

# Método de múltiplas comparações

UM PROCEDIMENTO GRÁFICO DE MÚLTIPLAS COMPARAÇÕES PARA DIVERSOS DESVIOS PADRÃO

Senin J. Banga e Gregory D. Fox  
18 de junho de 2013

## RESUMO

É fornecido um novo procedimento gráfico para várias comparações de desvios padrão de  $k$ . Como teste para homogeneidade de variâncias, o novo procedimento tem um propriedades do tipo similar I e propriedades de erro II, como as versões de Brown e Forsythe (1974) do teste de Levene (1960),  $W_{50}$ . A exibição gráfica associada ao teste de várias comparações, contudo, fornece uma ferramenta visual útil para triagem de amostras com diferentes desvios padrão.

*Termos do índice: homogeneidade de variâncias, teste de Levene, teste de Brown-Forsythe, teste de Layard, várias comparações*

## 1. Introdução

A modificação de Brown e Forsythe (1974) do teste de Levene (1960), comumente chamado de teste de  $W_{50}$  é, talvez, um dos procedimentos mais amplamente usados para testar a homogeneidade (igualdade) de variâncias. Em parte, o teste  $W_{50}$  é popular devido a ele ser robusto e ser assintoticamente livre de distribuição. Comparado com outros testes de homogeneidade de variâncias, o teste  $W_{50}$  também é fácil de calcular. (Para uma comparação desses testes, consulte Conover et al. (1981).) Além disso, o teste  $W_{50}$  é facilmente acessível porque ele está disponível em diversos pacotes de softwares como SAS, Minitab, R e JMP.

Contudo, para algumas distribuições, o teste  $W_{50}$  pode ser muito baixo, particularmente em pequenas amostras. Por exemplo, a Pan (1999) mostra que para algumas distribuições, incluindo a distribuição normal, o teste  $W_{50}$  pode não ter poder suficiente para detectar diferenças entre dois desvios padrão, independentemente da magnitude das diferenças.

Não está claro a partir da análise de Pan se a mesma limitação seria aplicada a designs de várias amostras. Espera-se que esta limitação não se aplique a experimentos com mais de duas amostras, simplesmente porque esses experimentos provavelmente incluem mais dados do que experimentos de duas proporções. O teste  $W_{50}$  é conhecido por ter boas propriedades de amostra maior (Miller, 1968, Brown e Forsythe, 1974, Conover et al., 1981).

Tornou-se uma prática comum seguir um teste significativo  $W_{50}$  com um procedimento de comparação pareado simultâneo, com base em uma correção de multiplicidade Bonferroni. Conforme indicado por Pan (1999), contudo, essa abordagem provavelmente irá falhar ou produzir resultados enganosos, devido ao baixo poder do teste  $W_{50}$  em experimentos de duas proporções. Usar a correção de Bonferroni piora o problema, porque ela é conservadora, particularmente quando o número de comparações pareadas é grande. Em contraste, diversos procedimentos de comparações múltiplas eficazes estão disponíveis para comparação de médias seguindo uma ANOVA para um fator. Para obter exemplos, consulte Tukey (1953), Hochberg et al. (1982) e Stoline (1981). Uma análise post-hoc análoga para comparações entre variâncias de amostras seria útil.

Neste teste, propomos um método gráfico para comparação das variâncias (ou desvios padrão) de múltiplas amostras. A análise está baseada em "intervalos de incerteza" para variâncias que são similares aos intervalos de incerteza descritos por Hochberg et al. (1982) para médias. Primeiro, um procedimento de múltiplas comparações pareadas está baseado na versão modificada de Bonett (2006) do teste de Layard (1973) para a igualdade de variâncias de experimentos de duas proporções. A correção de multiplicidade usada nas comparações pareadas está baseada em uma generalização do método de Tukey-Kramer (Tukey, 1953; Kramer, 1956), proposto por Nakayana (2009). Intervalos de incerteza, que nos referimos como "intervalos de múltiplas comparações" ou "intervalos de MC", são derivados do procedimento de comparação pareada usando o melhor procedimento aproximado descrito por Hochberg et al. (1982). O teste de MC resultante rejeita a hipótese nula se, e somente se, pelo menos um par dos intervalos de MC não se sobrepuserem. Intervalos de MC não sobrepostos identificam as amostras que têm variâncias significativamente diferentes (ou desvios padrão).

Realizamos estudos de simulação para avaliar as propriedades de pequenas amostras do teste de MC. Para comparação, nós também incluímos o teste  $W_{50}$  nos estudos de simulação.

## 2. Procedimento de múltiplas comparações gráficas

Permita que o  $Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$  seja amostras independentes de  $k$ , cada amostra sendo independente e identicamente distribuída com a média de  $E(Y_{il}) = \mu_i$  e a variância de  $\text{Var}(Y_{il}) = \sigma_i^2 > 0$ . Além disso, suponha que as amostras se originem de populações com uma curtose comum de  $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$ .

Além disso, permita que o  $\bar{Y}_i$  e o  $S_i$  sejam a média e o desvio padrão da amostra  $i$ , respectivamente. Permita que o  $m_i$  seja a média aparada da amostra  $i$  com a proporção aparada do  $1/[2\sqrt{n_i - 4}]$  e permita que o  $\hat{\gamma}_{ij}$  seja um estimador da curtose combinada de amostras  $(i, j)$ , dado como

$$\begin{aligned}\hat{\gamma}_{ij} &= (n_i + n_j) \frac{\sum_{l=1}^{n_i} (Y_{il} - m_i)^4 + \sum_{l=1}^{n_j} (Y_{jl} - m_j)^4}{\left[ \sum_{l=1}^{n_i} (Y_{il} - \bar{Y}_i)^2 + \sum_{l=1}^{n_j} (Y_{jl} - \bar{Y}_j)^2 \right]^2} \\ &= (n_i + n_j) \frac{\sum_{l=1}^{n_i} (Y_{il} - m_i)^4 + \sum_{l=1}^{n_j} (Y_{jl} - m_j)^4}{\left[ (n_i - 1)S_i^2 + (n_j - 1)S_j^2 \right]^2}\end{aligned}$$

Observe que o  $\hat{\gamma}_{ij}$  é assintoticamente equivalente ao estimador da curtose combinada do Layard (1973), onde a média da amostra do  $\bar{Y}_i$  foi substituída pela média aparada do  $m_i$ . Desta forma, o  $\hat{\gamma}_{ij}$  é um estimador consistente da curtose comum desconhecida do  $\gamma$ , contanto que as variâncias de população sejam iguais. Bonett (2006) propõe este estimados em vez do estimador da curtose combinada de Layard para melhorar o desempenho de pequenas amostras do teste de Layard em problemas de duas proporções. Todo o tempo, nos referimos à versão modificada de Bonett (2006) do teste de Layard simplesmente como teste de Bonett.

Supponha que haja mais de dois grupos independentes ou amostras para comparar ( $k > 2$ ). O procedimento de comparação múltipla gráfica que propomos é derivado de múltiplas comparações pareadas que são baseadas no teste de Bonett. Uma abordagem alternativa é basear as comparações pareadas no teste  $W_{50}$ . Em experimentos de duas amostras, o poder de desempenho do teste de  $W_{50}$  é problemático para algumas distribuições incluindo a distribuição normal (Pan, 1999). Além disso, Banga e Fox (2013) mostram que os intervalos de confiança da razão das variâncias que são baseados no teste de Bonett são, geralmente, superiores àqueles baseados no teste de  $W_{50}$ .

Dado qualquer par  $(i, j)$  de amostras, um teste de Bonett bilateral com nível de significância  $\alpha'$  rejeita as hipóteses nulas da igualdade de variâncias se, e somente se,

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > z_{\alpha'/2} \sqrt{\frac{\hat{\gamma}_{ij} - k_i}{n_i - 1} + \frac{\hat{\gamma}_{ij} - k_j}{n_j - 1}}$$

onde  $z_{\alpha'/2}$  é o ponto do percentil superior de  $\alpha'/2 \times 100^{\text{th}}$  da distribuição normal padrão,

$$k_i = \frac{n_i - 3}{n_i}, k_j = \frac{n_j - 3}{n_j}, c_i = \frac{n_i}{n_i - z_{\alpha'/2}}, c_j = \frac{n_j}{n_j - z_{\alpha'/2}}$$

Uma vez que há múltiplas comparações pareadas, exatamente comparações  $k(k - 1)/2$ , é necessário um ajuste de multiplicidade. Por exemplo, se um nível de significância geral ou por família alvo,  $\alpha$ , for dado, uma abordagem comum, conhecida como a correção Bonferroni, é para escolher o nível de significância de cada uma das comparações pareadas de  $k(k - 1)/2$ ,  $\alpha' = 2\alpha/(k(k - 1))$ . A correção Bonferroni, contudo, é bem conhecida por produzir procedimentos de comparação pareados cada vez mais conservadores, conforme aumenta o número de amostras para comparação. Uma abordagem alternativa e melhor é proposta por Nakayama (2009) e é baseada em uma aproximação de grandes amostras do método Tukey-Kramer (Tukey, 1953; Kramer, 1956). Especificamente, o teste de múltiplas comparações pareadas gerais é significativo se, e somente se, o seguinte é verdadeiro para algum par de amostras de  $(i, j)$ :

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > \frac{q_{k,\alpha}}{\sqrt{2}} \sqrt{\frac{\hat{y}_{ij} - k_i}{n_i - 1} + \frac{\hat{y}_{ij} - k_j}{n_j - 1}}$$

onde  $q_{\alpha,k}$  é o ponto superior de  $\alpha$  do intervalo de variáveis independentes e aleatórias normais padrão identicamente distribuídas de  $k$ . Isto é,  $q_{\alpha,k}$  satisfaz

$$\Pr\left(\max_{1 \leq i < j \leq k} |Z_i - Z_j| \leq q_{\alpha,k}\right) = 1 - \alpha$$

onde  $Z_1, \dots, Z_k$  variáveis independentes e aleatórias normais padrão identicamente distribuídas. Barnard (1978) fornece um algoritmo numérico simples com base em uma quadratura Gaussian de 16 pontos para cálculo da função de distribuição do intervalo normal.

Conforme sugerido por Hochberg et al. (1982), um procedimento de comparação múltipla gráfica que se aproxima do procedimento de comparação pareada múltipla descrito acima irá rejeitar a hipótese nula se, e somente se,

$$|\ln(c_i S_i^2) - \ln(c_j S_j^2)| > q_{\alpha,k}(V_i + V_j)/\sqrt{2}$$

onde os  $V_i$  estão selecionados para minimizar o seguinte:

$$\sum_{i \neq j} \sum (V_i + V_j - b_{ij})^2$$

onde

$$b_{ij} = \sqrt{\frac{\hat{y}_{ij} - k_i}{n_i - 1} + \frac{\hat{y}_{ij} - k_j}{n_j - 1}}$$

A solução para este problema, conforme ilustrado em Hochberg et al. (1982), é escolher

$$V_i = \frac{(k - 1) \sum_{j \neq i} b_{ij} - \sum_{1 \leq j < l \leq k} b_{jl}}{(k - 1)(k - 2)}$$

Conclui-se que um teste de homogeneidade de variâncias com base neste procedimento aproximado rejeita a hipótese nula se, e somente se, pelo menos um par dos intervalos dados a seguir não se sobrepuserem:

$$\left[ S_i \sqrt{c_i \exp(-q_{\alpha,k} V_i / \sqrt{2})}, S_i \sqrt{c_i \exp(q_{\alpha,k} V_i / \sqrt{2})} \right], i = 1, \dots, k$$

O procedimento MC gráfico consiste na exibição desses intervalos em um gráfico para identificar visualmente as amostras com intervalos não sobrepostos. Além disso, o valor-p do teste geral da homogeneidade de variância (ou desvio padrão) pode ser determinado. Fornecemos algoritmos detalhados para cálculo do valor-p na próxima seção. Mas, primeiro, indicamos alguns fatos simples sobre o procedimento MC.

#### OBSERVAÇÃO

1. O estimador de curtose combinada,  $\hat{\gamma}_{ij}$ , com base no par  $(i, j)$  de amostras, poderia ter sido substituído pelo estimador da curtose combinada geral, com base em todas as amostras de  $k$ . Apesar desta abordagem simplificar um pouco os cálculos, os resultados da simulação que não são mostrados aqui, indicam que usar o  $\hat{\gamma}_{ij}$  produz resultados melhores.
2. O intervalo correspondente à amostra  $i$  não é um intervalo de confiança para o desvio padrão da população pai da amostra. Hochberg et al. (1982) refira-se a esse intervalo como um "intervalo de incerteza". Nos referimos a ele como um "intervalo de múltipla comparação" ou um "Intervalo de MC". Os intervalos de MC são úteis para a comparação de desvios padrão ou variâncias de experimentos de múltiplas amostras.
3. Os intervalos de MC que são descritos neste documento podem ser usados apenas para comparar mais de dois desvios padrão. Quando existem apenas duas amostras, os intervalos de comparação podem ser criados, mas eles transmitem as mesmas informações que são fornecidas pelos resultados dos testes. É muito mais informativo criar um intervalo de confiança da razão dos desvios padrão, como o descrito por Banga e Fox (2013) e fornecido com o comando do Minitab, Variância de duas amostras.

## 3. Valor-p do método de múltiplas comparações gráficas

Antes de descrever o algoritmo para calcular o valor-p do método gráfico (MC), primeiro derivamos o valor-p associado à modificação de Bonett (2006) do teste de Layard em experimentos de duas amostras. A seguir mostraremos como aplicar os resultados do experimento de duas amostras ao procedimento de comparações múltiplas.

### 3.1 Valor-p em experimentos de duas amostras

Como mencionado anteriormente, os ajustes de Bonett (2006) do teste de Layard, em experimentos de duas amostras, rejeitam a hipótese nula de homogeneidade de variâncias se, e somente se

$$|\ln(c_1 S_1^2) - \ln(c_2 S_2^2)| > z_{\alpha/2} se$$

ou equivalente

$$|\ln(c_{\alpha/2} S_1^2 / S_2^2)| > z_{\alpha/2} se$$

onde

$$se = \sqrt{\frac{\hat{y}_{12} - k_1}{n_1 - 1} + \frac{\hat{y}_{12} - k_2}{n_2 - 1}}$$

$$c_{\alpha/2} = \frac{c_1}{c_2} = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Bonett introduziu a constante  $c_{\alpha/2}$  como um ajuste de pequenas amostras para mitigar o efeito de probabilidades de erro de caudas diferentes em experimentos não balanceados de pequenas amostras. O efeito da constante, no entanto, é insignificante em experimentos não balanceados de grandes amostras e a constante não tem efeito em experimentos balanceados.

Conclui-se que, se o experimento é balanceado, o valor-p do teste bilateral de homogeneidade das variâncias é simplesmente calculado como

$$P = 2 \Pr(Z > |Z_0|)$$

onde

$$Z_0 = \frac{\ln(S_1^2) - \ln(S_2^2)}{se}$$

Se o experimento não é balanceado, então  $P = 2 \min(\alpha_L, \alpha_U)$ , onde  $\alpha_L$  é a menor solução para  $\alpha$  na equação,

$$\exp[\ln(c_\alpha S_1^2/S_2^2) - z_\alpha se] = 1 \quad (1)$$

e  $\alpha_U$  é a menor solução de  $\alpha$  na equação,

$$\exp[\ln(c_\alpha S_1^2/S_2^2) + z_\alpha se] = 1 \quad (2)$$

Algoritmos para encontrar  $\alpha_L$  e  $\alpha_U$  são fornecidos a seguir. Os detalhes matemáticos dos algoritmos são adiados para a seção Apêndice.

Permitir

$$L(z, n_1, n_2, S_1, S_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z se + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

Também permita

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

As soluções  $\alpha_L$  e  $\alpha_U$  são calculados nos seguintes passos:

Caso 1:  $n_1 < n_2$

- Calcular  $z_m$  como fornecido no resultado acima e avaliar  $L(z_m, n_1, n_2, S_1, S_2)$ .
- Se  $L(z_m) \leq 0$ , encontre a raiz  $z_L$ , de  $L(z, n_1, n_2, S_1, S_2)$  no intervalo,  $(-\infty, z_m]$  e calcule  $\alpha_L = \Pr(Z > z_L)$ .
- Se  $L(z_m) > 0$ , então a função  $L(z, n_1, n_2, S_1, S_2)$  não tem raiz. Definir  $\alpha_L = 0,0$ .

Caso 2:  $n_1 > n_2$

- Calcule  $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2/S_2^2$ .
- Se  $L(0, n_1, n_2, S_1, S_2) \geq 0$ , encontre a raiz,  $z_o$ , de  $L(z, n_1, n_2, S_1, S_2)$  no intervalo  $(0, n_2)$ , caso contrário encontra a raiz  $z_L$  no intervalo  $(-\infty, 0)$ .
- Calcule  $\alpha_L = \Pr(Z > z_L)$ .

Para calcular  $\alpha_U$ , simplesmente aplicamos os passos acima usando a função,  $L(z, n_2, n_1, S_2, S_1)$ , em vez da função,  $L(z, n_1, n_2, S_1, S_2)$ .

## 3.2 Valor-p de múltiplas comparações gráficas

Supondo-se que existem  $k$  ( $k > 2$ ) amostras no experimento, permitimos que  $P_{ij}$  seja o valor-p do teste associado com qualquer par  $(i, j)$  de amostras. Recordamos que o teste de comparações múltiplas rejeita a hipótese nula da homogeneidade de variâncias se, e somente se, pelo menos um par de intervalos de comparação de  $k$  não se sobrepuserem. Conclui-se que o valor-p geral associado com o procedimento de comparações múltiplas é

$$P = \min\{P_{ij}, 1 \leq i < j \leq k\}$$

Para calcular  $P_{ij}$  realizamos o algoritmo de experimentos de duas amostras usando

$$se = V_i + V_j$$

onde  $V_i$  é definido anteriormente.

Se  $n_i \neq n_j$ , então

$$P_{ij} = \min(\alpha_L, \alpha_U)$$

onde  $\alpha_L = \Pr(Q > z_L\sqrt{2})$ ,  $\alpha_U = \Pr(Q > z_U\sqrt{2})$  e  $z_L$  são as menores raízes da função,  $L(z, n_i, n_j, S_i, S_j)$  e  $z_U$  são as menores raízes da função e  $L(z, n_j, n_i, S_j, S_i)$  e  $Q$  são as variáveis aleatórias conforme definido anteriormente. As quantidades  $z_L$  e  $z_U$  são encontradas aplicando-se o algoritmo do experimento de duas amostras descrito anteriormente para o par  $(i, j)$  de amostras.

Se  $n_i = n_j$  então  $P_{ij} = \Pr(Q > |z_o|\sqrt{2})$ , onde

$$z_o = \frac{\ln S_i^2 - \ln S_j^2}{V_i + V_j}$$

## 4. Estudos de simulação e resultados

Dois estudos de simulação principal são realizados para investigar o desempenho de pequenas amostras do teste de MC como um teste geral para a homogeneidade das variâncias. Todas as simulações foram conduzidas usando-se a versão 8 do pacote de software Matemática.

# Estudo 1

O primeiro estudo é criado para avaliar e comparar as propriedades de erro do tipo I do teste de MC e do teste de  $W_{50}$ . Comparamos o desempenho de dois testes com amostras geradas a partir de várias distribuições em três experimentos diferentes: um experimento de 3 amostras, um experimento de 4 amostras e um experimento de 6 amostras. Em cada experimento, o tamanho das amostras variam de 10 a 50 em incrementos de 10. As amostras são traçadas a partir das seguintes distribuições pai:

- a distribuição normal
- distribuições simétricas de cauda leve, representadas pela distribuição uniforme e a distribuição Beta com os parâmetros de (3, 3)
- distribuições simétricas de cauda pesada, representadas pela distribuição t com 5 graus de liberdade ( $t(5)$ ) e a distribuição Laplace
- distribuições de cauda pesada e assimétrica, representadas pela distribuição exponencial, uma distribuição qui-quadrado com 1 grau de liberdade ( $\chi^2(1)$ ) e uma distribuição qui-quadrado com 5 graus de liberdade ( $\chi^2(5)$ )
- uma distribuição normal contaminada (CN(0,9, 3)) para a qual 90% de observações são traçadas a partir da distribuição padrão normal e os 10% restantes são traçados a partir de uma distribuição normal com uma média de 0 e um desvio padrão de 3.

Cada simulação consiste em 10.000 amostras replicadas. O nível nominal e alvo de  $\alpha$  alcançado é de 0,05. O erro da simulação é de 0,002 aproximadamente. Os níveis de significância simulados para cada teste são relatados na Tabela 1.

**Tabela 1** Comparação dos Níveis de significância simulada ( $\alpha = 0,05$ )

Descrição	Distribuição [Curtose]	$n_i$	$k = 3$		$k = 4$		$k = 6$	
			MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$
Normal	Normal [3,0]	10	0,038	0,033	0,038	0,031	0,036	0,029
		20	0,039	0,038	0,040	0,038	0,041	0,033
		30	0,043	0,041	0,044	0,038	0,046	0,039
		40	0,046	0,043	0,046	0,041	0,048	0,041
		50	0,046	0,046	0,046	0,044	0,052	0,047
Simétricas com cauda leve	Uniforme [1,8]	10	0,029	0,029	0,025	0,024	0,023	0,020
		20	0,028	0,026	0,030	0,026	0,028	0,023
		30	0,037	0,035	0,034	0,032	0,034	0,030
		40	0,038	0,037	0,037	0,037	0,035	0,033
		50	0,041	0,041	0,036	0,036	0,036	0,036
	Beta(3, 3) [2,5]	10	0,031	0,032	0,031	0,029	0,031	0,025
		20	0,035	0,031	0,036	0,027	0,037	0,026



Descrição	Distribuição [Curtose]	$n_i$	$k = 3$		$k = 4$		$k = 6$	
			MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$
		30	0,041	0,035	0,037	0,034	0,037	0,032
		40	0,040	0,036	0,039	0,035	0,040	0,033
		50	0,044	0,039	0,044	0,037	0,044	0,035
Simétricas com caudas pesadas	Laplace [6,0]	10	0,056	0,038	0,063	0,041	0,071	0,039
		20	0,054	0,044	0,058	0,043	0,059	0,041
		30	0,051	0,042	0,053	0,043	0,052	0,044
		40	0,048	0,045	0,048	0,045	0,048	0,046
		50	0,045	0,045	0,051	0,046	0,049	0,047
	$t(5)$ [9,0]	10	0,042	0,032	0,044	0,031	0,042	0,031
		20	0,043	0,039	0,045	0,038	0,045	0,040
		30	0,039	0,040	0,040	0,040	0,041	0,040
		40	0,041	0,042	0,040	0,041	0,039	0,038
		50	0,040	0,050	0,039	0,046	0,038	0,046
Assimétricas com caudas pesadas	$\chi^2(5)$ [5,4]	10	0,040	0,039	0,046	0,040	0,048	0,039
		20	0,040	0,043	0,040	0,040	0,042	0,039
		30	0,039	0,047	0,042	0,044	0,043	0,042
		40	0,040	0,046	0,041	0,044	0,039	0,042
		50	0,037	0,047	0,038	0,047	0,040	0,048
	Exponencial [9,0]	10	0,063	0,051	0,073	0,049	0,076	0,048
		20	0,051	0,049	0,053	0,048	0,057	0,046
		30	0,042	0,048	0,046	0,051	0,049	0,049
		40	0,034	0,050	0,038	0,046	0,037	0,049
		50	0,033	0,045	0,037	0,047	0,038	0,046
	$\chi^2(1)$ [15,0]	10	0,084	0,048	0,098	0,050	0,118	0,050
		20	0,053	0,046	0,060	0,047	0,068	0,046
		30	0,041	0,041	0,045	0,045	0,050	0,047
		40	0,044	0,049	0,046	0,047	0,045	0,047
		50	0,038	0,050	0,037	0,049	0,040	0,049
Normal contaminado	CN(0,9, 3) [8,3]	10	0,020	0,016	0,018	0,012	0,016	0,010
		20	0,014	0,015	0,012	0,013	0,008	0,007

Descrição	Distribuição [Curtose]	$n_i$	$k = 3$		$k = 4$		$k = 6$	
			MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$
		30	0,012	0,014	0,010	0,011	0,007	0,008
		40	0,009	0,017	0,009	0,014	0,006	0,008
		50	0,009	0,016	0,007	0,012	0,006	0,009

Os resultados mostram que ambos os testes apresentam bom desempenho para a maioria das distribuições. A maioria dos níveis de significância simulados estão perto do alvo de 0,05. Entretanto, os níveis de significância simulados para ambos os testes tendem a ser conservadores (menor do que 0,05) quando pequenas amostras são traçadas a partir de distribuições de cauda leve normal e simétrica. Para essas distribuições, os níveis de significância simulados do teste de MC estão mais próximos do nível de significância alvo do que aqueles do teste de  $W_{50}$ .

Quando amostras pequenas são traçadas a partir de distribuições de cauda pesada, o teste de  $W_{50}$  tende a ser conservador e o teste de MC tende a ser liberal. O teste de MC é ainda mais liberal quando pequenas amostras são traçadas de distribuições extremamente assimétricas. Por exemplo, quando amostras de tamanho 10 são traçadas a partir de uma distribuição qui-quadrado com 1 grau de liberdade, os níveis de significância simulados do teste de MC são 0,084, 0,098 e 0,118 para distribuições de 3, 4 e 6 amostras, respectivamente.

Ambos os testes são influenciados por outliers. Os níveis de significância da distribuição normal contaminada são extremamente conservadores mesmo quando os tamanhos das amostras são tão grandes quanto 50.

## Estudo 2

O segundo estudo avalia e compara as propriedades de erro do tipo II (poder) dos dois procedimentos em um experimento de 4 amostras. Usamos as mesmas amostras para este estudo que aquelas usadas para as amostras de tamanho 20 e a condição  $k = 4$  no Estudo 1. As observações são escalonadas por um fator de 1, 2, 3 ou 4. Por exemplo, na condição denotada como 1:1:4:4, as observações para as amostras 1 e 2 são as mesmas que aquelas usadas no Estudo 1. As observações nas amostras 3 e 4 são escalonadas por um fator de 4.

Incluimos a condição 1:1:1:1 para comparação. Observe que os resultados desta condição são os mesmos que aqueles relatados no Estudo 1 para amostras de tamanho 20 e  $k = 4$ . Escolhemos amostras de tamanho 20 porque os resultados do Estudo 1 sugerem que, para ambos os testes, as amostras de tamanho 20 produzem níveis de significância alcançados que estão próximos do nível alvo para a maioria das distribuições.

Os níveis de potência simulados nesses experimentos são calculados como a proporção de réplicas que levam a rejeições da hipótese nula da homogeneidade de variâncias.

Os resultados estão relatados na Tabela 2.

**Tabela 2** Comparação dos Níveis de significância simulados ( $\alpha = 0,05$ )

Description	Distribuição	Razão do desvio padrão							
		1:1:1:1		1:1:2:2		1:2:3:4		1:1:4:4	
		MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$	MC	$W_{50}$
	Normal	0,040	0,038	0,846	0,853	0,998	0,994	1,000	1,000
Simétricas com cauda leve	Uniforme	0,030	0,026	0,985	0,962	1,000	0,999	1,000	1,000
	Beta(3, 3)	0,036	0,027	0,938	0,916	1,000	0,999	1,000	1,000
Simétricas com cauda pesada	Laplace	0,058	0,043	0,597	0,629	0,931	0,921	0,996	0,998
	$t(5)$	0,045	0,038	0,657	0,703	0,952	0,949	0,997	0,998
Assimétricas com cauda pesada	$\chi^2(5)$	0,040	0,040	0,625	0,704	0,949	0,949	0,996	0,999
	Exponencial	0,053	0,048	0,431	0,507	0,804	0,779	0,963	0,978
	$\chi^2(1)$	0,060	0,047	0,298	0,291	0,602	0,504	0,838	0,824
Contaminado	CN(0,9, 3)	0,012	0,013	0,499	0,612	0,889	0,917	0,989	0,998

Os resultados sugerem que as propriedades de erro do tipo II (poder) do teste de MC e o teste de  $W_{50}$  são similares. Em geral, os níveis de poder simulados alcançados com ambos os testes são da mesma ordem de magnitude. Em apenas um caso o poder dos dois testes difere em mais de 0,1.

Os níveis de poder simulados do teste de MC são ligeiramente melhores do que aqueles do teste de  $W_{50}$  quando amostras são traçadas de distribuições simétricas com caudas de leve a moderadas. Por outro lado, o teste de  $W_{50}$  parece ser ligeiramente mais poderoso do que o teste de MC quando as amostras são traçadas de distribuições com caudas pesadas.

## 5. Exemplo

Nesta seção, aplicamos o procedimento de MC gráfico e o teste de  $W_{50}$  a um conjunto de dados obtidos de Ott et al. (2010), página 397. Os dados são descritos como a seguir:

*Uma empresa de fundição tem vários fornos nos quais eles aquecem a matéria-prima antes de vertê-la em um molde de cera. É muito importante que esses metais sejam aquecidos a uma temperatura precisa com muito pouca variação. Três fornos são selecionados aleatoriamente e suas temperaturas são registradas (°C) com muita exatidão em 10 aquecimentos sucessivos. Os dados coletados são os seguintes:*

<b>Forno 1</b>	1670,87	1670,88	1671,51	1672,01	1669,63	1670,95	1668,70	1671,86	1669,12	1672,52
<b>Forno 2</b>	1669,16	1669,60	1669,76	1669,18	1671,92	1669,69	1669,45	1669,35	1671,89	1673,45
<b>Forno 3</b>	1673,08	1672,75	1675,14	1674,94	1671,33	1660,38	1679,94	1660,51	1668,78	1664,32

A Figura 1 mostra boxplots das temperaturas para cada forno. Os boxplots sugerem que não há outliers nas temperaturas registradas, e que a variabilidade da temperatura do Forno 3 é diferente daquela do Forno 1 ou do Forno 2.

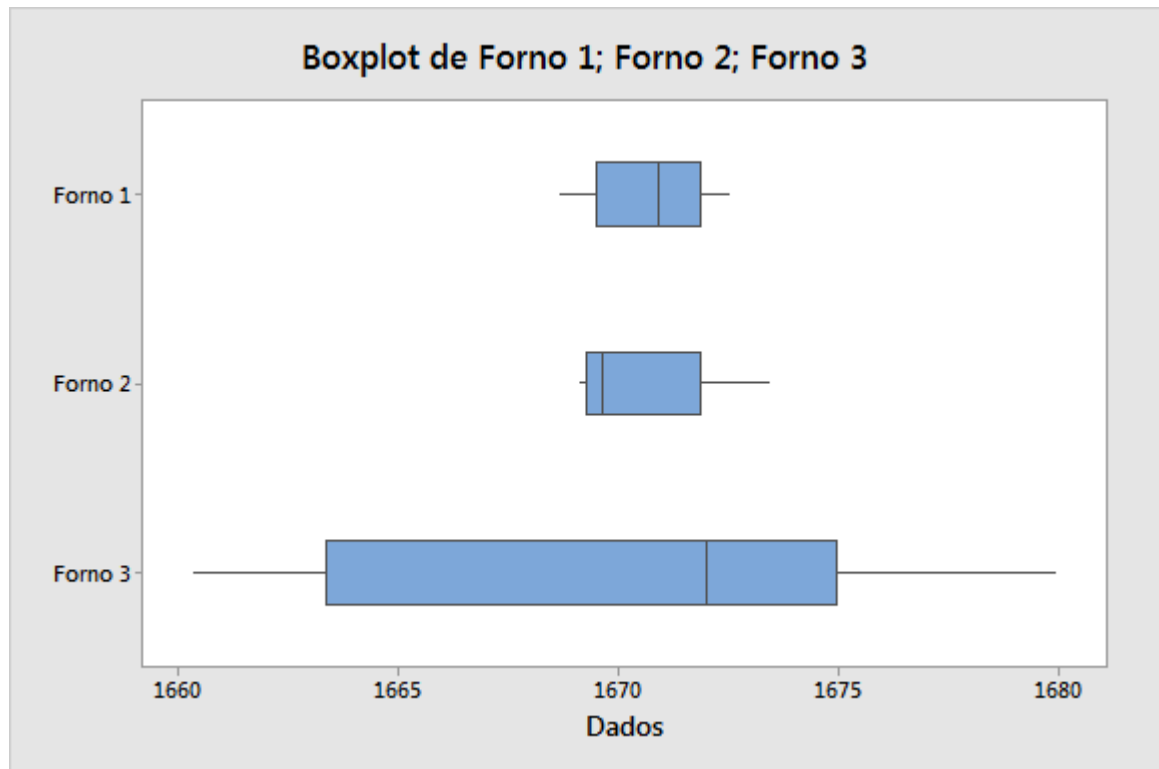


Figura 1 Boxplots de temperatura de forno (°C)

A Figura 2 mostra os intervalos de MC dos mesmos dados, e também os resultados do teste de MC geral e o teste de  $W_{50}$ , que é referido na legenda como teste de Levene. Os valores-p significativos para ambos os testes indicam que a variabilidade nas temperaturas é diferente entre os três fornos. Os intervalos de MC não sobrepostos confirmam que a variabilidade do Forno 3 é diferente daquela do Forno 2 ou do Forno 1. Os intervalos de MC são (0,896, 2,378), (1,072, 2,760) e (4,366, 12,787) para os Fornos 1, 2 e 3, respectivamente.

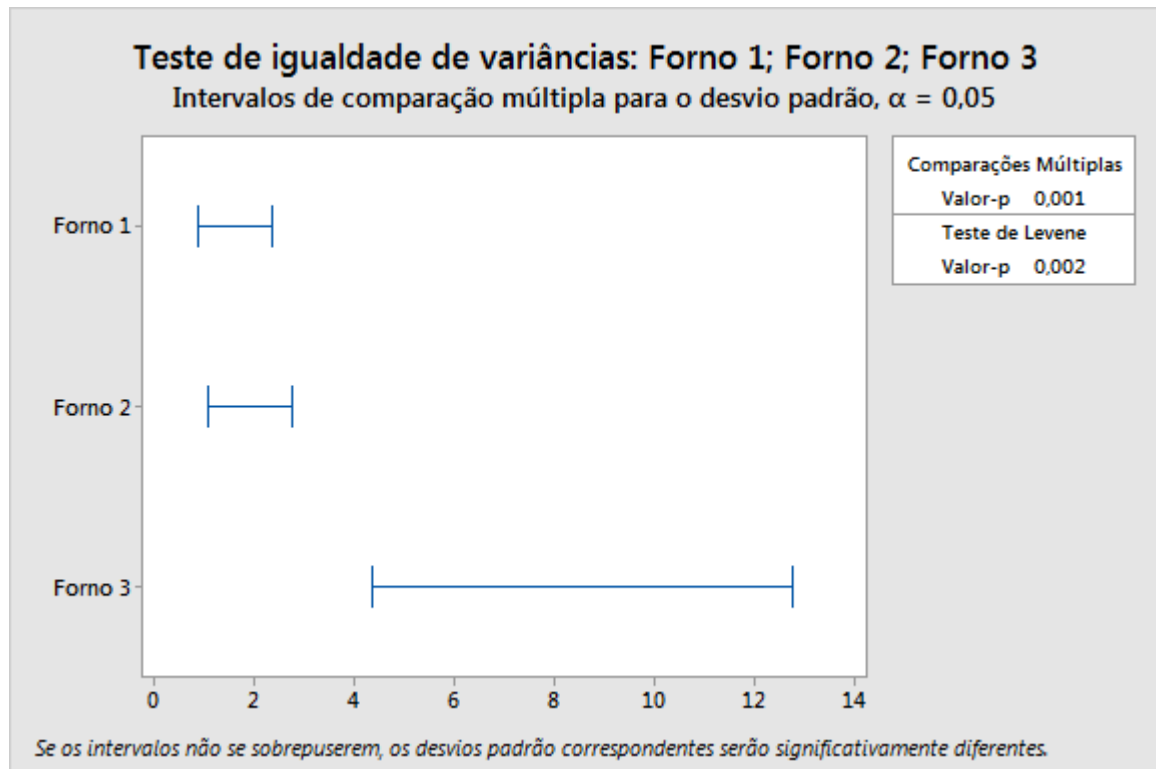


Figura 2 Os intervalos de MC e os valores-p do teste de MC e do teste de  $W_{50}$  (teste de Levene)

## 6. Conclusão

No geral, os resultados da simulação mostram que, para experimentos com múltiplas amostras pequenas, o desempenho do teste de MC é similar àquele do teste  $W_{50}$ . O teste de MC é ligeiramente mais ajustado às distribuições simétricas ou quase simétricas com caudas de leves a moderadas, enquanto o teste de  $W_{50}$  pode ser preferido quando os dados são traçados a partir de distribuições altamente assimétricas e de distribuições com caudas pesadas. Uma vantagem clara do procedimento de MC é que ele fornece uma ferramenta visual poderosa para triagem de amostras com desvios padrão ou variâncias diferentes quando o teste geral da homogeneidade de desvios padrão é significativo. O procedimento de MC gráfico está disponível no Minitab, versão 17.

## 7. Apêndice

O ajuste de Bonett (2006) do teste de Layard em experimentos de duas amostras rejeita a hipótese nula de homogeneidade de variâncias se, e somente se,

$$|\ln(c_1 S_1^2) - \ln(c_2 S_2^2)| > z_{\alpha/2} se$$

ou equivalente

$$|\ln(c_{\alpha/2} S_1^2 / S_2^2)| > z_{\alpha/2} se$$

onde

$$se = \sqrt{\frac{\hat{Y}_{12} - k_1}{n_1 - 1} + \frac{\hat{Y}_{12} - k_2}{n_2 - 1}}$$

$$c_{\alpha/2} = \frac{c_1}{c_2} = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Assim, se o experimento é balanceado, então  $c_{\alpha/2} = 1$ , de forma que o valor-p do teste seja simplesmente

$$P = 2 \Pr(Z > |Z_0|)$$

onde

$$Z_0 = \frac{\ln(S_1^2) - \ln(S_2^2)}{se}$$

Se o experimento não é balanceado, então  $P = 2 \min(\alpha_L, \alpha_U)$ , onde

$\alpha_L$  é a menor solução para  $\alpha$  na equação

$$\exp[\ln(c_{\alpha} S_1^2 / S_2^2) - z_{\alpha} se] = 1 \quad (1)$$

e  $\alpha_U$  é a menor solução  $\alpha$  da equação

$$\exp[\ln(c_{\alpha} S_1^2 / S_2^2) + z_{\alpha} se] = 1 \quad (2)$$

A abordagem para solucionar essas equações de  $\alpha$  é, primeiro, solucionar as equações de  $z \equiv z_{\alpha}$  e, depois, obter  $\alpha = \Pr(Z > z)$  onde a variável aleatória  $Z$  tem a distribuição normal padrão. Antes de descrevermos como solucionar essas equações, observamos que a equação (1) pode ser expressa novamente como a equação  $L(z) = 0$  onde

$$L(z, n_1, n_2, S_1, S_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z se + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

Similarmente, a equação (2) é equivalente à equação  $U(z) = 0$ , onde

$$U(z, n_1, n_2, S_1, S_2) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} + z se + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

Observamos que  $L(z, n_2, n_1, S_2, S_1) = -U(z, n_1, n_2, S_1, S_2)$ . Consequentemente, as raízes de apenas uma das duas funções deve ser encontrada.

O algoritmo para solução da equação (1) ou (2) é derivado do seguinte resultado:

## Resultado

*Permita que o  $n_1, n_2, S_1$  e o  $S_2$  sejam dados e corrigidos. Para experimentos não balanceados, a função,  $L(z, n_1, n_2, S_1, S_2)$ , tem, no máximo, duas raízes.*

4. Se  $n_1 < n_2$ , então  $L(z, n_1, n_2, S_1, S_2)$  é convexo: ele satisfaz  $L(-\infty, n_1, n_2, S_1, S_2) = L(n_1, n_1, n_2, S_1, S_2) = +\infty$  e alcança seu mínimo em

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

Assim, se  $L(z_m, n_1, n_2, S_1, S_2) \leq 0$ , então há duas raízes: uma no intervalo  $(-\infty, z_m]$  e a outra no intervalo  $[z_m, n_1)$ . Por outro lado, se  $L(z_m, n_1, n_2, S_1, S_2) > 0$ , então a função  $L(z, n_1, n_2, S_1, S_2)$  não tem nenhuma raiz.

5. Se  $n_1 > n_2$ , então  $L(z, n_1, n_2, S_1, S_2)$  reduz monotonicamente de  $+\infty$  para  $-\infty$  e, portanto, tem uma raiz única. Se  $L(0, n_1, n_2, S_1, S_2) = \ln S_1^2/S_2^2 \geq 0$ , então a raiz está no intervalo de  $(0, n_2)$ ; caso contrário, ele reside no intervalo de  $(-\infty, 0)$ .

## Prova

A seguir, permitimos  $L(z) \equiv L(z, n_1, n_2, S_1, S_2)$ .

Primeiro, queremos provar que, se  $n_1 < n_2$ , então  $L(z)$  é convexo e alcança seu mínimo em

$$z_m = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

Como anteriormente definido

$$L(z) = \ln \frac{n_1}{n_2} + \ln \frac{n_2 - z}{n_1 - z} - z se + \ln \frac{S_1^2}{S_2^2}, z < \min(n_1, n_2)$$

Então, temos  $\lim_{z \rightarrow -\infty} L(z) = +\infty$  e

$$\lim_{z \rightarrow \min(n_1, n_2)} L(z) = \begin{cases} +\infty & (n_1 < n_2) \\ -\infty & (n_2 < n_1) \end{cases}$$

Além disso, observe que derivativo de  $L(z)$  satisfaz

$$-\frac{(n_1 - z)(n_2 - z)}{se} L'(z) = z^2 - (n_1 + n_2)z + n_1 n_2 + \frac{n_1 - n_2}{se}$$

Permitir

$$Q(z) = -\frac{(n_1 - z)(n_2 - z)}{se} L'(z)$$

Se  $n_1 < n_2$ , então o  $Q(z)$  quadrático tem duas raízes dadas como

$$z_1 = \frac{n_1 + n_2 - \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

e

$$z_2 = \frac{n_1 + n_2 + \sqrt{(n_1 - n_2)(n_1 - n_2 - \frac{4}{se})}}{2}$$

Uma vez que  $Q(n_1) = \frac{n_1 - n_2}{se} < 0$ , temos  $z_1 < n_1 = \min(n_1, n_2) < z_2$ , de forma que  $Q(z) > 0$  para  $z$  em  $(-\infty, z_1)$  e, de forma que  $Q(z) < 0$  para  $z$  em  $(z_1, n_1)$ . Segue-se que  $L'(z) < 0$  para

$z$  em  $(-\infty, z_1)$  e que  $L'(z) > 0$  para  $z$  em  $(z_1, n_1)$ . Desta forma,  $L(z)$  é convexo no domínio de  $(-\infty, \min(n_1, n_2))$  e alcança seu valor mínimo em  $z_1 \equiv z_m$ .

Se  $n_1 > n_2$ , então há dois casos, o caso onde  $n_1 - n_2 > 4/se$  e o caso onde

$0 < n_1 - n_2 < 4/se$ . No primeiro caso,  $z_1$  e  $z_2$  são as raízes de  $Q(z)$  de tal modo que  $n_2 = \min(n_1, n_2) < z_1 < z_2$ . (Isso é porque  $n_2 - \frac{z_1+z_2}{2} = \frac{n_2-n_1}{2} < 0$ ). Assim  $Q(z) > 0$  para  $z$  no domínio de  $(-\infty, \min(n_1, n_2))$ . No segundo caso,  $Q(z)$  não tem raízes, de forma que  $Q(z) > 0$  no domínio.

Segue-se que se  $n_1 > n_2$ , então  $L'(z) < 0$  de forma que  $L(z)$  diminui monotonicamente de  $+\infty$  para  $-\infty$ .

## 8. Referências

- Banga, S. J. e Fox, G. D. (2013). Em Robust Confidence Interval for a Ratio of Standard Deviations de Bonett. No prelo.
- Barnard, J. (1978). Probability Integral of the Normal Range. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 27, 197–198.
- Bonett, D. G. (2006). Robust Confidence Interval for a Ratio of Standard Deviations. *Applied Psychological Measurements*, 30, 432–439.
- Brown, M. B., and Forsythe A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, 69, 364–367.
- 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.
- Hochberg, Y., Weiss, G., and Hart S. (1982). On Graphical Procedures for Multiple Comparisons. *Journal of the American Statistical Association*, 77, 767–772.
- Kramer, C. Y. (1956). Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications. *Biometrics*, 12, 307–310.
- Layard, M. W. J. (1973). Robust Large-Sample Tests for Homogeneity of Variances. *Journal of the American Statistical Association*, 68, 195–198.
- Levene, H. (1960). "Robust Tests for Equality of Variances," em I. Olkin, ed., *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 278–292.
- Miller, R. G. (1968). Jackknifing Variances. *Annals of Mathematical Statistics*, 39, 567–582.
- Nakayama, M. K. (2009). Asymptotically Valid Single-Stage Multiple-Comparison Procedures. *Journal of Statistical Planning and Inference*, 139, 1348–1356.
- Ott, R. L. and Longnecker, M. (2010). *An introduction to Statistical Methods and Data Analysis, sixth edition*, Brooks/Cole, Cengage Learning.
- Pan, G. (1999). On a Levene Type Test for Equality of Two Variances. *Journal of Statistical Computation and Simulation*, 63, 59–71.



Stoline, M. R. (1981). The Status of Multiple of Comparisons: Simultaneous Estimation of All Pairwise Comparisons in One-Way ANOVA Designs. *The American Statistician*, 35, 134–141.

Tukey, J. W. (1953). *The Problem of Multiple Comparisons*. Monografia mimeografada.

Wolfram, S. (1999). *The Mathematica Book*, 4th ed. Wolfram Media/Cambridge University Press.

© 2015, 2017 Minitab Inc. All rights reserved.

Minitab®, Quality. Analysis. Results.® and the Minitab® logo are all registered trademarks of Minitab, Inc., in the United States and other countries. See [minitab.com/legal/trademarks](http://minitab.com/legal/trademarks) for more information.