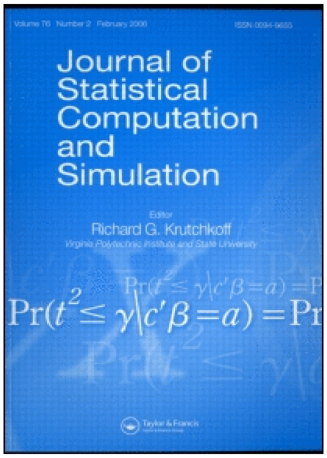

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713650378

On the size of the **F**-test for the one-way random model with heterogeneous error variances

To cite this Article: Lee, Juneyoung, Khuri, André I., Kim, Kee Whan and Lee, Sangkon , 'On the size of the **F**-test for the one-way random model with heterogeneous error variances', Journal of Statistical Computation and Simulation, 77:6, 443 - 455
To link to this article: DOI: 10.1080/10629360600569105
URL: http://dx.doi.org/10.1080/10629360600569105

Journal of Statistical Computation and Simulation
Vol. 77, No. 6, June 2007, 443–455

On the size of the F-test for the one-way random model with heterogeneous error variances

JUNEYOUNG LEE*†, ANDRÉ I. KHURI‡, KEE WHAN KIM§ and SANGKON LEE¶

†Department of Biostatistics, Korea University, Seoul 136-705, Korea
‡Department of Statistics, University of Florida, Gainesville, FL 32611-8545, USA
§Department of Informational Statistics, Korea University, Chungnam 339-700, Korea
¶Department of Computer Science, Jeonju University, Chonbuk 560-759, Korea

(Revised 9 February 2004; in final form 3 June 2005)

Traditional analysis of variance tests are based on the assumption of homogeneous error variances, which often fails in real experimental situations. Violation of this assumption affects not only the power of the standard F-test, but also its size. When a design is unbalanced, the effect of unequal error variances is even more complex. In this paper, we study the effect of heterogeneous error variances on the size of the F-test concerning the among-group variance component in an unbalanced random one-way model. We also provide a method for computing the true critical value of the F-test for a given level of significance.

Keywords: Among-group variance component; Davies' algorithm; Level of significance; Measure of imbalance; Unbalanced design

1. Introduction

In many experimental situations, the error variances in a given model can be heterogeneous for a variety of reasons. This can seriously affect the associated analysis of variance (ANOVA) tests. In particular, the power and size of the ANOVA F-test for the treatment effects in a fixed-effects one-way model can be affected. The assumption of homogeneous error variances should therefore be checked before applying the F-test. A simple way to check this assumption is the use of graphs, such as the plots of residuals against predicted values. Departures from the assumption would be indicated by a pattern of greater dispersion in the residuals on one side of the graph than the other. Several parametric tests are also available [see, for example, refs. 1–4]. Hartley's [1] so-called $F_{\max}$-test requires equally replicated data. When the sample sizes are unequal, or the variance estimates are based on different numbers of degrees of freedom, a procedure developed by Bartlett [2] is generally recommended. Both procedures require the variance estimators to be independent. Furthermore, both tests are extremely sensitive to departures from normality and should not be used with non-normally

*Corresponding author. Email: jyleeuf@korea.ac.kr

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10629360600569105

distributed data. A more robust procedure to non-normal distributions was given by Levene [3] who suggested using the standard one-way ANOVA on the absolute values of the residuals to test the hypothesis of equal variances. Conover *et al.* [4] recommended using the median rather than the sample mean of each treatment in Levene's test statistic.

The analysis of the fixed-effects one-way model under heterogeneous error variances has for many years been given a great deal of attention by researchers. Welch [5], James [6], and Box [7] were among the first to address the heterogeneity problem. It has long been established that the standard F-test for testing equality of the treatment means is sensitive to the assumption of heterogeneous error variances, especially if the cell sample sizes are unequal (see, for example, ref. [8, p. 351] and refs. [9–11]). Several test procedures have been proposed to solve this problem. Brown and Forsythe [9] used Monte Carlo simulation to compare the size and power of four test statistics, including the standard F-test statistic. Bishop and Dudewicz [10] proposed a two-stage testing procedure that was independent of the unknown error variances. Draper and Guttman [12] used a Bayesian approach in situations involving several treatment groups, but only two different group variances were suspected. Several other approximate tests were introduced by Wilcox [13], Alexander and Govern [14], and more recently, by Smith and Peddada [15].

Handling unequal error variances in random or mixed models has also been studied by some authors. Rao *et al.* [16] discussed point estimation of the parameters of the one-way random model with unequal error variances. Using the same model, Jeyaratnam and Othman [17] suggested an approximate test for the among-group variance component. Singh [18] investigated the effect of heterogeneity of variances and data imbalance on the power of the standard F-test for the among-group variance component. The corresponding power values were computed using a doubly infinite series of incomplete beta functions. Singh [18] concluded that heteroscedasticity increases the power of the test in situations where the larger variances are associated with the smaller group sizes. If larger variances are associated with the larger group sizes, the power is decreased. A small increase in the power, due to heteroscedasticity, was noted with balanced data. Recently, Singh [19] considered an estimation of heritability with heterogeneous error variances in the one-way random model.

As was noted earlier, because the heteroscedasticity of the error variances affects the size of the test as well as its power, its impact on the test size should be investigated. In this study, we investigate the effect of heterogeneous error variances on the size of the standard F-test for the among-group variance component in the one-way random model. We consider several designs having different degrees of imbalance and different sizes of error variances. The level of significance of the test is then calculated for a given design and a given pattern of error variances. Moreover, a method for achieving a nominal size for the test is proposed by calculating the true critical value of the test. Using this value, one can proceed with a test which provides a desired level of significance concerning the among-group variance component.

2. Level of significance of the F-test

Consider the unbalanced one-way random model,

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, n_i, \tag{1}$$

where α_i and ε_{ij} are independently distributed as normal variates with zero means and variances $\sigma_\alpha^2, \sigma_i^2$ $(i = 1, 2, \ldots, k)$, respectively. Note that the error variances are equal within groups, that is, for fixed i, but are unequal among groups, that is, for different values of i, $i = 1, 2, \ldots, k$.

Model (1) can be written as

$$\boldsymbol{y} = \mu \mathbf{1}_N + \left[\bigoplus_{i=1}^{k} \mathbf{1}_{n_i}\right]\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{2}$$

where $N = \sum_{i=1}^{k} n_i$, $\mathbf{1}_{n_i}$ is a column vector of n_i ones $(i = 1, 2, \ldots, k)$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)'$, $\oplus$ is the symbol of direct sum of matrices, that is, $\oplus_{i=1}^{k}\mathbf{1}_{n_i} = \operatorname{diag}(\mathbf{1}_{n_1}, \mathbf{1}_{n_2}, \ldots, \mathbf{1}_{n_k})$, and $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are the vectors of observations and random errors, respectively. Hence, $\boldsymbol{y}$ is distributed as $\boldsymbol{y} \sim N(\mu\mathbf{1}_N, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the variance–covariance matrix of $\boldsymbol{y}$, which can be expressed as

$$\boldsymbol{\Sigma} = \sigma_\alpha^2 \bigoplus_{i=1}^{k} \mathbf{J}_{n_i} + \bigoplus_{i=1}^{k} (\sigma_i^2\, \mathbf{I}_{n_i}), \tag{3}$$

where $\mathbf{I}_{n_i}$ is the identity matrix and $\mathbf{J}_{n_i}$ is the matrix of ones of order $n_i \times n_i$. For testing the significance of σ_α^2, the standard F-test statistic is given by

$$F = \left(\frac{N-k}{k-1}\right)\frac{\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}}, \tag{4}$$

where $\boldsymbol{Q}$ and $\boldsymbol{R}$ are defined as

$$\boldsymbol{Q} = \bigoplus_{i=1}^{k}\left(\frac{1}{n_i}\mathbf{J}_{n_i}\right) - \frac{1}{N}\mathbf{J}_N \tag{5}$$

and

$$\boldsymbol{R} = \mathbf{I}_N - \bigoplus_{i=1}^{k}\left(\frac{1}{n_i}\mathbf{J}_{n_i}\right), \tag{6}$$

respectively. This F-statistic has, under the null hypothesis H_0: $\sigma_\alpha^2 = 0$, the F-distribution with $k-1$ and $N-k$ degrees of freedom when the error variances are homogeneous, that is, when the σ_i^2's are equal. It is also known that this test is uniformly most powerful invariant if the associated data set is balanced. However, when the error variances are heterogeneous, as in model (1), neither $\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}$ nor $\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}$ are distributed as scalar multiples of chi-squared variates and are not independent. This follows from the fact that neither the matrix $\boldsymbol{Q}\boldsymbol{\Sigma}$ nor $\boldsymbol{R}\boldsymbol{\Sigma}$ are scalar multiples of idempotent matrices. Moreover, the $\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{R}$ matrix is not equal to $\mathbf{0}$ (see ref. [20, Theorems S2 and S3 in Appendix S5, p. 467). Consequently, the F-statistic does not follow an exact F-distribution. There is no exact procedure for testing the significance of σ_α^2 in this case [17].

2.1 *Level of significance of the standard F-test*

The level of significance of the test statistic in equation (4) is given by

$$\alpha_f = P[F \geq F_{\alpha,k-1,N-k} | \sigma_\alpha^2 = 0], \tag{7}$$

where $F_{\alpha,k-1,N-k}$ denotes the upper α-quantile of the F-distribution with $k-1$ and $N-k$ degrees of freedom. Under H_0: $\sigma_\alpha^2 = 0$, however, the matrices, $\boldsymbol{Q}\boldsymbol{\Sigma}_0$ and $\boldsymbol{R}\boldsymbol{\Sigma}_0$, where $\boldsymbol{\Sigma}_0 = \oplus_{i=1}^{k}\sigma_i^2\mathbf{I}_{n_i}$, are not scalar multiples of idempotent matrices, nor $\boldsymbol{Q}\boldsymbol{\Sigma}_0\boldsymbol{R} = \mathbf{0}$. Consequently, the F-statistic does not follow the F-distribution. As a result, $F_{\alpha,k-1,N-k}$ is not the true α-critical value of the test statistic. The purpose of using formula (7) is to determine the extent of the difference between α_f and the nominal level of significance, α.

The value of α_f changes not only with the error variances, but is also affected by the chosen design. Although the exact distribution of the F-statistic in equation (4) under the null hypothesis is unknown, we can obtain the exact value of α_f using the following theorem.

THEOREM 2.1 *Let $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\mathbf{A}$ be a symmetric matrix of order $p \times p$. Then $\boldsymbol{x}'\mathbf{A}\boldsymbol{x}$ can be expressed as a linear combination of independent central chi-squared variates of the form*

$$\boldsymbol{x}'\mathbf{A}\boldsymbol{x} \overset{D}{=} \sum_{i=1}^{r} \lambda_i\, \chi^2_{\nu_i}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the distinct non-zero eigenvalues of $\mathbf{A}\boldsymbol{\Sigma}$ with multiplicities $\nu_1, \nu_2, \ldots, \nu_r$, respectively.

Proof See ref. [21, chapter 29, pp. 150–153]. ■

COROLLARY 2.1 *The result of Theorem 2.1 remains true if $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ provided that $\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = 0$ and $\mathbf{A}$ is positive semidefinite.*

Proof Assuming $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, let $\boldsymbol{y} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$. Then, $\boldsymbol{y} \sim N(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}, \mathbf{I}_N)$. By the spectral decomposition theorem, we have

$$\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \boldsymbol{y}'\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'\boldsymbol{y},$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\delta_1, \delta_2, \ldots, \delta_N)$ with δ_i being the ith eigenvalue of $\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}$, and $\mathbf{P}$ is an orthogonal matrix whose columns are orthonormal eigenvectors of $\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}$ corresponding to the δ_i's. By letting $\boldsymbol{w} = \mathbf{P}'\boldsymbol{y}$, $\boldsymbol{w} \sim N(\mathbf{P}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}, \mathbf{I})$, we obtain

$$\boldsymbol{x}'\mathbf{A}\boldsymbol{x} = \boldsymbol{w}'\boldsymbol{\Lambda}\boldsymbol{w} = \sum_{i=1}^{r} \lambda_i\, w_i^2, \tag{8}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the distinct non-zero eigenvalues of $\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}$, and hence of $\mathbf{A}\boldsymbol{\Sigma}$, with multiplicities $\nu_1, \nu_2, \ldots, \nu_r$, respectively. The w_i^2's are distributed as $\chi'^2_{\nu_i}(\theta_i)$ with non-centrality parameters $\theta_i = (1/2)(\mathbf{P}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu})'(\mathbf{P}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu})$, $i = 1, 2, \ldots, r$. Here, $\mathbf{P}_i$ is a matrix of order $N \times \nu_i$ whose columns are eigenvectors corresponding to λ_i. Thus, $\boldsymbol{x}'\mathbf{A}\boldsymbol{x} \overset{D}{=} \sum_{i=1}^{r} \lambda_i\, \chi^2_{\nu_i}$ if and only if $\theta_i = 0$ for all $i = 1, 2, \ldots, r$, that is, $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1/2}\mathbf{P}_i\mathbf{P}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = 0$ for all $i = 1, 2, \ldots, r$, or equivalently, if and only if $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1/2}\,(\sum_{i=1}^{r} \lambda_i\mathbf{P}_i\mathbf{P}_i')\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = 0$ (the sufficiency part is due to $\lambda_i > 0$, which is true if $\mathbf{A}$ is positive semidefinite), that is, $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1/2}\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = 0$. This is equivalent to $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = 0$. Hence, the result follows. ■

THEOREM 2.2 *For the unbalanced one-way random model* (1), *the size of the F-test in equation* (7) *can be expressed as*

$$\alpha_f = P\left[\sum_{i=1}^{s} \tau_i\,\chi^2_{\kappa_i} \geq 0\right],$$

where $\tau_1, \tau_2, \ldots, \tau_s$ are the distinct non-zero eigenvalues of $(\boldsymbol{Q} - (k-1)/(N-k)\, F_{\alpha,k-1,N-k}\boldsymbol{R})\boldsymbol{\Sigma}_0$ with multiplicities $\kappa_1, \kappa_2, \ldots, \kappa_s$, respectively, which satisfy the relationship $\sum_{i=1}^{s} \kappa_i = \mathrm{rank}\,(\boldsymbol{Q} - (k-1)/(N-k)\,F_{\alpha,k-1,N-k}\boldsymbol{R})$. Here, the $\chi^2_{\kappa_i}$ represent independent χ^2-distributions with κ_i degree of freedom, $i = 1, 2, \ldots, s$.

Proof Using equations (3) and (4) in (7), the size of the F-test is given by

$$\begin{aligned}\alpha_f &= P\left[\left(\frac{N-k}{k-1}\right)\frac{\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}} \geq F_{\alpha,k-1,N-k} \mid \sigma_\alpha^2 = 0\right] \\ &= P\left[\boldsymbol{y}'\left(\boldsymbol{Q} - \frac{k-1}{N-k}F_{\alpha,k-1,N-k}\boldsymbol{R}\right)\boldsymbol{y} \geq 0 \mid \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \oplus_{i=1}^{k}\sigma_i^2\mathbf{I}_{n_i}\right]. \end{aligned} \tag{9}$$

Let us apply the representation (8) to $\boldsymbol{y}'\mathbf{A}\boldsymbol{y}$, where $\mathbf{A} = \boldsymbol{Q} - (k-1)/(N-k)F_{\alpha,k-1,N-k}\boldsymbol{R}$, and $E(\boldsymbol{y}) = \mu\mathbf{1}_N$. In this case, the non-centrality parameter θ_i associated with w_i^2 must be zero for all i. This is true because $\boldsymbol{\Sigma}^{-1/2}\mathbf{1}_N$ is an eigenvector of $\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}$ with eigenvalue zero by the fact that $\boldsymbol{Q}\mathbf{1}_N = 0$ and $\boldsymbol{R}\mathbf{1}_N = 0$. Hence, $\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{-1/2}\mathbf{1}_N) = \boldsymbol{\Sigma}^{1/2}\mathbf{A}\mathbf{1}_N = \mathbf{0}$. Consequently, $\boldsymbol{\Sigma}^{-1/2}\mathbf{1}_N$ must be orthogonal to the columns of $\mathbf{P}_i$, which are the eigenvectors of $\boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}$ corresponding to the non-zero eigenvalue $\tau_i (i = 1, 2, \ldots, s)$. It follows that, for $i = 1, 2, \ldots, s$, $\mathbf{P}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = \mathbf{P}_i'\boldsymbol{\Sigma}^{-1/2}(\mu\mathbf{1}_N) = \mathbf{0}$. Hence, $w_i^2 \sim \chi^2_{\kappa_i}, i = 1, 2, \ldots, s$. ■

Davies [22] developed an algorithm for calculating probabilities associated with the distribution of a linear combination of independent χ^2 random variables. The algorithm itself (identified as AS155) is accessible via STATLIB, an E-mail and file-transfer protocol-based retrieval system for statistical software. Thus, when a design $D = \{n_1, n_2, \ldots, n_k\}$ is given and the error variances σ_i^2, $i = 1, 2, \ldots, k$, are known, we can obtain values of α_f using Davies' algorithm [22] on the basis of Theorem 2.2.

2.2 *Method for computing the true critical value of the F-test*

From equation (9), it can be seen that, for a given design $D = \{n_1, n_2, \ldots, n_k\}$ and specified values of σ_i^2, $i = 1, 2, \ldots, k$, the level of significance α_f depends on $F_{\alpha,k-1,N-k}$, the upper α-quantile of the F-distribution. Since the error variances may be heterogeneous, the F-statistic in equation (4) no longer has the F-distribution under the null hypothesis H_0: $\sigma_\alpha^2 = 0$. Thus, to obtain the true level of significance, a value other than $F_{\alpha,k-1,N-k}$ should be used. Let F_α^{adj} be such a value, that is,

$$P(F \geq F_\alpha^{\mathrm{adj}}) = \alpha. \tag{10}$$

This can be written as in equation (9) in the form

$$P\left[\sum_{i=1}^{r}\lambda_i^*\,\chi^2_{\nu_i^*} \geq 0\right] = \alpha \tag{11}$$

where $\lambda_1^*, \lambda_2^*, \ldots, \lambda_r^*$ are the distinct non-zero eigenvalues of $(\boldsymbol{Q} - (k-1)/(N-k)F_\alpha^{\mathrm{adj}}\boldsymbol{R})\boldsymbol{\Sigma}_0$ with multiplicities $\nu_1^*, \nu_2^*, \ldots, \nu_r^*$, respectively, which satisfy the relationship $\sum_{i=1}^{r}\nu_i^* = \mathrm{rank}(\boldsymbol{Q} - (k-1)/(N-k)F_\alpha^{\mathrm{adj}}\boldsymbol{R})$ and $\boldsymbol{\Sigma}_0 = \oplus_{i=1}^{k}\sigma_i^2\,\mathbf{I}_{n_i}$.

The true critical value of the F-test, namely F_α^{adj}, that satisfies equation (11) for a given value of α is obtained in the following manner: Davies' algorithm [22] is utilized to compute the value of α_f in equation (9) using the $F_{\alpha,k-1,N-k}$ value when the design and the error variances are given as was stated in section 2.1. If the ensuing value of α_f value is larger, or smaller, than the nominal level of significance, α, we increase, or decrease, the $F_{\alpha,k-1,N-k}$ value until the nominal value of α is achieved, which is also done by using Davies' algorithm [22]. This process can be repeated several times until the difference between α_f and α is < 0.00001, which results in the value of F_α^{adj} as in equation (10).

3. Simulation

3.1 *Measuring design imbalance and generating designs having a specified degree of imbalance*

To evaluate the impact of heterogeneous error variances on the size of the standard F-test concerning σ_α^2 for model (1), it is necessary to consider designs having different degrees of imbalance.

Measures of data imbalance were considered by Ahrens and Pincus [23], Khuri [24], and Lera Marques [25]. One of Ahrens and Pincus' [23] measures for the one-way model is given by

$$\phi = \frac{\left(\sum_{i=1}^{k} n_i\right)^2}{k \sum_{i=1}^{k} n_i^2}, \tag{12}$$

where ϕ satisfies $(1/k) < \phi \leq 1$. Large values of ϕ indicate 'near balance', whereas values close to the lower bound indicate severe imbalance. This measure can be utilized to assess the efficiency of an unbalanced design as compared to a balanced design with the same number of observations.

To study the effect of imbalance on data analysis, a method for generating designs having a specified degree of imbalance is needed. Three such methods were introduced in the statistical literature [26–28]. In this study, Khuri's method [28] for constructing designs for the one-way

Table 1. Generated designs, $D = \{n_1, n_2, \ldots, n_k\}$, for the one-way model for each combination of levels of the number of groups (k), the total number of observations (N), and the degree of imbalance (ϕ).

k	N	ϕ	n_1	n_2	n_3	n_4	n_5	n_6	n_7	ϕ_a
3	50	0.5	1	9	40	–	–	–	–	0.495
3	50	0.5	4	40	6	–	–	–	–	0.504
3	50	0.5	40	8	2	–	–	–	–	0.500
	...					...				...
3	100	0.7	56	4	40	–	–	–	–	0.701
3	100	0.7	6	35	59	–	–	–	–	0.703
	...					...				...
3	250	0.9	44	103	103	–	–	–	–	0.900
3	250	0.9	122	57	71	–	–	–	–	0.899
5	50	0.7	14	21	6	4	5	–	–	0.700
5	50	0.7	1	7	9	21	12	–	–	0.698
5	50	0.7	3	14	10	3	20	–	–	0.700
	...					...				...
5	100	0.9	26	12	14	19	29	–	–	0.902
5	100	0.9	15	15	26	14	30	–	–	0.900
	...					...				...
5	250	0.5	7	86	12	14	131	–	–	0.501
5	250	0.5	73	19	16	4	138	–	–	0.500
7	50	0.9	7	8	3	6	9	11	6	0.902
7	50	0.9	8	6	9	2	9	9	7	0.902
7	50	0.9	8	5	8	8	3	7	11	0.902
	...					...				...
7	100	0.5	10	3	15	4	9	48	11	0.500
7	100	0.5	6	3	45	24	8	12	2	0.500
	...					...				...
7	250	0.7	37	8	31	76	61	9	28	0.700
7	250	0.7	12	13	52	42	76	46	9	0.700

Note: A value of ϕ_a denotes the actual value of ϕ for a ϕ-generated design.

model is used. Given that the n_i's must be integer-valued, it may not be possible to find a design that satisfies (12) for a specified value of ϕ. Khuri's method [28], however, guarantees to find a design whose degree of imbalance is closest to ϕ. It should also be noted that for specified values of k, N, and ϕ, there can be several designs that satisfy equation (12) [see ref. 28 for more details].

Table 2. The calculated size of the standard F-test, α_f, and the F_α^{adj} values in equation (10) for $\alpha = 0.05$.

(a) $k = 3$[†]

N	ϕ	σ_1^2	σ_2^2	σ_3^2	n_1	n_2	n_3	$\alpha_f^{\ddagger}$	F_α^{adj}
50	0.5	0.1	0.1	1	1	9	40	0.0001	0.7666
50	0.5	1	0.1	0.1	1	9	40	0.4727	20.4685
50	0.5	10	0.1	0.1	1	9	40	0.8184	198.9138
			...			...		...	...
100	0.7	10	0.1	0.1	56	4	40	0.0068	1.5748
100	0.7	0.1	10	0.1	56	4	40	0.6370	68.2543
100	0.7	10	0.1	10	56	4	40	0.0167	2.0611
			...			...		...	...
250	0.9	1	0.1	0.1	114	47	89	0.0260	2.3661
250	0.9	10	0.1	0.1	114	47	89	0.0258	2.3361
250	0.9	0.1	10	1	114	47	89	0.2117	7.3781

(b) $k = 5$[§]

N	ϕ	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	n_1	n_2	n_3	n_4	n_5	$\alpha_f^{\P}$	F_α^{adj}
50	0.7	10	10	0.1	0.1	0.1	14	21	6	4	5	0.0119	1.5719
50	0.7	10	0.1	0.1	1	0.1	14	21	6	4	5	0.0630	2.8932
50	0.7	0.1	0.1	10	0.1	1	14	21	6	4	5	0.2926	10.0104
				...					...			...	...
100	0.9	0.1	0.1	0.1	1	1	26	12	14	19	29	0.0500	2.4675
100	0.9	10	1	0.1	0.1	10	26	12	14	19	29	0.0334	2.1401
100	0.9	0.1	10	0.1	0.1	0.1	26	12	14	19	29	0.2655	8.3342
				...					...			...	...
250	0.5	10	0.1	0.1	0.1	0.1	145	18	32	51	4	0.0004	0.7158
250	0.5	0.1	0.1	0.1	0.1	10	145	18	32	51	4	0.6790	52.6429
250	0.5	0.1	10	0.1	1	0.1	145	18	32	51	4	0.3555	10.2210

(c) $k = 7$[‖]

N	ϕ	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	σ_6^2	σ_7^2	n_1	n_2	n_3	n_4	n_5	n_6	n_7	$\alpha_f^{\#}$	F_α^{adj}
50	0.9	10	10	1	10	10	10	0.1	7	8	3	6	9	11	6	0.0381	2.1535
50	0.9	0.1	0.1	10	1	0.1	0.1	1	7	8	3	6	9	11	6	0.4057	12.2677
50	0.9	0.1	0.1	10	0.1	0.1	0.1	0.1	7	8	3	6	9	11	6	0.4477	22.0451
					...							...				...	...
100	0.5	0.1	0.1	0.1	0.1	0.1	10	0.1	10	3	15	4	9	48	11	0.0008	0.7113
100	0.5	0.1	10	0.1	0.1	0.1	0.1	1	10	3	15	4	9	48	11	0.5561	20.7700
100	0.5	0.1	10	0.1	0.1	0.1	0.1	0.1	10	3	15	4	9	48	11	0.5772	31.0727
					...							...				...	...
250	0.7	1	0.1	10	1	1	0.1	10	25	37	64	25	20	4	75	0.0114	1.3930
250	0.7	10	10	10	0.1	0.1	0.1	1	25	37	64	25	20	4	75	0.0515	2.1554
250	0.7	0.1	0.1	0.1	0.1	0.1	10	0.1	25	37	64	25	20	4	75	0.6412	35.2063

Note: These values are obtained by using Davies' algorithm [22].
[†]A total of 1080 different settings are evaluated based on 45 generated designs.
[‡]The corresponding $F_{0.05,k-1,N-k}$ values are 3.1951 for $N = 50$, 3.0902 for $N = 100$, and 3.0324 for $N = 250$.
[§]A total of 10,800 different settings are evaluated based on 45 generated designs.
[¶]The corresponding $F_{0.05,k-1,N-k}$ values are 2.5787 for $N = 50$, 2.4675 for $N = 100$, and 2.4085 for $N = 250$.
[‖]A total of 98,280 different settings are evaluated based on 45 generated designs.
[#]The corresponding $F_{0.05,k-1,N-k}$ values are 2.3185 for $N = 50$, 2.1977 for $N = 100$, and 2.1360 for $N = 250$.

3.2 *Effects of heterogeneous error variances and design imbalance on the size of the F-test*

Once unbalanced designs are generated, by calculating the values of α_f and F_α^{adj}, the true critical value, using formulas (9) and (11), respectively, the effects of the error variances and the degree of design imbalance on the test size can be evaluated. In this study, we consider a factorial combination of the levels of k, N, ϕ and σ_i^2's, namely, $k = 3, 5, 7$; $N = 50, 100, 250$; $\phi = 0.5, 0.7, 0.9$; and $\sigma_i^2 = 0.1, 1, 10, i = 1, 2, \ldots, k$, excluding the three cases $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = 0.1, 1$, or 10. The region of interest R for this study is therefore given by

$$R = \{(k, N, \phi, \sigma_i^2) \mid 3 \le k \le 7, 50 \le N \le 250, 0.5 \le \phi \le 0.9,$$
$$0.1 \le \sigma_i^2 \le 10, \quad i = 1, 2, \ldots, k\}. \tag{13}$$

For each combination of the levels of k, N, and ϕ values, five different unbalanced designs were generated using Khuri's method [28]. The numbers of designs so generated

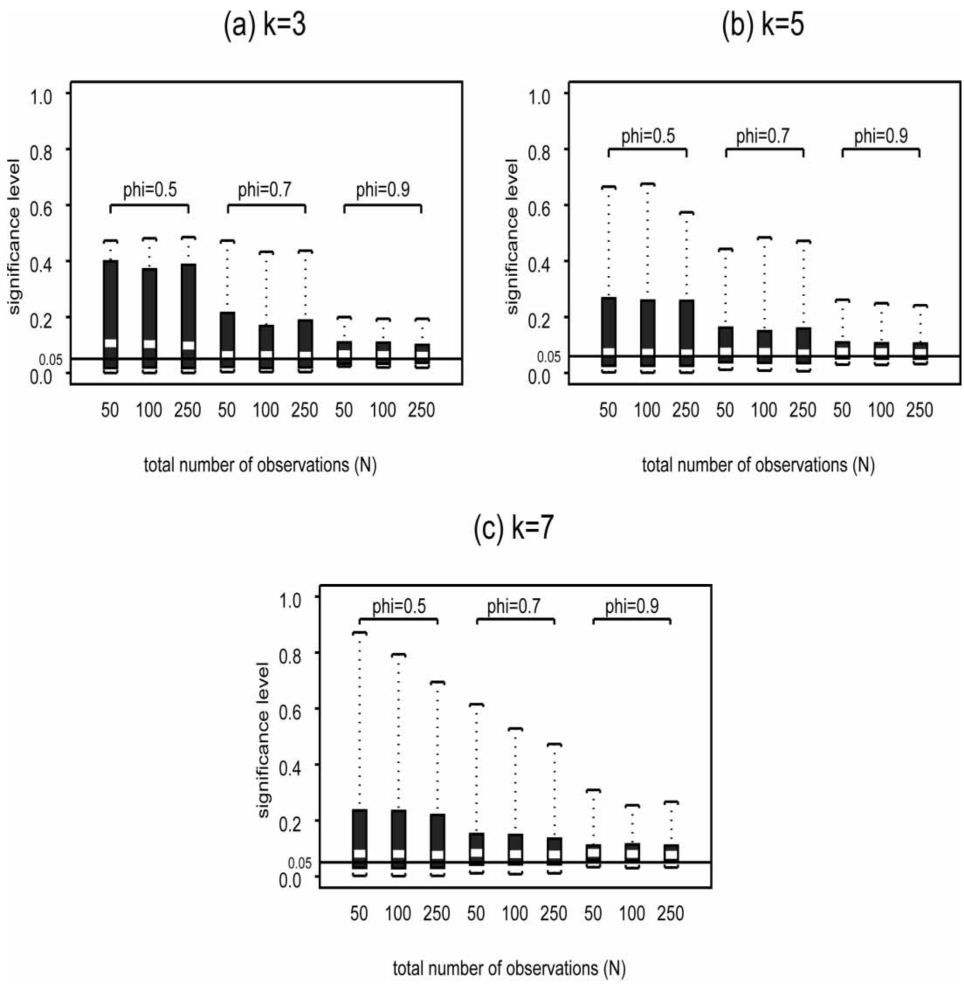

Figure 1. Box plots of calculated levels of significance, α_f, of the F-test when the maximum error variance is 10 times as large as the minimum.

is 45(= 3[for N] × 3[for ϕ] × 5[for replicates]) for each value of $k = 3$, 5, and 7, respectively, giving a total of 135. Some of the generated designs are listed in table 1 along with the actual value of ϕ, namely ϕ_a, for an (k, N, ϕ)-generated design. The value of ϕ_a is obtained from equation (12) using the n_i's from the generated design. For each of the generated designs, we consider a factorial combination of the levels of unequal $\sigma_i^2 (i = 1, 2, \ldots, k)$ values. This results in 1080 (= 45[number of designs] × $(3^3 - 3)$[number of unequal combinations of σ_i^2's]) combinations for $k = 3$; 10,800 (= 45 × $(3^5 - 3)$) for $k = 5$; and 98,280 (= 45 × $(3^7 - 3)$) for $k = 7$.

Let us consider a 5% level of significance in this simulation study. For each of the aforementioned combinations, values of α_f and F_α^{adj} are given in table 2. Note that α_f ranges from 0.0001 to 0.8244 for $k = 3$; from 0.0003 to 0.9610 for $k = 5$; and from 0.0008 to 0.9981 for $k = 7$. The corresponding F_α^{adj} values range from 0.4953 to 198.9183 for $k = 3$; from 0.7044 to 157.6412 for $k = 5$; and from 0.7113 to 169.6310 for $k = 7$. Box plots of α_f are shown in figures 1 and 2. Figure 1 corresponds to the case where the maximum error variance is 10 times as large as the minimum error variance, whereas figure 2 corresponds to when it is 100 times as large. A reference line of 0.05 was drawn on the box plots. Similar box plots were obtained for values of F_α^{adj} as shown in figures 3 and 4.

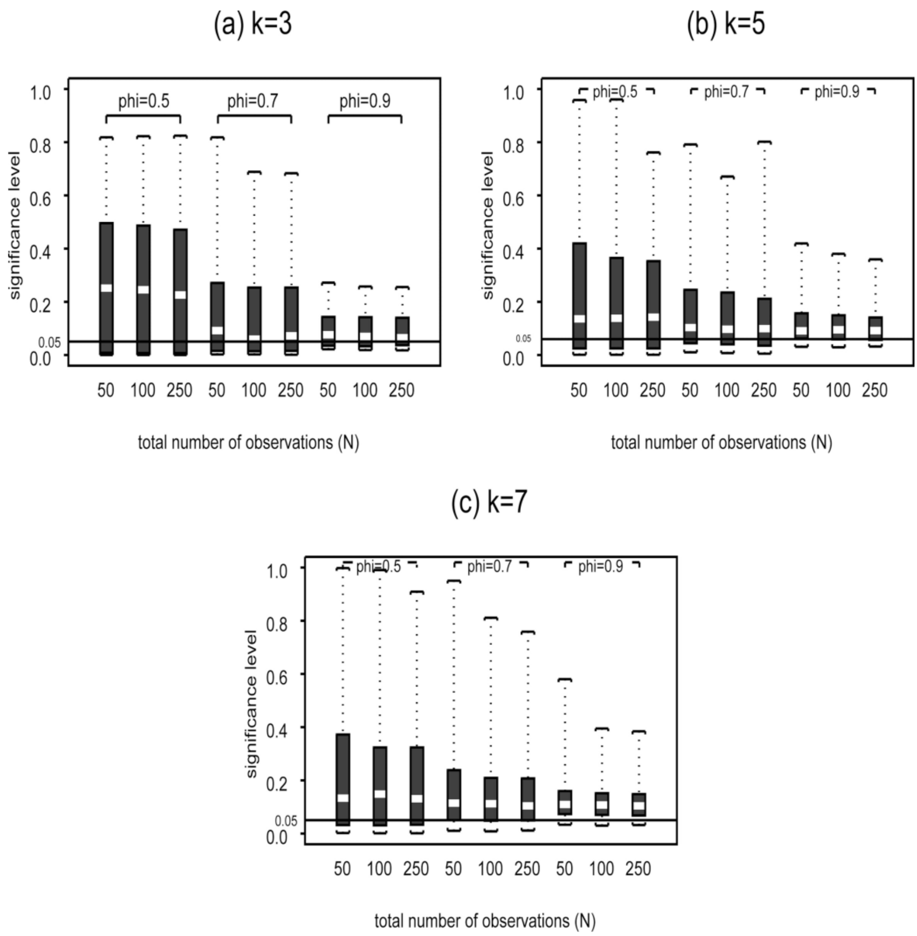

Figure 2. Box plots of calculated levels of significance, α_f, of the F-test when the maximum error variance is 100 times as large as the minimum.

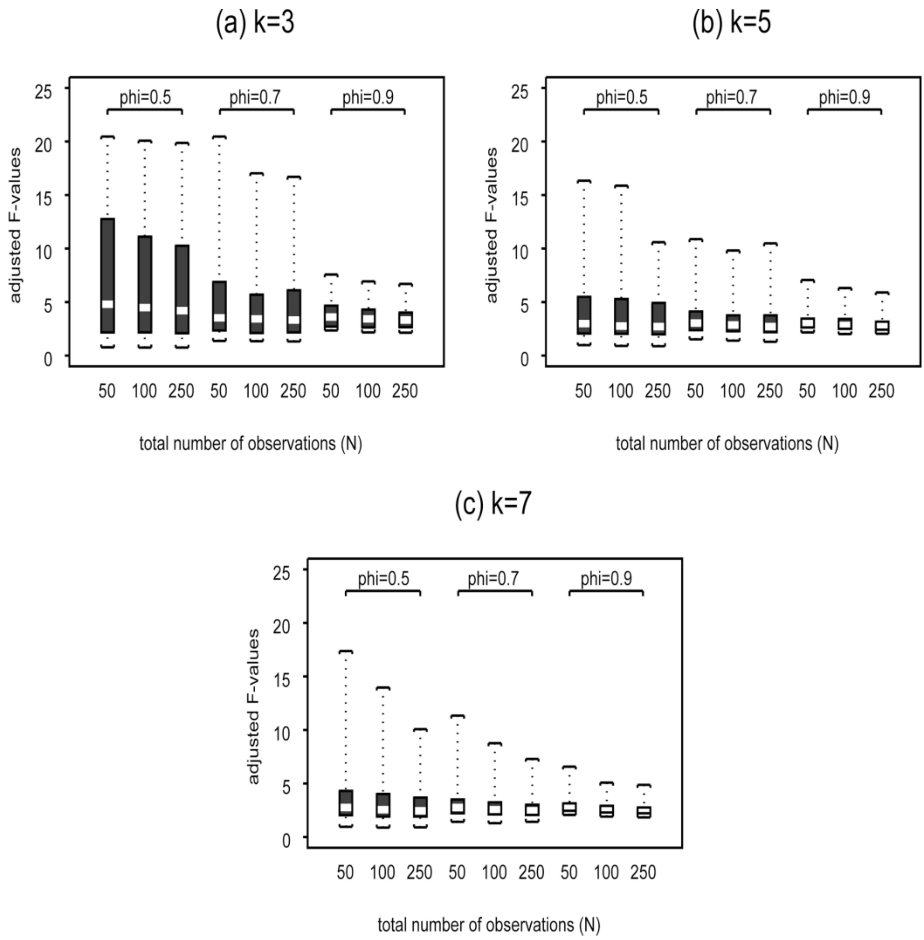

Figure 3. Box plots of the true α-critical values of the F-test, $F_{\alpha}^{\mathrm{adj}}$, for $\alpha = 0.05$ when the maximum error variance is 10 times as large as the minimum.

From figures 1 and 2, we can see that whenever the error variances are heterogeneous, the computed level of significance, α_f, tends to be, in most cases, larger than 0.05. This pattern is more apparent when the design is more unbalanced, regardless of the values of k or N. The interquartile range does not change much when the ratio of the maximum to the minimum error variance changes from 10 to 100, although there is a tendency to have more extreme values observed as the ratio increases. The total number of observations does not seem to have much effect on the level of significance for given values of k and ϕ. The effect of design imbalance on the significance level is also of interest. Both moderately ($\phi = 0.7$) and severely ($\phi = 0.5$) unbalanced designs produce values of α_f larger than those for mildly ($\phi = 0.9$) unbalanced designs. The interquartile range of α_f gradually increases as the degree of design imbalance increases. Overall, by looking at the interquartile ranges, we can see that design imbalance has a larger effect on α_f than does heterogeneity of the error variances.

From figures 3 and 4, it can be seen that the true α-critical values of the F-test, $F_{\alpha}^{\mathrm{adj}}$, for $\alpha = 0.05$ range almost from 0 to 20, when the maximum error variance is 10 times as large as the minimum, and from 0 to 200, when it is 100 times as large. By comparison, the table F-value, $F_{0.05,k-1,N-k}$, under the assumption of homogeneous error variances, varies from

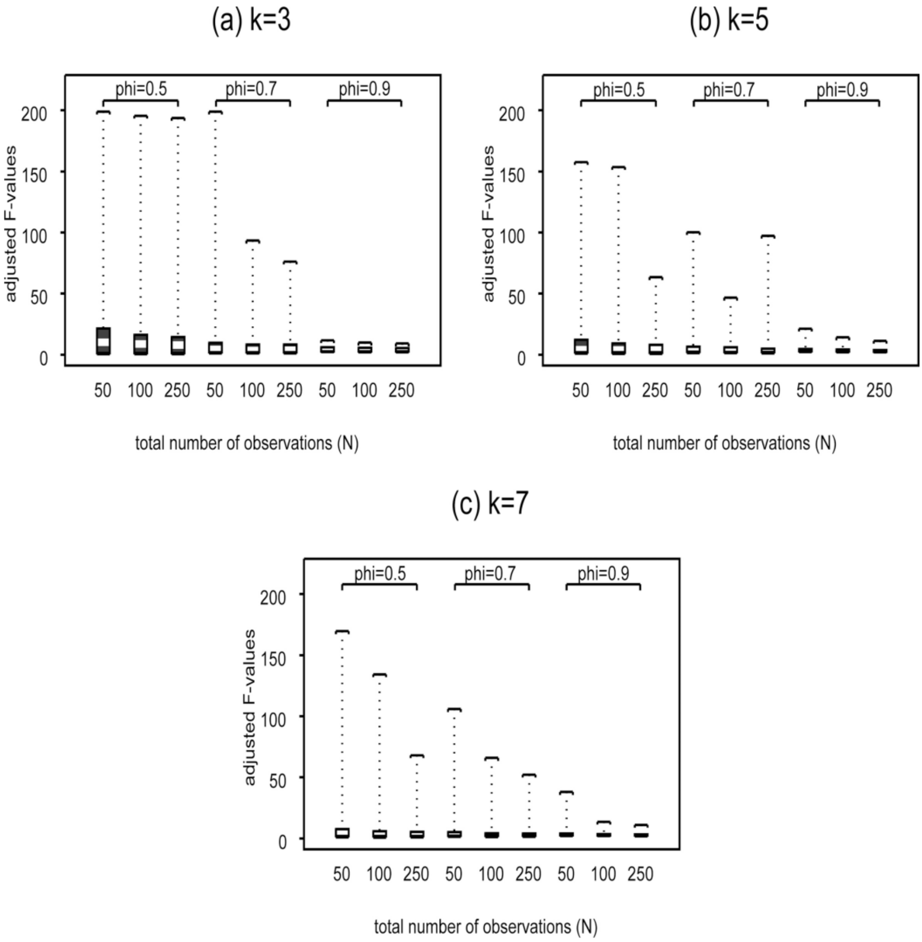

Figure 4. Box plots of the true α-critical values of the F-test, F_α^{adj}, for $\alpha = 0.05$ when the maximum error variance is 100 times as large as the minimum.

2.136 to 3.195 (see the footnotes in table 2), depending on the values of k and N. Hence, the effect of heterogeneous error variances on F_α^{adj} can be quite considerable.

4. Extension to the fixed-effects model

The use of Davies' algorithm [22] can be extended to the one-way model with fixed effects, where α_i in model (1) is a fixed unknown parameter ($i = 1, 2, \ldots, k$). In this case, the hypothesis of interest is

$$H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_k,$$

or equivalently,

$$H_0\colon \boldsymbol{\alpha} = \alpha_0 \mathbf{1}_k,$$

where α_0 is the common value of the α_i's. The variance–covariance matrix for model (2) is $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, where, if we recall, $\boldsymbol{\Sigma}_0 = \oplus_{i=1}^k \sigma_i^2 \mathbf{1}_{n_i}$. Thus, under H_0, $\boldsymbol{y} \sim N[(\mu + \alpha_0)\mathbf{1}_N, \boldsymbol{\Sigma}_0]$. As in the random one-way model case, the standard F-test statistic is given by formula (4), which, as was seen before, does not follow the exact F-distribution, even under H_0. The corresponding level of significance can be calculated using the formula

$$\alpha_f = P\left[\boldsymbol{y}'\left(\boldsymbol{Q} - \frac{k-1}{N-k}F_{\alpha,k-1,N-k}\boldsymbol{R}\right)\boldsymbol{y} \geq 0 \mid E(\boldsymbol{y}) = (\mu + \alpha_0)\mathbf{1}_N\right]. \tag{14}$$

where $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}_0$. This is basically the same as formula (9) except that $E(\boldsymbol{y}) = (\mu + \alpha_0)\mathbf{1}_N$ instead of $\mu\mathbf{1}_N$. Consequently, the representation of the quadratic form in equation (14) as $\sum_{i=1}^{s} \tau_i \chi^2_{\kappa_i}$ (see Theorem 2.2) is still valid, where, as before, $\tau_1, \tau_2, \ldots, \tau_s$ are the distinct non-zero eigenvalues of $(\boldsymbol{Q} - (k-1)/(N-k)F_{\alpha,k-1,N-k}\boldsymbol{R})\boldsymbol{\Sigma}_0$ with multiplicities $\kappa_1, \kappa_2, \ldots, \kappa_s$, respectively. Thus Davies' algorithm can be applied resulting in the same values for α_f as in the random-effects case. The same conclusion applies to the computation of the true critical value of the F-test as in section 2.2.

5. Concluding remarks

When the traditional assumption of the homogeneous error variances is violated, the ANOVA-based F-test for testing the significance of the random effect's variance component, in a one-way random model, does not provide the desired level of significance. Unbalancedness of a design makes it more difficult to study the effect of unequal error variances on the level of significance of the test.

This study explored the effects of heterogeneous error variances and design imbalance on the level of significance of the ANOVA-based F-test. This was done by obtaining values of α_f, the level of significance in equation (7), for different combinations of k, the number of groups; N, the total number of observations; ϕ, the degree of imbalance; and the error variances. It was found that data imbalance had a larger effect on α_f than did heterogeneity of error variances.

We also provided a method for achieving a desired level of significance of the F-test. This was done by computing the true α-critical value of the F-test using Davies' algorithm [22]. Further study of the use of this critical value in investigating the power of the test is needed.

The case of the one-way model with fixed-effects was also considered. It was demonstrated that the values of α_f and the true critical values of the F-test, for comparing treatment means, are the same as in the random-effects case.

Acknowledgement

J. Lee was supported by grant A06-0171-B51004-06N1-00050B from the Korea health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea; S. Lee was supported by a grant no. R05-2003-000-10690-0 from the Basic Research Program of the Korea Science and Engineering Foundation.

References

[1] Hartley, H.O., 1950, The maximum F ratio as a short cut test for heterogeneity of variance. *Biometrika*, **37**, 308–312.

[2] Bartlett, M.S., 1937, Some examples of statistical methods of research in agriculture and applied biology. *Journal of the Royal Statistical Society*, Supplement **4**, 137–147.

[3] Levene, H., 1960, Robust tests for equality of variances. In: I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madows, and H.B. Mann (Eds) *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (Stanford, California: Stanford University Press).

[4] Conover, W.J., Johnson, M.E. and Johnson, M.M., 1981, A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, **23**, 351–361.

[5] Welch, B.L., 1947, The generalization of Student's problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

[6] James, G.S., 1951, Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, **38**, 19–43.

[7] Box, G.E.P., 1954, Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variances in the one-way classification. *Annals of Mathematical Statistics*, **25**, 290–302.

[8] Scheffé, H., 1959, *The Analysis of Variance* (New York: John Wiley).

[9] Brown, M.B. and Forsythe, A.B., 1974, The small sample behavior of some statistics which test the equality of several means. *Technometrics*, **16**, 129–132.

[10] Bishop, T.A. and Dudewicz, E.J., 1978, Exact analysis of variance with unequal variances: test procedures and tables. *Technometrics*, **20**, 419–430.

[11] Krutchkoff, R.G., 1988, One-way fixed effects analysis of variance when the error variances may be unequal. *Journal of Statistical Computation and Simulation*, **30**, 259–271.

[12] Draper, N.R. and Guttman, I., 1966, Unequal group variances in the fixed-effects one-way analysis of variance: a Bayesian sidelight. *Biometrika*, **53**, 27–35.

[13] Wilcox, R.R., 1988, A new alternative to the ANOVA F and new results on James' second-order method. *British Journal of Mathematical and Statistical Psychology*, **41**, 109–117.

[14] Alexander, R.A. and Govern, D.M., 1994, A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, **19**, 91–101.

[15] Smith, T. and Peddada, S.D., 1998, Analysis of fixed effects linear models under heteroscedastic errors. *Statistics and Probability Letters*, **37**, 399–408.

[16] Rao, P.S.R.S, Kaplan, J. and Cochran, W.G., 1981, Estimators for the one-way random effects model with unequal error variances. *Journal of the American Statistical Association*, **76**, 89–97.

[17] Jeyaratnam, S. and Othman, A.R., 1985, Test of hypothesis in one-way random effects model with unequal error variances. *Journal of Statistical Computation and Simulation*, **21**, 51–57.

[18] Singh, B., 1991, Distribution of variance ratio in unbalanced random model under heterogeneity of error variances. *Communications in Statistics, Part A – Theory and Methods*, **20**, 3021–3028.

[19] Singh, B., 2003, Estimation of heritability in heteroscedastic one-way unbalanced random model. *Biometrical Journal*, **45**, 527–540.

[20] Searle, S.R., Casella, G. and McCulloch, C.E., 1992, *Variance Components* (New York: John Wiley).

[21] Johnson, N.L. and Kotz, S., 1970, *Continuous Univariate Distributions* – 2 (New York: John Wiley).

[22] Davies, R.B., 1980, The distribution of a linear combination of χ^2 random valiables. *Applied Statistics*, **29**, 323–333.

[23] Ahrens, H. and Pincus, R., 1981, On two measures of unbalancedness in a one-way model and their relation to efficiency. *Biometrical Journal*, **23**, 227–235.

[24] Khuri, A.I., 1987, Measures of imbalance for unbalanced models. *Biometrical Journal*, **29**, 383–396.

[25] Lera Marques, L., 1994, Measures of imbalance for higher order designs. *Biometrical Journal*, **36**, 481–490.

[26] Ahrens, H. and Sanchez, J., 1992, Imbalance and its influence on variance component estimation. *Biometrical Journal*, **34**, 539–555.

[27] Donner, A. and Koval, J.J., 1987, A procedure for generating group sizes from a one-way classification with a specified degree of imbalance. *Biometrical Journal*, **29**, 181–187.

[28] Khuri, A.I., 1996, A method for determining the effect of imbalance. *Journal of Statistical Planning and Inference*, **55**, 115–129.