \prod_{i=1}^{k} \frac{e^{-\lambda_i}\frac{{\lambda_i}^{n_i}}{n_i!}}{1-e^{-\lambda_i}}.$$

However, when the null hypothesis $H_0 : \mu_1 = ... = \mu_k$ holds, $\delta(\mathbf{n}) = 0, \forall \mathbf{n} > 0$ and

$$\mathcal{F} \sim \dot{F}_0(z) = \sum_{\mathbf{n}>\mathbf{0}} q(\mathbf{n})F(z|k-1, n-k). \tag{2}$$

Notice that in eq. (1), the n for the degrees of freedom of the denominator does not change within the sum.

To compute the values of $\dot{F}_0(z)$ the corresponding series in eq. (2) must be obviously truncated. The truncation error of this series when restricting ourselves to samples with $\mathbf{n} \leq \mathbf{n}^o$ was studied in a previous paper (Mexia and Moreira, 2010). In that paper, it was shown that the truncation error does not exceed a bound

$$b(\mathbf{n}^o) < \frac{k\varepsilon}{(1-e^{-\lambda^o})^k},$$

if each n_i^o, the components of $\mathbf{n}^o$, are chosen such

$$\sum_{n_i=0}^{n_i^o} e^{-\lambda_i}\frac{\lambda_i^{n_i}}{n_i!} > 1-\varepsilon, i = 1, ..., k, \tag{3}$$

with ε small (Mexia and Moreira, 2010). So, from this inequality we can obtain the minimal samples sizes that allow us to control the truncation error for the distribution $\dot{F}_0(z)$.

In Table 1 is presented for different values of the minimal average sample size, $\lambda^o = Min\{\lambda_1, ..., \lambda_k\}$ and different small ε, the minimal samples dimension n^o to have satisfy inequality (3). These truncation errors are perfectly controlled even if the samples dimensions are relatively small, i.e., for instance for $k = 3$, if the minimum of the $\lambda_i, i = 1, 2, 3$ is $\lambda_0 = 1$, for $\varepsilon = 10^{-6}$ the sample dimensions should not be less than 9. In this example, the truncation error has an upper bound of 0.000012. The critical values tables in common use have in general 3 decimal places of precision, thus truncating the distribution with an error of 0.000012 is more than needed to get a critical value with good accuracy.

Table 1: The minimal dimension of the samples n^o required to have the truncation error controlled

	λ^o				
	1	2	5	10	20
$\varepsilon = 10^{-4}$	6	9	15	24	39
$\varepsilon = 10^{-6}$	9	12	19	28	45
$\varepsilon = 10^{-8}$	11	14	22	32	50

Theses formulas enable us to solve the equation $\dot{F}_0(z) = 1 - \alpha$ for z, to obtain the $(1 - \alpha)$-th quantiles, i.e, the critical values of α level tests. After obtaining the quantiles for the distribution $\dot{F}_0(z)$, the inference for the one-way ANOVA to compare the k pathologies proceeds in the usual way (Scheffé, 1959; Montgomery, 1997).

In order to demonstrate the advantage of using F tests with random samples size, abbreviated $\dot{F}$ tests, we were led to obtain several critical values for 5% of probability for the case of a test with 3 pathologies and compare these with the critical values using an usual F test. The values, presented in Table 3, were computed assuming increasing different values for λ_1, λ_2 and λ_3 that originate consequently different values of n. In Figure 1 the curves for the critical values in Table 2 are presented. From that Figure we see that the critical values for the $\dot{F}$ test are always lower than the critical values for the usual F test. Although, the difference between both decrease and tend to 0 as n increases.

Table 2: Critical values for F tests with random sample size and for usual F tests, case for compare 3 pathologies

				Critical Values (5%)	
λ_1	λ_2	λ_3	n	$\dot{F}$	F
0,05	0,05	0,05	9	4.256	5.143
0,1	0,1	0,1	12	3.885	4.256
0,1	0,3	0,5	16	3.634	3.806
1	1	1	27	3.354	3.403
2	2	1	33	3.285	3.316
4	3	2	43	3.214	3.232
5	4	4	53	3.172	3.183
6	5	4	57	3.159	3.168
8	8	6	71	3.126	3.132
11	11	10	88	3.100	3.104
16	16	16	114	3.076	3.078

3. Power analysis

Power analysis using the statistical power of a hypothesis test also called sample size analysis, is an important tool for deciding what sample size is required to guarantee a reasonable chance to reject a false null hypothesis. In the next two sub-sections, we are going use the power analysis under the context of random samples size to obtain first, the minimum dimensions n_i, $i = 1, ..., k$ that the samples must have and second, the minimum duration for the data collection in order to obtain that minimal dimensions.

3.1. Obtaining the global sample dimension

Suppose that we are working with F tests for fixed effect models for which, a null hypotheses H_0 and an alternative one H_a were previously formulated. Considering a study with k pathologies and that n represents the sum of sample dimensions for all pathologies, the power function of this tests is given by

$$Pow_\alpha(\delta) = Pr(\mathcal{F} > f_{1-\alpha,k-1,n-k}|H_0 \text{ false}) = 1 - \beta,$$

where $f_{1-\alpha,k-1,n-k}$ is the $(1-\alpha)$-th quantile of the central F distribution with $k-1$ and $n-k$ degrees of freedom. Since $Pr(\mathcal{F} > f_{1-\alpha,k-1,n-k}|H_0 \text{ false}) = 1 - F(f_{1-\alpha,k-1,n-k}|k-1, n-k, \delta)$ we have

$$F(f_{1-\alpha,k-1,n-k}|k-1, n-k, \delta) = \beta. \tag{4}$$

Since the power of the F text is a monotonically increasing function of the parameter δ, from equation (4) we can get the pairs of (n, δ) that allow

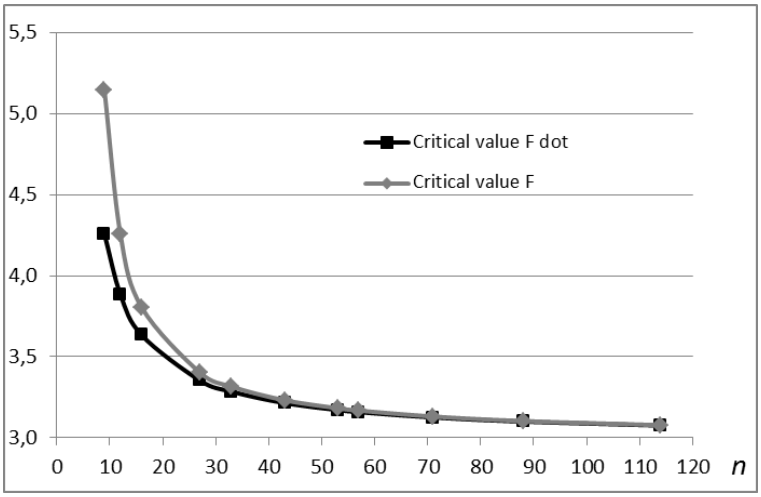

Figure 1: Critical valor for 5% probability level $\dot{F}$ versus F test

obtaining a predetermined power $1-\beta$ for the α-level F test with $k-1$ and $n-k$ degrees of freedom. From those pairs, we can choose a minimum global sample size n that guarantees, for all alternatives with non-centrality parameter not less than a given δ, a test power of at least $1-\beta$. We point out that the parameter δ measures the "distance" between the alternatives and the tested hypothesis. Thus, in an usual F test we can look for the minimum sample size n that, for a given δ, ensures a sufficient power for the test.

However, for obtaining the power of the F test with random samples size, we should use the $\dot{F}$ distribution and the $\dot{f}_{1-\alpha,k-1,n-k}$ quantil, that is

$$\dot{F}(\dot{f}_{1-\alpha,k-1,n-k}) = \sum_{\mathbf{n>0}} q(\mathbf{n}) F(\dot{f}_{1-\alpha,k-1,n-k} | k-1, n-k, \delta(\mathbf{n})). \qquad (5)$$

For the computation of $\dot{F}$ distribution, as described in section 2, we have to know the frequencies $\lambda_1, ..., \lambda_k$ and obtain the minimal sizes $n_1, ..., n_k$ of the samples that allow to truncate the series and compute with enough precision the distribution values. The sum of all n_i give us a minimum n and then we can obtain the quantil $\dot{f}_{1-\alpha,k-1,n-k}$. Moreover, in computing the non-central $\dot{F}$, since in eq. (5) the non-centrality parameter $\delta(\mathbf{n})$ has to be computed in each step of the sum, we need also to fixate the values of the sample means $\mu_1, ..., \mu_k$ and variance σ^2 based on previous knowledge. The power function for this specific $\dot{F}$ test is given by

$$Pow_\alpha(\delta(\mathbf{n})) = 1 - \dot{F}(\dot{f}_{1-\alpha,k-1,n-k}). \qquad (6)$$

If the obtained power is not enough, we should increase the $n_1, ..., n_k$ proportionally without changing the $\lambda_1, ..., \lambda_k$ and recalculate the power until we have obtained a reasonable value. A reasonable value for the power will be for instance 80%. This procedure may be repeated with different assumptions of sample means values $\mu_1, ..., \mu_k$ and σ^2, in order to obtain a final conclusion about the minimum n that guaranties sufficient power for the test. Lets call to these minimum dimensions $\dot{n}_1, ..., \dot{n}_k$ and to their sum $\dot{n}$.

In Table 3 we present the comparative results for the power of an $\dot{F}$ and F usual test with 3 pathologies ($k = 3$). These results were obtained assuming different values of $\lambda_1, \lambda_2, \lambda_3$ and with μ_1, μ_2, μ_3 and σ^2 constant. Notice that the n_1, n_2, n_3 obtained are the minimum that allow a good precision in the $\dot{F}$ distribution values. For all the cases presented in Table 3 the power of the $\dot{F}$ test is considerably higher than the power of the usual F test, this meaning that the F tests considering F distributions that account on the randomness of samples size have an higher probability of reject a false null hypothesis.

Table 3: Comparison between the power of $\dot{F}$ tests (F tests with random samples size) and the usual F tests; case for comparing 3 pathologies and $\mu_1 = 0.5$, $\mu_2 = 0.3$, $\mu_3 = 1.2$, $\sigma^2 = 1$.

								Power	
λ_1	λ_2	λ_3	n_1	n_2	n_3	n	δ	$\dot{F}$	usual F
0.05	0.05	0.05	3	3	3	9	1.34	42.14%	11.82%
0.1	0.1	0.1	4	4	4	12	1.79	53.07%	15.92%
0.1	0.3	0.5	4	5	7	16	2.68	69.06%	23.82%
1	1	1	9	9	9	27	4.02	83.19%	37.35%
2	2	1	12	12	9	33	4.43	85.77%	41.61%
4	3	2	17	14	12	43	5.71	91.36%	52.84%
5	4	4	19	17	17	53	7.65	95.71%	66.85%
8	8	6	25	25	21	71	9.96	97.92%	79.55%

Considering now, for instance, the second example in Table 3 ($\lambda_1 = \lambda_2 = \lambda_3 = 0.1$), the power obtained 53.07% for the minimum n_1, n_2, n_3 and those $\mu_1, \mu_2, \mu_3, \sigma^2$, is not sufficient high, thus we proportionally increase the n_1, n_2, n_3 without changing the $\lambda_1, \lambda_2, \lambda_3$ until obtaining a power near 80% as showed in the Table 4. This power is obtained for an $n = 24$. If we think the sample means and variance can change significantly in order to cause a decrease of the δ, we must also calculate the power for several different values of these parameters and obtain the n that allows a power of at least 80%.

Table 4: Power of an $\dot{F}$ test with $\lambda_1 = \lambda_2 = \lambda_3 = 0.1$ and $\mu_1 = 0.5, \mu_2 = 0.3, \mu_3 = 1.2, \sigma^2 = 1$ for increasing n_1, n_2, n_3 until obtaining 80% in comparison with the power of the the usual F Test

					Power	
n_1	n_2	n_3	n	δ	$\dot{F}$	usual F
4	4	4	12	1.79	53.1%	15.9%
5	5	5	15	2.23	61.9%	20.1%
7	7	7	21	3.13	75.3%	28.8%
8	8	8	**24**	3.57	**80.2**%	37.3%

3.2. Obtaining the minimum duration by sampling

In the last section, the minimum sample dimensions $\dot{n}_i, i = 1, ..., k$ and global dimension $\dot{n}$ to get an $\dot{F}$ test with power larger than or equal to $1 - \beta$ were obtained. Now, we may want to know how much time is necessary to wait in order to get this minimum number of observations.

Assuming the existence of k independent Poisson counting processes $N_i(t)$; $t \geq 0, i = 1, ..., k$, corresponding to the arrivals of the patients with the k pathologies, and reliable lower bounds $\dot{\lambda}_1, ..., \dot{\lambda}_k$ for the processes intensities, the probability of getting at least $\dot{n}_i$ number of patients for the i-th pathology, for all $i = 1, .., k$, is

$$\begin{aligned} pr(\mathbf{N}(t) \geq \dot{\mathbf{n}}) &= \prod_{i=1}^{k} pr(N_i(t) \geq \dot{n}_i) = \prod_{i=1}^{k} [1 - pr(N_i(t) < \dot{n}_i)] \\ &\geq \prod_{i=1}^{k} [1 - \sum_{j=1}^{\dot{n}_i - 1} e^{-\dot{\lambda}_i t} \frac{(\dot{\lambda}_i t)^j}{j!}]. \end{aligned} \tag{7}$$

since for the lower bounds $\dot{\lambda}_1, ..., \dot{\lambda}_k$ we have

$$pr(N_i(t) < \dot{n}_i) < \sum_{j=1}^{\dot{n}_i - 1} e^{-\dot{\lambda}_i t} \frac{(\dot{\lambda}_i t)^j}{j!},$$

for $i = 1, ..., k$.

Now, setting a probability p to have $\mathbf{N}(t) \geq \dot{\mathbf{n}}$, the time t can be chosen such that

$$\prod_{i=1}^{k} [1 - \sum_{j=1}^{\dot{n}_i - 1} e^{-\dot{\lambda}_i t} \frac{(\dot{\lambda}_i t)^j}{j!}] > p. \tag{8}$$

4. An example of application

In order to illustrate the usefulness of the method, in this section an example of application with non real data is presented.

Consider the people arriving at an hospital with pathologies 1, 2 and 3 during a given time span, from which a blood sample is taken to register a specific value. Suppose that is known from previous studies that the arrivals rates for the 3 pathologies are $\lambda_1 = 8.56$, $\lambda_2 = 0.64$ and $\lambda_3 = 0.34$ cases per day, in that hospital.

Before moving to data collection, we may want to answer to the following questions:

1. What are the minimum samples sizes required to have a controlled truncation error for the series distribution in eq. (2), therefore enough precision for the $\dot{F}$ distribution?
2. Are these minimum samples sizes sufficient to obtain a pre-fixed test power of 80%? Do we need more observations in order to get that?
3. How long is required to wait to obtain these minimum samples sizes?

In the next section the answers to those questions will be given. After that, we can proceed with the analysis of the collected data.

4.1. Planning the study

In order to answer to question 1, we use the inequality (3) to computed the minimum samples sizes for the 3 pathologies. The results are presented in Table 5 taking $\varepsilon = 10^{-6}$. If these minimum samples size are satisfied, we are

Table 5: Minimum samples sizes

pathology	1	2	3
λ_i	8.56	0.64	0.34
n_i min.	26	7	6

guaranteed that the truncation error will have an upper bound of 0.0001230, which give us enough precision on the critical values.

An F test to compare $k = 3$ different pathologies with $\alpha = 5\%$ of significance will have an $\mathcal{F}$ statistic with 2 and $n - 3$ degrees of freedom. So, we start by obtaining the power of the corresponding $\dot{F}$ test for the minimum samples sizes in Table 5 ($n = 39$). To this purpose, we used the known estimates for the sample means values: $\hat{\mu}_1 = 0.91, \hat{\mu}_1 = 1.5, \hat{\mu}_1 = 0.7$; and

variance: $\hat{\sigma}^2 = 0.4$. The power obtained, 92.9% for a $\delta = 6.23$, is high enough, thus we do not need to increase the $n_i, i = 1, 2, 3$. As a result, the minimum global sample size of 39 corresponding to the sum of the minimum samples sizes in Table 5, enables us to conduct an $\dot{F}$ test with a power much higher that 80%.

Now, let's consider a probability $p = 1 - \varepsilon$ of getting $n_1 \geq 26 \cap n_2 \geq 7 \cap n_3 \geq 6$, with $\varepsilon = 10^{-4}$. We computed the probability given by expression (8) for t increasing from 1 to 100 and stopping when $pr(n_1 \geq 0 \cap n_2 \geq 0 \cap n_3 \geq 0) > p$ and found out that for $t \simeq 58$ the value for that probability was 0.99991. In other words, the data collection period should be at least 58 days in order to obtain that number of observations per pathology.

4.2. Data analysis

A study was conducted with a collection period of 58 days and the 42 observations presented in Table 6 were collected.

Table 6: Observations

n_i	pathology 1	pathology 2	pathology 3
1	1.12	1.96	0.65
2	0.06	1.83	0.46
3	1.09	0,88	0.67
4	0.27	1.62	0.49
5	1.58	1.98	0.85
6	0.40	1.78	0.64
7	0.03	1.66	
8	0.36	1.80	
9	1.03	1.99	
10	1.86	0.35	
11	1.68		
12	0.45		
13	1.59		
14	0.81		
15	0.71		
16	0.44		
17	1.33		
18	0.70		
19	0.29		
20	1.26		
21	0.06		
22	1.30		
23	1.23		
24	0.74		
25	1.78		
26	1.71		
mean	0.9179	1.5833	0.6266

In Table 7 is presented the one way ANOVA used to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$.

The distribution $\dot{F}_0(z) = \sum_{\mathbf{n}>\mathbf{0}} q(\mathbf{n}) F(z|2, 39)$ of the $\mathcal{F}$ statistic was computed and the corresponding series was truncated, that is, the terms such $n_1 > 26 \wedge n_2 > 10 \wedge n_3 > 6$ were not considered. This condition guaranties that the truncation error is less than 0.0001230.

Table 7: One way ANOVA table

ANOVA	df	SS	$\mathcal{F}$ statistic	δ
Treatments	2	4.365	7.515	11.2
Error	39	11.326		
Total	41	15.691		

Using a numerical method the equation $\dot{F}_0(z) = 0.95$ was solved in order to z and obtained the quantile $\dot{f}_{(0.95;2;39)}$, which is the critical value of the 0.05 level test. Since the obtained critical value 3.220 is less than the $\mathcal{F} = 7.515$, the null hypothesis is rejected and we can conclude that the 3 pathologies are significantly different. Moreover, since the obtained samples led to an higher $\delta = 11.2$ (Table 7) and $n = 42$, the power for this test is 99.29% even higher then planed, in contrast with the 82.96% of the usual F test.

Had we considered the samples sizes n_1, n_2, n_3 as fixed numbers rather than as Poisson random variables, the critical value for this test with the usual central F distribution would be 3.238 instead of 3.220. As a result, when using F tests with random samples size, a slightly smaller value for the $\mathcal{F}$ statistic is required. However, the power of the test using the $\dot{F}$ distribution, i.e., considering random samples size, increase considerably comparatively with the power of the usual F test.

All the computations were performed using the EXCEL and programming in Microsoft Visual Basic for applications.

5. Conclusions

In comparison studies, when there is a specific time span to collect the observations and the sizes of the samples cannot be decided previously, it is appropriate to consider the sample dimensions as realizations of independent Poisson variables. The F tests developed from this assumption have the advantage of being more powerful, i.e., F tests considering F distributions that account on the randomness of samples size have an higher probability of reject a false null hypothesis. Moreover, they generate smaller critical values than the F usual distribution, the difference tending to zero as the number of observations increase. Furthermore, the approach is useful in the planning phase of a study to obtain the minimum duration for the data collection that ensures the chosen power and level of significance.

Acknowledgements

This work was partially supported by CMA/FCT/UNL, under the project PEst-OE/MAT/UI0297/2011.

References

Hocking, R., 2003. Methods and Applications of Linear Models. John Willey&Sons. New York.

Kendall, M. and Stuart, A., 1961. The Advanced Theory of Statistics. Vol.II. Charles Griffin & Co. Londres.

Montgomery, D.C., 1997. Design and Analysis of Experiments - 5th edition. John Willey & Sons. New York.

Mexia, J.T., Nunes, C., Ferreira, D., Ferreira, S., Moreira, E., 2011. Ortogonal fixed efects ANOVA with random sample sizes. WSEAS proceedings of the 5th International conference on Applied Mathematics, Simulation, Modelling, 84-90.

Mexia, J.T., Moreira, E.E., 2010. Randomized sample size F tests for the one-way layout. AIP Conference Proceedings, volume 1281- ICNAAM 2010-8th International Conference of Numerical Analysis and Applied Mathematics, 1248-1251,doi:10.1063/1.3497917.

Mexia, J.T., 1990. Best linear unbiased estimates, duality of F tests and Sheff multiple comparison method in presence of controlled heteroscedasticity. Computational statistics and data analisys 10(3), 271-281.

Nunes, C., Ferreira, D., Ferreira, S., Mexia, J.T., 2011. F-tests with a rare pathology. Journal of Applied Statistics. DOI:10.1080/02664763.2011.603293.

Scheffé, H., 1959. The Analysis of Variance. John Willey & Sons. New York.