Can the Bartlett test really detect the heterogeneity of variances?

Crysttian Arantes Paixão*
Universidade Federal de Santa Catarina, Curitibanos/SC, Brasil
crysttian.arantes.paixao@ufsc.br

Katia Alves Campos*
Instituto Federal de Educação, Ciência e Tecnologia, Machado/MG, Brasil
katia.campos@ifsuldeminas.edu.br

Augusto Ramalho de Morais*
Universidade Federal de Lavras, Departamento de Ciências Exatas, Lavras/MG, Brasil
armorais@dex.ufla.br

Abstract

The hypotheses about the variances were established to propose a favorable situation (simulation on H_0 is true, $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma^2$ and an unfavorable one (on H_0 is false, $H_1 : 4\sigma_i^2$, with i = 0, 1, 2 e 3) in order to assess the efficiency of Bartlett's Test (1937). Six scenarios with seven response variables were generated without considering correlation between the variables: two under the hypothesis about the true averages, without differences as to the variance simulation ($H_0 - H_0$), and about heteroscedasticity ($H_0 - H_1$), two under H_1 homoscedastic, with two standard deviations of difference between the averages, $H_1 : 2s - H_0$ and also with eight standard deviations of difference, $H_1 : 8s - H_0$, and two under H_1 heterocedastic, $H_1 : 2s - H_1$ e $H_1 : 8s - H_1$, respectively. In order to evaluate the tests in each simulated response variable, the number of accepted null hypotheses was calculated. The results were presented as percentage (accepted number of $H_0/1000$). Overall, Bartlett's Test obtained high acceptance values of homoscedasticity, even for the simulated scenario under means difference of eight standard deviations (95.8%). This result suggests that in addition to Bartletts test other means of evaluating the variance homogeneity should also be used.

Keywords: Multivariate Analysis; simulation; Bartletts Test; variance homogeneity.

1. Introduction

The homogeneity of variances is one of assumptions for the Analysis of Variance and, its infringement may make impracticable the use of this statistical method. The Bartlett test is highlighted among parametric statistical tests of homoscedasticity (BARTLETT, 1937), which is a modified form of testing the Likelihood ratio.

According to RIBOLDI et al. (2014), the Bartlett test should be made standard, but they stated, in the end, that it should be only used when data are approximately normal.

The purpose of this study was to test the power of the Bartlett test, by using six simulated scenarios under the null and alternative hypotheses, both for means and for variances, which show differences of about eight standard deviations.

2. Material and Methods

Six scenarios containing seven response variables, namely, V_1, V_2, V_3, V_4, V_5, V_6 and V_7 were obtained, seeking to assess the power of the Bartlett test.

The scenarios created for the simulation are combinations between hypotheses about means, and hypotheses about variances of errors, which are described in the sequence. The data simulation was carried out considering the following two situations about hypotheses of means:

1. Simulation under the assumption that H_0 is true: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$,
2. Simulation under the assumption that H_0 is false:

$$H_1 : \begin{cases} \mu_1 & = & \mu \\ \mu_2 = \mu_3 = \mu_4 & = & \mu + 2\sigma \end{cases} \quad (1)$$

where:

μ_i is the population mean for each variable in the ith treatment, containing i = 1, 2, 3 and 4 for each simulated response variable;

μ is the population mean of each simulated response variable; and

σ is the square root of the population variance for each simulated response variable.

Values of μ and σ are shown in the Table 1, and were obtained from an experiment performed using the randomized block design aiming to assess the effect of organic fertilization in the growth of coffee seedlings.

Table 1: Values of population mean and variance used as basis for the simulations

	V1	V2	V3	V4	V5	V6	V7
Mean	21.03	2.56	21.22	0.67	0.18	20.24	7.41
Variance	22.35	0.10	4.44	0.10	0.01	60.08	1.17

Seeking to test the reaction of decisions about the Bartlett test with the increasing of the difference among means, were also simulated data about the alternative hypothesis for the mean of treatments, on which the difference among means was increased to eight standard deviations:

$$H_1 : \begin{cases} \mu_1 & = & \mu \\ \mu_2 = \mu_3 = \mu_4 & = & \mu + 8\sigma \end{cases} \quad (2)$$

Hypotheses about variances were established in order to propose a favourable situation (simulation about the true H_0), and other unfavourable (about the false H_0), aiming to asses the efficiency of multivariable methods tested on data which infringe the homoscedasticity:

1. Simulation under the assumption that H_0 is true: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma^2$
2. Simulation under the assumption that H_0 is false:

$$H_1 : \begin{cases} \sigma_1^2 & = & \sigma^2 \\ \sigma_2^2 & = & 4\sigma^2 \\ \sigma_3^2 & = & 16\sigma^2 \\ \sigma_4^2 & = & 64\sigma^2 \end{cases} \quad (3)$$

where:

σ_i^2 is the population variance for each variable in the ith treatment, containing i = 1, 2, 3 and 4 for each simulated response variable; and

σ^2 is the population variance for each simulated response variable.

Table 2: Description of combinations of hypotheses chosen to establish simulated scenarios, and their names

Hypothesis about means (μ)	Differences among means	Hypothesis about variances (σ^2)	Ratio between variances	Name of simulated scenarios
true	0 σ	true	0 σ^2	$H_0 - H_0$
true	0 σ	false	$4^i\sigma^2$	$H_0 - H_1$
false	2 σ	true	0 σ^2	$H_1 : 2s - H_0$
false	8 σ	true	0 σ^2	$H_1 : 8s - H_0$
false	2 σ	false	$4^i\sigma^2$	$H_1 : 2s - H_1$
false	8 σ	false	$4^i\sigma^2$	$H_1 : 8s - H_1$

i = 0, 1, 2, 3

A binomial, on which the first expression is related to hypothesis about means of treatments, and the second is about the variance of errors, was named aiming to make easy the presentation of results of analysis about each proposed scenario. These binomials are described in the Table 2, summarizing the relative hypotheses about means and variances.

For each of six scenarios shown in the Table 2, about 1,000 simulations were carried out using the R Statistical Package (R CORE TEAM, 2014).

For each of seven response variables (V_1, V_2, V_3, V_4, V_5, V_6 and V_7) simulated in each of different scenarios proposed here was carried out an Univariate Analysis of Variance, according to PIMENTEL-GOMES (2009), and was assessed the homogeneity of variance by means of the Bartlett test (BARTLETT, 1937).

For the tests assessment, the number of times on which the nullity hypothesis was accepted, for each simulated response variable, showing results in percentage [(number of H_0 acceptances/1,000)$\times$100%].

3. Results and Discussion

From all simulations for each tested scenario, the pattern of data simulated without considering the correlation was carried out. The percentage of acceptances of null hypotheses are shown in the Tables 3, 4 and 5, both for the effect of treatments ($H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$), and for variances ($H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$) from the Univariate Analysis of Variances of seven response variables. However, after estimating values of tests and the p value for each test, the count of tests which validated such hypotheses was carried out. For the ANOVA assumptions about the homoscedasticity, the count was carried out at 5% of significance and, for the F-test, was both at 1% and 5% of significance.

It may see in the Table 3 that there was, in general, a good control of rate of Type I error, because all simulated response variables showed an acceptance of equality of treatments effect over 95% for the scenario $H_0 - H_0$.

The Bartlett test detected homoscedasticity acceptance for over 99% of cases, and for the F-test was recorded the acceptance of the equality hypothesis of treatments close to 95% and 99% of significance, respectively for 5% and 1% of significance.

Regarding simulated data under $H_0 - H_1$, the Bartlett and F-tests showed the acceptance percentage of effects equality for treatments slightly lower to those simulated under $H_0 - H_0$. The acceptance percentage of homoscedasticity varied from 95.97% to 96.77%, which may be considered high, once was expected detecting many heteroscedasticity cases from the simulation. However, they are lower than homoscedasticity cases on which was detected over 99% of acceptance of H_0.

Table 3: Acceptance percentage of homogeneity of variances (Bartlett), and of F-test for each treatment at 1% and 5% of significance, under the null hypothesis ($H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$), obtained for data simulated with $\rho = 0$, and under homoscedasticity ($H_0 - H_0$) and heteteroscedasticity ($H_0 - H_1$) scenarios.

Scenario	Variable	Bartlett test	F-test ($p < 0.01$)	F-test ($p < 0.05$)
	V_1	99.47	99.43	95.47
	V_2	99.57	99.10	95.53
	V_3	99.60	98.83	95.07
$H_0 - H_0$	V_4	99.30	98.83	94.97
	V_5	99.53	99.40	96.03
	V_6	99.43	98.60	95.30
	V_7	99.67	99.03	95.07
	V_1	95.97	96.47	90.17
	V_2	96.77	95.90	90.43
	V_3	96.93	96.03	90.80
$H_0 - H_1$	V_4	95.97	96.03	90.60
	V_5	96.67	96.30	90.47
	V_6	96.80	96.33	90.33
	V_7	96.50	96.03	91.07

In general, the Bartlett test was found to show high values of homoscedasticity acceptance, even for the scenario simulated under difference of four standard deviations. This result suggests the need of using other tests for assessing the homogeneity of variances, instead.

Scenarios simulated under equality of treatments variances ($H_0 - H_0$, $H_1 : 2s - H_0$ and $H_1 : 8s - H_0$) were found to show a minimum value of about 99.30% from all samples considered homoscedastic, V_1 (Table 4, the scenario $H_1 : 8s - H_0$), when the difference among means was estimated in eight standard deviations.

For data simulated under the heteroscedasticity ($H_0 - H_1$, $H_1 : 2s - H_1$ and $H_1 : 8s - H_1$), containing an amplitude of about 64 times the population variance (σ^2), the acceptance percentage of homoscedasticity was high; about 97.37% for the variable V_2, and 95.80% for the variable V_3, both estimated for the scenario $H_1 8s - H_1$ (Table 5).

Likewise, there was a reduction in the acceptance percentage of H_0 for the F-test. There was the acceptance percentage about 96% at 1% of significance, and about 90% at 5%. In relation to the scenario $H_0 - H_0$, this reduction was due to the presence of heterogeneous variances.

For data simulated under the false H_0 scenario for means of treatments, both homoscedastic ($H_1 - H_0$), the acceptance percentage of errors normality was not affected with the increasing of differences among means of treatments (Table 4). A similar situation occurred in relation to the acceptance of variance homogeneity (Bartlett test). For the F-test, the difference of two standard deviations, at 1% of significance, showed minimum acceptance values of treatments equality about 47.77% for the response variable V_2, and maximum about 51.20% for V_4. This detection error of differences of treatments effects, the Type II error, may be explained in relation to data variability, showing that difference between means of two standard deviations is not easily detected. However, when the difference between means for eight standard deviation increased, there was a 100% rejection of the dataset, both at 1% and at 5% of significance (Table 5).

For the heteroscedasticity scenario (Table 5), when the difference between means was established in two standard deviations, even at 5% significance, the acceptance percentage of equality among effects of treat-

Table 4: Acceptance percentage of homogeneity of variances (Bartlett), and of F-test for each treatment at 1% and 5% of significance, under the null hypothesis ($H_0 : \mu_1 = \mu 2 = \mu_3 = \mu_4$), obtained for data simulated with $\rho = 0$, and under homoscedasticity hypothesis, with differences for means given for two ($H_1 : 2s - H_0$) and for eight standard deviations ($H_1 : 8s - H_0$)

Scenario	Variable	Bartlett test	F-test ($p < 0.01$)	F-test ($p < 0.05$)
	V_1	99.53	49.77	19.20
	V_2	99.37	47.77	19.13
	V_3	99.43	49.97	< 0.0001
$H_1 : 2s - H_0$	V_4	99.73	51.20	21.67
	V_5	99.50	49.53	19.77
	V_6	99.43	50.83	20.50
	V_7	99.60	50.33	20.20
	V_1	99.30	< 0.0001	< 0.0001
	V_2	99.57	< 0.0001	< 0.0001
	V_3	99.60	< 0.0001	< 0.0001
$H_1 : 8s - H_0$	V_4	99.43	< 0.0001	< 0.0001
	V_5	99.63	< 0.0001	< 0.0001
	V_6	99.53	< 0.0001	< 0.0001
	V_7	99.60	< 0.0001	< 0.0001

ments was greater than 86% of cases. The same inference may be made for data simulated using different and homoscedastic means. Even when the difference among means of eight standard deviations increased, there was an acceptance over 14%. This reinforces the assumption of that the heteroscedasticity of variances interferes in the detection of treatments equality, because the scenario $H_0 - H_1$ also showed the decreasing of the acceptance percentage of the null hypothesis for treatments means.

It was also found that the acceptance percentages of homoscedasticity of variances were lower than those found under H_0. For the Bartlett test, maybe it might be expected a lesser acceptance percentage, once simulation was carried out under H_1. With the increase of means difference, from two to eight standard deviations (Tables 4 and 5), there was no significant differences in relation to acceptance percentages of homoscedasticity.

4. Conclusions

The Bartlett test was found to show high acceptance values of homoscedasticity, even for the scenario simulated under differences of means of eight standard deviations, which suggests the need of using other tests to verify the homogeneity of variances.

References

BARTLETT, M. S. Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London, Series A, London, v. 160, n. 2, p. 268-282, 1937.

PIMENTEL-GOMES, F. Curso de estatística experimental. 15. ed. Piracicaba: FEALQ, 2009. 451 p.

R CORE TEAM (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

RIBOLDI, Joo et al. Precisão e poder de testes de homocedasticidade paramétricos e não-paramétricos avaliados por simulação. Rev. Bras. Biom, v. 32, n. 3, p. 334-344, 2014.

Table 5: Acceptance percentage of homogeneity of variances (Bartlett), and of F-test for each treatment at 1% and 5% of significance, under the null hypothesis ($H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$), obtained for data simulated with $\rho = 0$, and under heteroscedasticity hypothesis, with differences for means given for two ($H_1 : 2s - H_1$) and for eight standard deviations ($H_1 : 8s - H_1$)

Scenario	Variable	Bartlett test	F-test ($p < 0.01$)	F-test ($p < 0.05$)
	V_1	96.13	95.37	88.10
	V_2	96.33	96.27	87.97
	V_3	96.43	95.40	88.30
$H_1 : 2s - H_1$	V_4	96.30	95.63	88.13
	V_5	96.60	96.00	88.57
	V_6	96.17	95.40	87.77
	V_7	96.17	95.07	87.23
	V_1	96.67	58.03	27.30
	V_2	97.37	59.50	28.37
	V_3	95.80	59.67	29.33
$H_1 : 8s - H_1$	V_4	96.43	60.40	28.37
	V_5	97.07	58.80	27.37
	V_6	96.57	57.90	27.83
	V_7	96.63	57.40	26.87