Avaliação do Poder de Alguns Testes de Homogeneidade das Variâncias

Marta Eliane Echeverria Borges
Departamento de Engenharia Química, UEM

Josmar Mazucheli
Departamento de Estatística,UEM

1 Introdução

A suposição de homogeneidade das variâncias de ($a \geqslant 2$) populações normais univariadas, dentre vários motivos, é extremamente importante para se testar a hipótese de igualdade de médias. Alternativamente, em situações de populações normais multivariadas ($a \geqslant 2$), tem-se como suposição básica a igualdade de matrizes de variâncias covariâncias para se testar a igualdade de vetores de médias. Uma vez disponíveis a amostras independentes de tamanhos $n_1, \ldots, n_a$ destas a populações, (univariadas/multivariadas), antes da aplicação das técnicas usuais de análise de variância (anova/manova) é importante verificar se a suposição homocedasticidade é válida, pelo menos aproximadamente. De acordo com o grau de severidade da heterogeneidade das variâncias e/ou com os tamanhos das amostras, a análise de variância pode ter sua significância estatística comprometida (ver, por exemplo, Milliken e Johnson 1984). Estudos de simulação Monte Carlo mostraram que em situações de heterocedasticidade e amostras desbalanceadas, mesmo em situações de normalidade, ocorre um inflacionamento no erro do Tipo I. Box (1954) discute a robustez da análise de variância para as situações balanceadas, $n_1 = \cdots = n_a$. Mais recentemente, Legendre e Borcard avaliaram, na análise de variância com uma classificação, o efeito da violação da suposição de homogeneidade de variâncias sobre o erro do Tipo I e sobre o poder do teste F. Uma solução para a condução de análise de variância na presença de heterocedasticidade é discutida em Babu et al.(1999). Na ausência da suposição de normalidade, muitos trabalhos discutem o uso de métodos robustos (ver, por exemplo, Wilcox 1997).

Na prática, recomenda-se avaliar a suposição de homogeneidade de variâncias priori a análise de variância usual. Entretanto, em muitas áreas de aplicações, não somente priori a análise de variância, é importante avaliar a homogeneidade de variâncias por si só. Este fato diz respeito a área de controle de qualidade em que é importante avaliar como os parâmetros do processo produtivo afetam a variabilidade e média do produto final.

Neste trabalho, dado a importância de se avaliar a hipótese de homogeneidade das variâncias, foi conduzido um estudo de simulação Monte Carlo com o intuito de estimar o poder dos testes: Levene, Bartlett, Fligner-Killeen e de Hartley. A escolha destes testes no estudo de simulação esta relacionada com o fato de que os mesmo são os mais conhecidos na literatura e disponíveis na maioria dos softwares estatísticos. Vale ressaltar que, para amostras univariadas, existem na literatura mais de 50 testes para se testar a hipótese de homogeneidade de variâncias. Conover (1998) apresenta e discute alguns testes baseados nos postos das observações. Uma ampla e importante revisão é apresentada em Conover et al. (1981). Haris (1985) propôs 4 testes para se avaliar a suposição de homogeneidade de matrizes de variâncias covariâncias. Recentemente, Diniz et al.(2005) conduziram um estudo de simulação considerando a situação multivariada e os testes M-Box (Box, 1949) e Jensen (Jensen, 1991).

2 Estudo de Simulação

Existem dois Tipos de erros que podem ser cometidos quando hipóteses estatísticas são testadas. O primeiro consiste em rejeitar a hipótese nula quando a mesma é verdadeira. Este erro é chamado de erro do Tipo I (a probabilidade de se cometer o erro do Tipo I é representada por α). O outro erro ocorre quando não é possível rejeitar H_0 dado que de fato H_0 é falsa. Este erro é chamado de erro do Tipo II (a probabilidade de cometer um erro do Tipo II é representada por β). O poder, $\tau = (1 - \beta)$, de qualquer teste estatístico é definido como sendo a probabilidade de rejeitar a hipótese nula dado que a mesma é falsa, o que caracteriza uma situação de acerto. Um teste estatístico é chamado de ideal quando $\alpha = 0$ e $\tau = (1 - \beta) = 1$. Em termos práticos isto jamais ocorre. Na prática o valor do nível de significância, $\alpha > 0$, é fixado e escolhe-se o teste com maior poder. Por outro lado, um teste estatístico é chamado de exato se a probabilidade de cometer um erro do Tipo I é α. Se tem um teste conservativo quando a probabilidade do Tipo I nunca excede o valor de α. Um teste exato é conservativo embora a recíproca possa não ser verdade.

Dado um problema de testar hipóteses e vários testes concorrentes, é importante que se conheça algumas destas propriedades. Entretanto, para a maioria dos testes, estas e outras propriedade somente podem ser avaliadas via estudos de simulação Monte Carlo, como conduzido neste artigo.

Gerando a amostras independentes sob a hipótese alternativa, tem-se a estimativa do poder calculando a proporção de vezes em que H_0 é acertadamente rejeitada. Definido $\hat{\tau} = \left(1 - \hat{\beta}\right)$ como sendo o poder empírico do teste tem-se:

$$\hat{\tau} = \frac{\text{Número de vezes em que } H_0 \text{ é rejeitada} \mid H_0 \text{ é falsa}}{M}.$$

Para todos os testes definidos, o poder empírico foram estimados gerando amostras a partir das distribuições: Normal, Laplace, Logística e Valor-Extremo (Tabela 1). Sem perdas de generalidades, as M amostras foram geradas considerando o parâmetro de locação $\mu = 0$. Para todas as configurações consideradas no estudo de simulação, foram adotados níveis de significância de 1% ($\alpha = 0.01$)e 5% ($\alpha = 0.05$). Todo o estudo de simulação foi conduzido no software R (Ihaka e Gentleman 1996).

Tabela 1: Distribuições consideradas no estudo de simulação.

Distribuição	Densidade	Variância
Normal	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	σ^2
Laplace	$\frac{1}{2\sigma} \exp\left(\frac{\lvert x-\mu\rvert}{\sigma}\right)$	$2\sigma^2$
Logística	$\frac{1}{\sigma} \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\left[1+\exp\left(\frac{x-\mu}{\sigma}\right)\right]^2}$	$\frac{\pi^2\sigma^2}{3}$
Valor Extremo	$\frac{1}{\sigma} \exp\left[\left(\frac{x-\mu}{\sigma}\right) - \exp\left(\frac{x-\mu}{\sigma}\right)\right]$	$\frac{\pi^2\sigma^2}{6}$

3 Estimação do Poder dos Testes

- **Distribuição Normal**

Para dados com distribuição Normal considerou-se as seguintes configurações: $a = 2, 3$ e 5, $n_i = 5, 10, 20, 30$ e 50 e $\sigma_i = 1, 2, 5, 10, 15$ e 20, em que $a =$ número de grupos, $n_i =$ tamanho do *i-ésimo* grupo e $\sigma_i =$ o desvio padrão do *i-ésimo* grupo, $i, \ldots, a$. Observando os resultados empíricos pode-se constatar que o teste de Bartlett é altamente poderoso para amostras com 10 ou mais elementos, ou seja na maioria das vezes, 100% das 20.000 simulações, a hipótese nula foi rejeitada quando falsa, exceto quando a análise homogeneidade de variância for para dois grupos com desvio padrão 1 e 2 respectivamente, para $\alpha = 1\%$ e 5%. Para $n_i = 5$ este teste apresentou poder para o nível de significância de 5%. O mesmo ocorreu para o teste de Hartley.

Para o teste de Fligner-Killeen e as formas do teste de Levene verificou-se que estes apresentaram um alto poder para $n = 20$ ou maiores, com nível de significância de $\alpha = 1\%$ e 5%, exceto para grupos de dois quando a variabilidade é pequena, ou seja com desvio padrão 1 e 2 respectivamente para $n = 20$ com $\alpha = 1\%$ e para $n = 10$ e $\alpha = 5\%$.

Vorapongsathorn, et al. (2004) numa simulação de 1000 vezes, verificou que o teste de Bartlett e Levene com amostra de 45 de tamanho em cada um dos três grupos foram altamente poderosos com variância ($\sigma = 1 : 2 : 4$) para $\alpha = 1\%$ e 5%.

- **Distribuição Laplace**

Para dados com distribuição Laplace considerou-se as seguintes configurações: $a = 2, 3$ e 5, $n_i = 5, 10, 20, 30$ e 50 e $\sigma_i = 0,7071$; $1,0000$; $1,5811$; $2,2361$; $2,7386$ e $3,1623$.

Analisando os resultados constatou-se que o teste de Bartlett manifestou poder acima de 60% quando a análise de dados constitui de 3 ou 5 grupos para amostras com pelo menos 10 observações, com os níveis de significância 1% e 5% , isto é quanto maior a heterogeneidade da variância entre os grupos e o tamanho da amostra maior o poder do teste.

Já o teste de Fligner-Killeen rejeitou H_0 quando ela falsa em mais de 80% das 20.000 simulações, quando trabalhado com 5 populações e amostras com no mínimo 20 elementos, sendo também poderoso para 3 grupos.

O teste de Hartley mostrou-se poderoso para a análise da homogeneidade das variâncias com 5 grupos para amostras pequenas ou grandes.

As formas do teste de Levene mostrou-se poderoso para grandes amostras ($n \geq 30$) em todas as configurações, exceto para dois grupos com desvio padrão $0,7071$ e $1,0$ respectivamente.

4 Considerações finais

O estudo mostrou que dados com distribuição Normal independente do tamanho da amostra, número de grupos e grau de severidade da variância é altamente poderoso quando utilizado o teste de Bartlett. O teste Hartley mostrou-se poderoso na maioria das configurações apresentadas. O mesmo ocorreu para dados com distribuição Laplace, porém com três e cinco grupos.

Constatou-se também que o comportamento do poder de todos os testes referente as distribuições consideradas, converge para 1 a medida que se aumenta o grau de heterogeneidade das variâncias e o número de grupos.

5 Referência Bibliográficas

Babu, G. J.; Padmanabhan, A. R. e Puri, M. L. (1999). Robust one-way anova under possibly non-regular conditions. Biometrical Journal, **3**, 321-339.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36, 317-346.

Box, G. E. P. (1954). Effects of inequality of variance and Correlation Between Errors in the Two Way Classification. Annals of Mathematical Statistics, **25**, 484-498.

Conover, W. J.; Johnson, M. E. e 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.

Conover, W. J.(1998). Practical Nonparametric Statistics. John Wiley e Sons, Inc. 584.

Diniz, C. A. R.; Mazucheli, J.; Caron, R. (2005) Power of two tests for the homogeneity of covariance matrices. Rev. Mat. Estatíst. 23, 2, 59-77.

Haris, P. (1985). Testing for variance homogeneity of correlated variables. Biometrika, **72**, 103-107.

Ihaka, R. e Gentleman, R. (1996). R: A language for lata analysis and graphics. Journal of Computational and Graphical Statistics, **5**, 3, 299-314.

Jensen, J. L. (1991). A large deviation-type approximation for the Box class of likelihood ratio criteria. Journal of the American. Statistical. Association, 86, 437-440.

Legendre, P. e Borcard, D. Statistical comparison of univariate tests of homogeneity of variances. Submitted to the Journal of Statistical Computation and Simulation. Diponível em: http://www.biol09.biol.umontreal.ca/BIO2042/MS_THV.pdf. Acessado em: 24/05/2008.

Milliken G. A. e Johnson, D. E. (1984). Analysis of messy data. New York: Van Nostrand Reinhold.

O´Neill, M. E. e Mathews, K. L. (2002). Levene test of homogeneity of variance for general block and treatment design. Biometrics, textbf58, 216-224.

Vorapongsathorn,T.; Taejaroenkul, S. e Viwatwongkasem, C. (2004). A comparison of type I error and power of Bartlett´s test, Levene´s test and Cochran´s test under violation of assumptions. Songklanakarin Journal Science Technologic, **26**, 4 , 537-547.

Wilcox, R. (1997). Introduction to robust estimation and hypothesis testing. Academic Press.