

TESTE DE STUDENT-NEWMAN-KEULS *BOOTSTRAP*: PROPOSTA, AVALIAÇÃO E APLICAÇÃO EM DADOS DE PRODUTIVIDADE DA GRAVIOLA

Bruna de Oliveira GONÇALVES¹
Patrícia de Siqueira RAMOS²
Fabrício Goeking AVELAR¹

- RESUMO: Este trabalho teve por objetivos propor uma modificação do teste de comparações múltiplas SNK com o uso de reamostragem *bootstrap* (SNK_B) e avaliar o desempenho das duas versões do teste. O desempenho foi avaliado pelas taxas de erro tipo I por experimento e pelo poder por meio de um estudo de simulação Monte Carlo em condições de normalidade e não normalidade dos resíduos. Os dois testes foram exatos sob H_0 completa e normalidade. Sob H_0 completa e não normalidade, os testes SNK e SNK_B controlaram as taxas de erro tipo I por experimento e são considerados exatos na maior parte dos casos simulados. Sob H_0 parcial, o teste SNK_B foi liberal em todos os cenários, enquanto que o teste SNK foi, na maioria dos casos, conservador e liberal em alguns casos. Em geral, o poder cresceu com o aumento da diferença entre as médias simuladas, sob H_0 parcial e H_1 . Além disso, os testes foram aplicados em dados reais de um experimento delineado para avaliar os controles químico e mecânico de pragas da gravioleira com o objetivo de comparar os resultados obtidos pelos dois testes.
- PALAVRAS-CHAVE: Comparações múltiplas; reamostragem; simulação Monte Carlo; erro tipo I; poder.

1 Introdução

Comumente, ao se analisar um conjunto de dados, o pesquisador precisa decidir se, em média, os contrastes ou tratamentos aplicados produziram resultados iguais ou quais produzem um resultado superior, ou seja, testar a rejeição ou não da

¹Universidade Federal de Alfenas - UNIFAL, Instituto de Ciências Exatas, CEP: 37130-000, Alfenas, MG, Brasil. E-mail: brunag_25@hotmail.com; fabricao@unifal-mg.edu.br

²Universidade Federal de Alfenas - UNIFAL, Instituto de Ciências Sociais Aplicadas, CEP: 37048-395, Varginha, MG, Brasil. E-mail: patricia.amos@unifal-mg.edu.br

hipótese nula (H_0), sendo $H_0 : \mu_1 = \mu_2 = \dots = \mu_n$. Segundo Freund (2006) para verificar se existe pelo menos uma diferença significativa entre as médias dos tratamentos utiliza-se a análise de variância, porém, esta não indica quais médias diferem entre si. Os procedimentos de comparações múltiplas (PCM) permitem identificar essas diferenças.

Ao testar H_0 , o pesquisador corre o risco de cometer erros de decisão denominados erro tipo I e erro tipo II. O erro tipo I é cometido ao se rejeitar a hipótese nula verdadeira, cuja probabilidade de ocorrência denotaremos por α . O erro tipo II é cometido ao não se rejeitar a hipótese nula falsa e sua probabilidade é β . Segundo Bolfarine e Sandoval (2001), quando se rejeita H_0 , e ela de fato é falsa, considera-se que uma decisão correta foi tomada. Isto ocorre com probabilidade $1 - \beta$, valor que recebe o nome de poder do teste.

Os PCM se diferem quanto ao controle de erro tipo I e poder (MOOD et al., 1974). Segundo Souza e Lima Junior e Ferreira (2012), para escolher o melhor teste, deve-se considerar as qualidades estatísticas dos procedimentos (tipo de erro que é controlado e a forma como esse erro é controlado) e, de acordo com a taxa de erro tipo I, os testes podem ser classificados como conservadores, liberais ou exatos. Os testes são considerados conservadores quando possuem taxas de erro tipo I por experimento inferiores ao nível nominal de significância, são considerados liberais se a taxa de erro tipo I observada for superior ao nível nominal, ou seja, se possuem elevadas taxas de erro tipo I por experimento e os testes são considerados exatos quando a taxa de erro tipo I é igual ao nível de significância adotado (BIASE e FERREIRA, 2011).

Em estudos de desempenho de testes estatísticos, devido à dificuldade de se obter analiticamente informações sobre as taxas de erro tipo I e poder, a simulação Monte Carlo é uma alternativa viável para comparar os testes de comparações múltiplas (CARMER e SWANSON, 1973). Vários trabalhos, como os de Carmer e Swanson (1973), Silva e Ferreira e Bearzoti (1999), Borges e Ferreira (2003), Ramos e Ferreira (2009) e Ramos e Vieira (2014), utilizaram simulação Monte Carlo na avaliação de procedimentos de comparações múltiplas.

O que se espera de um teste é que ele controle o erro tipo I na maior parte das situações, mas apresente altas taxas de poder. Para melhorar o seu desempenho, em relação ao erro tipo I e poder, métodos de reamostragem *bootstrap* têm sido utilizados em alguns estudos sobre testes de comparações múltiplas de médias. Ramos e Ferreira (2009) e Ramos e Vieira (2014) utilizaram *bootstrap* para os procedimentos de comparações múltiplas de Caliński e Corsten (1985) e seu desempenho foi considerado superior ao dos testes originais.

A ideia básica de *bootstrap*, na ausência de qualquer conhecimento sobre a distribuição de probabilidade que gerou os dados, é realizar reamostragem com reposição de tamanho n da amostra original. A distribuição *bootstrap* de algum estimador de interesse é utilizada no lugar da distribuição teórica deste mesmo estimador, em função da dificuldade de desenvolvê-la ou do desconhecimento da distribuição da população de onde foi obtida a amostra aleatória (MANLY, 1998). Isso se refere ao *bootstrap* não paramétrico, mas há ainda o paramétrico, quando

se conhece a distribuição que gerou os dados, e o semiparamétrico, apropriado para algumas formas de regressão, que envolve reamostragem não paramétrica dos resíduos do modelo paramétrico (CARPENTER e BITHELL, 2000).

O teste de Student-Newman-Keuls (SNK) é um teste de comparações múltiplas, proposto por Newman (1939) e modificado por Keuls (1952), que controla as taxas de erro tipo I por experimento sob H_0 completa mas se torna liberal sob H_0 parcial. Além disso, seu poder é superior ao de alguns testes (RAMOS e FERREIRA, 2009). Por ser um teste sequencial, em que as médias são ordenadas antes de se aplicar o procedimento, nem todas as comparações duas a duas são realizadas, agilizando sua execução e identificando melhor grupos de médias consideradas iguais, se existirem. Portanto, é um teste com boas qualidades mas que poderiam ser melhoradas com o uso do *bootstrap*.

Assim, o objetivo desse trabalho foi propor uma outra versão do teste de comparações múltiplas SNK usando *bootstrap* paramétrico, avaliar e comparar o seu desempenho com o teste SNK original.

2 Metodologia

O desempenho de dois procedimentos de comparações múltiplas, SNK original e em sua versão *bootstrap*, foi avaliado por meio de simulação Monte Carlo. Foram delineadas simulações Monte Carlo sob a hipótese nula (H_0 completa), para mensurar as taxas de erro tipo I por experimento (TPE), e sob a hipótese alternativa (H_1), para mensurar o poder. Foram realizadas também simulações sob H_0 parcial, em que o erro tipo I foi medido dentro dos grupos de médias iguais e o poder foi medido entre os grupos de médias diferentes. Todas as rotinas necessárias para a implementação e avaliação dos testes foram realizadas utilizando o programa R em sua versão 3.0.2 (R CORE TEAM, 2014). O código em R utilizado para realizar as simulações se encontra em ANEXO, bem como alguns exemplos de aplicações dos testes a partir de dados simulados e dados reais. Uma das funções utilizadas, `order.SNK`, foi baseada na função `order.stat.SNK`, que pode ser encontrada no pacote ExpDes (2014).

Posteriormente, as taxas de erro tipo I e poder obtidas foram comparadas com os testes t de Student, Tukey, Duncan, avaliados por Girardi, Cargnelutti Filho e Storck (2009), como o teste de Scott-Knott, avaliado por Borges e Ferreira (2003) e Silva, Ferreira e Bearzoti (1999), e os procedimentos de Caliński e Corsten baseados na amplitude estudentizada e na distribuição F , em suas versões original e *bootstrap*, avaliados respectivamente por Ramos e Ferreira (2009) e Ramos e Vieira (2014). Os cenários simulados neste trabalho foram semelhantes aos simulados pelos autores citados.

2.1 Teste de Student-Newman-Keuls (SNK) original

Para a aplicação do teste SNK original pode ser usada a diferença mínima significativa (DMS) ou o valor- p . O roteiro a seguir ilustra como a decisão pode ser

tomada utilizando o valor- p para decidir se o contraste é significativo, considerando-se α o nível de significância adotado:

Primeiramente as médias de tratamentos ordenadas são representadas por $\bar{Y}_{(1)}, \dots, \bar{Y}_{(k)}$. Essas médias devem ser organizadas em ordem decrescente de magnitude. A seguir, devem ser obtidos os valores dos contrastes $|\hat{c}|_p$, o número de médias p abrangidas por esse contraste e os valores- p , seguindo o seguinte roteiro:

1. $\bar{Y}_{(k)}$ é a primeira média base;
2. Calcular o valor do contraste $|\hat{c}|_p$ entre a média base e a menor média e o valor- p do contraste:
 - a. Se valor- $p > \alpha$, todas as médias abrangidas pelo contraste recebem a mesma letra e a primeira diferente recebe outra letra;
 - b. Se valor- $p \leq \alpha$, repete-se o passo 2 tomando a média anterior na comparação com a média base até obter um valor- $p > \alpha$ ou até terminarem as médias;
3. Muda-se a média base para a próxima e repete-se até que a média base seja a penúltima.

O valor- p a ser calculado em cada contraste utiliza a distribuição da amplitude estudentizada ($q_\alpha(p, \nu)$) e é obtido por

$$\text{valor-}p = Pr \left(q_\alpha(p, \nu) > \frac{|\hat{c}|_p}{\sqrt{QMR/r}} \right), \quad (1)$$

em que p é o número de médias abrangidas pelo contraste, r número de repetições, ν são os graus de liberdade associados ao QMR , sendo este o quadrado médio do resíduo da análise de variância.

2.2 Teste SNK *bootstrap*

O teste SNK *bootstrap* difere do teste original no cálculo do valor- p para decidir se as médias abrangidas pelo contraste em questão podem ser consideradas iguais ou diferentes, pois foi obtida uma distribuição da amplitude estudentizada q para cada reamostragem. Foi utilizado o *bootstrap* na versão paramétrica porque se as pressuposições do modelo estiverem, pelo menos, aproximadamente corretas, os resultados do *bootstrap* paramétrico tendem a ser mais precisos do que a versão não paramétrica (KRISHNAMOORTHY et al., 2007).

Para implementar o *bootstrap* paramétrico, amostram-se rk observações de uma distribuição normal com média μ e variância σ^2 , compondo novas amostras de cada tratamento, em que r é o número de repetições e k o número de tratamentos. Novas médias são calculadas a partir das novas amostras obtidas e um valor q_b , quantil da distribuição *bootstrap*, é obtido por:

$$q_b = \frac{\bar{Y}_{(k)}^b - \bar{Y}_{(1)}^b}{\sqrt{\frac{QMR_b}{r}}}, \quad (2)$$

em que QMR_b é o quadrado médio do resíduo da b -ésima amostra *bootstrap* e $\bar{Y}_{(1)}^b$ e $\bar{Y}_{(k)}^b$ são as médias amostrais ordenadas obtidas na b -ésima amostra. O conjunto com todos os valores q_b obtidos ($b = 1, 2, \dots, B$) será utilizado para obter os valores- p por meio da expressão:

$$\text{valor-}p = \frac{\sum_{b=1}^B I\left(\frac{|\hat{c}|_p}{\sqrt{QMR_b/r}} < q_b\right)}{B}, \quad (3)$$

em que p é o número de médias abrangidas pelo contraste e $I(x)$ é a função indicadora de x que satisfaz a seguinte regra:

$$\begin{cases} I(x) = 1 & \text{se } x \text{ for verdadeira} \\ I(x) = 0 & \text{se } x \text{ for falsa} \end{cases}.$$

O valor- p obtido em cada contraste entre médias foi comparado com o nível nominal de significância α e o mesmo critério do teste original foi aplicado para decidir se o contraste poderia ser considerado significativo ou não.

2.3 Simulações

Foram consideradas $N = 1.000$ simulações de k tratamentos qualitativos e número de repetições r . Além disso, foram considerados diferentes números de erros padrão δ de diferenças entre médias consecutivas, já que o poder foi avaliado sob H_1 , em que as médias são todas diferentes, e sob H_0 parcial, em que algumas médias diferem entre si por δ erros padrão da média. O erro padrão da média é dado por

$$\sigma_{\bar{Y}} = \sqrt{\frac{\sigma^2}{r}}. \quad (4)$$

Os valores de k utilizados foram $k = 5; 10; 20$ e 80 tratamentos e valores de $r = 4; 10$ e 20 repetições. Essas combinações de valores tentam ilustrar o que acontece em situações reais, com poucos e muitos tratamentos e com números de repetições variados. Em todos os casos, o nível nominal de significância considerado foi de 5%.

Para cada combinação de k e r utilizou-se o delineamento inteiramente casualizado (DIC), considerando o modelo:

$$y_{ij} = \mu + \tau_i + e_{ij}, \quad (5)$$

em que y_{ij} é o valor observado para a variável resposta obtido para o i -ésimo tratamento em sua j -ésima repetição, μ é a constante associada a todas as observações, τ_i é o efeito do i -ésimo tratamento e e_{ij} erro aleatório associado a cada observação.

Foram considerados os modelos probabilísticos normal, lognormal e exponencial para descrever a variável aleatória do erro experimental e, assim, avaliar a robustez dos testes em situações adversas.

Sob H_0 completa, a média da normal foi considerada $\mu = 10$ sem perda de generalidade, bem como a variância $\sigma^2 = 1$. No modelo exponencial, o valor de λ foi 0,1 e, no lognormal, a média foi igual a 0 e a variância igual a 1, na escala logarítmica.

Além disso, ainda foram consideradas diferentes hipóteses sobre as médias. Sob H_0 completa, as médias dos tratamentos foram consideradas todas iguais. Dessa forma, para o modelo normal, o coeficiente de variação (CV) utilizado foi 10%. Uma ressalva deve ser feita em relação à escolha de apenas um valor de CV. De acordo com Silva e Ferreira e Bearzoti (1999) e Girardi e Cargnelutti Filho e Storck (2009), o CV não afetou as taxas de erro tipo I dos testes avaliados por esses autores, indicando que o esforço por parte de pesquisadores em diminuir o valor do CV não interfere no controle do erro tipo I dos testes de comparações múltiplas.

Para cada um dos valores obtidos das taxas de erro tipo I foram aplicados testes binomiais exatos considerando um nível nominal de significância de 0,01 para as hipóteses $H_0 : \alpha = 0,05$ e $H_1 : \alpha \neq 0,05$. Se a hipótese nula for rejeitada (valor- $p \leq 0,01$) e as taxas de erro tipo I observadas forem consideradas significativamente inferiores ao nível nominal de significância considerado, o teste é considerado conservador. Se a hipótese nula for rejeitada (valor- $p \leq 0,01$) e as taxas de erro tipo I observadas forem consideradas significativamente superiores ao nível nominal de significância, o teste é considerado liberal. Porém, se as taxas de erro tipo I observadas não forem consideradas significativamente diferentes do nível nominal de significância considerado (valor- $p > 0,01$), o teste é considerado exato, ou seja, controla o erro tipo I.

Sob a hipótese alternativa H_1 foram consideradas todas as k médias diferentes, mas a variância σ^2 foi constante. Entre duas médias de tratamentos consecutivos foi fixada a diferença de um erro padrão da média.

Sob a hipótese nula (H_0) parcial foram considerados dois grupos cujas médias são diferentes entre si por δ erros padrão, sendo $\delta = 2, 4, 8$ e 16. Dentro do grupo de médias iguais o mesmo procedimento para H_0 completa foi aplicado.

Os testes SNK original e *bootstrap* foram aplicados em todas os cenários simulados e as taxas de erro tipo I por experimento e poder foram estimadas em todos os N experimentos gerados. Sob as hipóteses nulas completa e parcial, as taxas de erro tipo I são obtidas e, sob as hipóteses parcial e alternativa, o poder é estimado.

3 Resultados e discussões

Na sequência são apresentados os resultados em relação ao erro tipo I e ao poder, em cenários formados por combinações de número de tratamentos, número de repetições em situações de normalidade e não normalidade dos resíduos, considerando o nível nominal de significância de 0,05, sob H_0 completa, sob H_1 e sob H_0 parcial.

3.1 Erro tipo I sob H_0 completa

Na Tabela 1 são apresentadas as taxas de erro tipo I por experimento (TPE) sob normalidade em função do número de tratamentos (k) e número de repetições (r) para H_0 completa dos testes SNK original (SNK) e sua versão *bootstrap* (SNK_B). Foram realizados testes binomiais exatos considerando um nível nominal de significância de 0,01 para cada taxa de erro tipo I observada.

Tabela 1 - Taxas de erro por experimento (TPE) dos testes SNK original (SNK) e SNK *bootstrap* (SNK_B) sob H_0 completa, considerando-se a distribuição normal (10, 1) e nível de significância 0,05, em função do número de tratamentos k e número de repetições r

k	r	SNK	SNK _B
5	4	0,045	0,043
	10	0,053	0,053
	20	0,052	0,051
10	4	0,052	0,051
	10	0,066	0,067
	20	0,054	0,055
20	4	0,047	0,050
	10	0,051	0,051
	20	0,057	0,056
80	4	0,047	0,053
	10	0,043	0,046
	20	0,050	0,048

Sob H_0 completa e normalidade dos resíduos, as taxas de erro tipo I obtidas pelos testes SNK e SNK_B não foram significativamente diferentes do nível nominal de significância adotado. Este resultado está de acordo com Machado et. al. (2005) que afirma que o teste SNK controla as taxas de erro tipo I por experimento sob H_0 completa. O mesmo resultado foi obtido por Ramos e Ferreira (2009) para os testes de Caliński e Corsten e sua versão *bootstrap*. Os resultados também se assemelham aos de Girardi e Cargnelutti Filho e Storck (2009) para o teste SNK, em que as taxas de erro tipo I por experimento não foram consideradas diferentes do nível nominal de significância, com exceção de alguns cenários que envolviam 3 e 5 tratamentos.

O comportamento dos testes aqui analisados, SNK e SNK_B , foram semelhantes, logo, o teste SNK seria mais indicado devido ao menor esforço computacional exigido. Porém, outras situações foram consideradas.

Na Tabela 2 são apresentadas as TPE sob H_0 completa considerando-se a distribuição lognormal com parâmetros de posição e escala iguais a 0 e 1, respectivamente, na escala logarítmica. Foram considerados diferentes números de tratamentos k , números de repetições r e nível de significância de 0,05.

Tabela 2 - Taxas de erro por experimento (TPE) dos testes SNK original (SNK) e SNK *bootstrap* (SNK_B) sob H_0 completa, considerando-se a distribuição lognormal (0, 1) e nível de significância 0,05, em função do número de tratamentos k e número de repetições r

k	r	SNK	SNK_B
	4	0,041	0,043
5	10	0,035 ⁺⁺	0,032 ⁺⁺
	20	0,040	0,041
	4	0,040	0,039
10	10	0,041	0,042
	20	0,044	0,039
	4	0,055	0,055
20	10	0,068 ^{**}	0,074 ^{**}
	20	0,062	0,065 ^{**}
	4	0,282 ^{**}	0,284 ^{**}
80	10	0,230 ^{**}	0,232 ^{**}
	20	0,176 ^{**}	0,177 ^{**}

⁺⁺ Significativamente diferente (valor- $p < 0,01$) e considerado menor do que o nível de significância $\alpha = 0,05$.

^{**} Significativamente diferente (valor- $p < 0,01$) e considerado maior do que o nível de significância $\alpha = 0,05$.

As TPE do teste SNK e SNK *bootstrap* com $k = 5$ e $r = 10$ são menores do que o nível de significância adotado e, para $k = 20$, $r = 10$ e $k = 80$ e todos os valores de r , são maiores do que o nível de significância adotado. Com 20 repetições e 20 tratamentos, o teste SNK_B apresentou taxa de erro tipo I maior que o nível significância, sendo considerado liberal. Este resultado confirma o que Borges e Ferreira (2003) obtiveram, considerando a distribuição lognormal, em que os testes de Tukey e SNK tenderam a apresentar taxas de erro por experimento bastante altas em situações de maior número de tratamentos.

Na Tabela 3 são apresentadas as TPE sob H_0 completa considerando a distribuição exponencial com parâmetro $\lambda = 0,1$. Foram considerados diferentes números de tratamentos k e número de repetições r .

Tabela 3 - Taxas de erro por experimento (TPE) dos testes SNK original (SNK) e SNK *bootstrap* (SNK_B) sob H_0 completa, considerando-se a distribuição exponencial (0, 1) e nível de significância 0,05, em função do número de tratamentos k e número de repetições r

k	r	SNK	SNK _B
5	4	0,035 ⁺⁺	0,035 ⁺⁺
	10	0,041	0,040
	20	0,045	0,048
10	4	0,046	0,047
	10	0,040	0,038
	20	0,055	0,054
20	4	0,058	0,056
	10	0,045	0,047
	20	0,046	0,043
80	4	0,071 ^{**}	0,071 ^{**}
	10	0,052	0,054
	20	0,062	0,064 ^{**}

⁺⁺ Significativamente diferente (valor- $p < 0,01$) e considerado menor do que o nível de significância $\alpha = 0,05$.

^{**} Significativamente diferente (valor- $p < 0,01$) e considerado maior do que o nível de significância $\alpha = 0,05$.

O teste SNK se mostrou conservador para 5 tratamentos e 4 repetições, apresentando taxa de erro tipo I por experimento significativamente inferior ao valor nominal de 0,05 e, para 80 tratamentos, o teste pode ser considerado liberal quando o número de repetições é igual a 4, apresentando taxa de erro tipo I por experimento significativamente maior que o valor nominal de 0,05. Nas demais situações, as taxas de erro tipo I por experimento não foram significativamente diferentes dos valores nominais de significância.

O teste de SNK_B, como o SNK, apresentou uma certa robustez no controle da taxa de erro tipo I por experimento sob H_0 completa, considerando-se a distribuição exponencial, em praticamente todos os casos analisados para 5, 10 e 20 tratamentos, sendo conservador apenas para $k = 5$ e $r = 4$ e liberal para $k = 80$, $r = 4$ e $r = 20$, com valores não muito distantes de 0,05. De maneira geral, o desempenho do teste SNK *bootstrap* foi semelhante ao do SNK original nas situações de não normalidade aqui consideradas e sob H_0 completa.

3.2 Erro tipo I sob H_0 parcial

Nas Tabelas 4 e 5 são apresentadas as taxas de erro tipo I por experimento em função do número de tratamentos k , números de repetições r e erros padrão δ de diferença entre as médias, sob H_0 parcial. Como, neste caso, havia duas situações

(dois grupos com médias que deveriam ser consideradas iguais dentro de cada grupo e médias que deveriam ser consideradas diferentes nas comparações entre os grupos), as taxas de erro tipo I foram mensuradas apenas nas comparações entre médias de um mesmo grupo. Assim, as diferenças de δ erros padrão são apenas consideradas para médias que se encontram em grupos diferentes e não dentro de um mesmo grupo. Foram realizados testes binomiais exatos considerando um nível nominal de significância de 0,01 para cada taxa de erro tipo I observada.

Tabela 4 - Taxas de erro tipo I dos testes SNK e SNK_B , em função do número de tratamentos ($k = 5$ e $k = 10$), número de repetições r , diferenças entre as médias δ , para o nível nominal de significância $\alpha = 0,05$, sob a distribuição normal e H_0 parcial

k	r	δ	SNK	SNK_B
5	4	2	0,067	0,351**
		4	0,091**	0,438**
		8	0,111**	0,433**
		16	0,116**	0,433**
	10	2	0,059	0,379**
		4	0,113**	0,479**
		8	0,088**	0,429**
		16	0,104**	0,432**
	20	2	0,047	0,373**
		4	0,105**	0,460**
		8	0,109**	0,456**
		16	0,101**	0,463**
10	4	2	0,039	0,749**
		4	0,082**	0,895**
		8	0,101**	0,878**
		16	0,101**	0,870**
	10	2	0,031 ⁺⁺	0,779**
		4	0,118**	0,893**
		8	0,100**	0,899**
		16	0,100**	0,904**
	20	2	0,031 ⁺⁺	0,809**
		4	0,106**	0,908**
		8	0,114**	0,910**
		16	0,108**	0,915**

⁺⁺ Significativamente diferente (valor- $p < 0,01$) e considerado menor do que o nível de significância $\alpha = 0,05$.

** Significativamente diferente (valor- $p < 0,01$) e considerado maior do que o nível de significância $\alpha = 0,05$.

Tabela 5 - Taxas de erro tipo I dos testes SNK e SNK_B , em função do número de tratamentos ($k = 20$ e $k = 80$), número de repetições r , diferenças entre as médias δ , para o nível nominal de significância $\alpha = 0,05$, sob a distribuição normal e H_0 parcial

k	r	δ	SNK	SNK_B
20	4	2	0,043	0,969**
		4	0,093**	0,997**
		8	0,100**	0,995**
		16	0,103**	0,999**
	10	2	0,048	0,982**
		4	0,105**	1,000**
		8	0,100**	1,000**
		16	0,086**	0,997**
	20	2	0,049	0,974**
		4	0,095**	0,999**
		8	0,089**	0,997**
		16	0,095**	0,998**
80	4	2	0,001 ⁺⁺	1,000**
		4	0,026 ⁺⁺	1,000**
		8	0,101**	1,000**
		16	0,090**	1,000**
	10	2	0,001 ⁺⁺	1,000**
		4	0,086**	1,000**
		8	0,089**	1,000**
		16	0,117**	1,000**
	20	2	0,035	1,000**
		4	0,089**	1,000**
		8	0,090**	1,000**
		16	0,100**	1,000**

⁺⁺ Significativamente diferente (valor- $p < 0,01$) e considerado menor do que o nível de significância $\alpha = 0,05$.

** Significativamente diferente (valor- $p < 0,01$) e considerado maior do que o nível de significância $\alpha = 0,05$.

O teste SNK_B foi considerado liberal em todas os cenários avaliados apresentando altas taxas de erro tipo I por experimento e essa taxa aumenta com o aumento do número de tratamentos k . Para o teste SNK pode-se observar, fixando-se $\delta = 2$, $k = 5$ e $k = 20$, que as TPE não são significativamente diferentes do nível nominal de significância, para os cenários $k = 10, r = 4$ e $k = 80, r = 20$, fixando $\delta = 2$ o teste também foi considerado exato, nos demais cenários em que $\delta = 2$ o teste SNK foi conservador. Para valores de δ diferentes de 2, as taxas de erro tipo I foram maiores que α , sendo então o teste SNK um teste liberal nesses cenários.

3.3 Poder sob H_1

Na Figura 1 são apresentados os gráficos de poder dos testes SNK e SNK_B , em função da diferença δ em erros padrão entre médias, para diferentes números de repetições (r) e tratamentos (k), considerando a distribuição normal sob H_1 e nível de significância de 0,05.

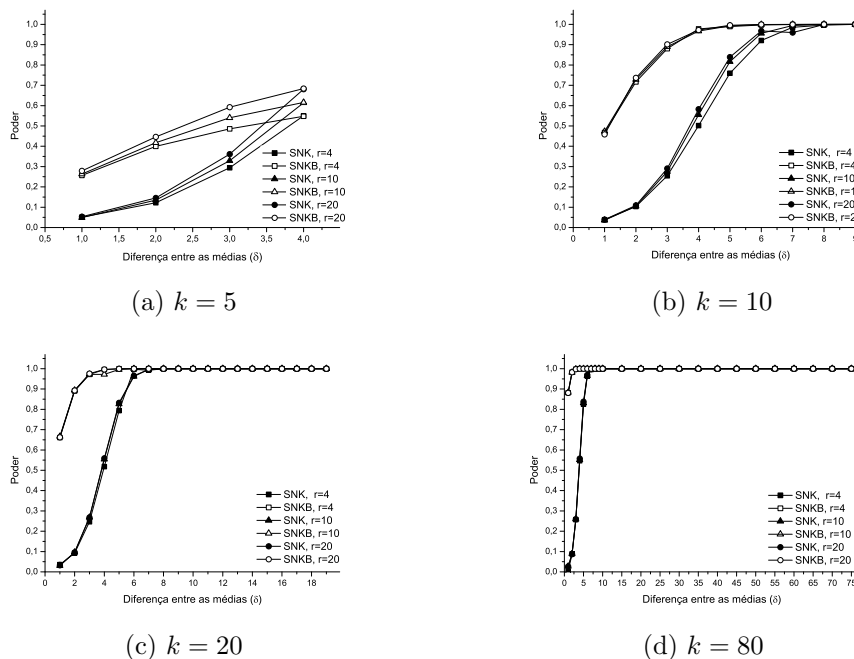


Figura 1 - Poder dos testes de Student-Newman-Keuls (SNK) e sua versão *bootstrap* (SNK_B), em função das diferenças entre médias δ , diferentes números de repetições r e diferentes números de tratamentos k , considerando-se a distribuição normal (10,1) sob H_1 e $\alpha = 0,05$.

Pode-se observar, comparando-se as Figuras 1a, 1b, 1c e 1d, que há um aumento no valor do poder à medida que o valor de k aumenta. Ao se considerar o efeito do número de repetições nos valores do poder dos testes, observou-se que o efeito de repetição só é evidente no caso de $k = 5$ e, para os demais casos, praticamente não houve diferenças entre esses valores.

Como já é preconizado pela teoria, o poder de um teste aumenta com o aumento de δ . Essa mesma tendência, porém em diferentes magnitudes, pode ser verificada em diversos trabalhos que avaliaram vários testes, como Ramos e Ferreira (2009), Girard e Cargnelutti Filho e Storck (2009), Conagin e Barbin e Demétrio (2008) e Silva e Ferreira e Bearzoti (1999). Quando as diferenças entre as médias são menores ($\delta \leq 4$), o poder do teste SNK_B é maior do que o do teste SNK que,

por sua vez, apresenta valores de poder abaixo de 0,05 quando $\delta = 1$ para todos os valores de k .

Ao se comparar tais resultados com os obtidos por Ramos e Ferreira (2009), pode-se verificar o mesmo padrão de comportamento do poder do teste de Caliński e Corsten (C) e sua versão *bootstrap* (CB), ocorrendo um aumento do poder à medida que aumentam os valores de r , k e δ . Porém, em quase todos os casos, praticamente não houve diferenças entre os valores de poder dos testes C e CB, enquanto que, nos testes SNK e SNK_B , essa diferença é evidente quando os valores de δ são pequenos.

Na Figura 2 estão apresentados os gráficos de poder dos testes SNK e SNK_B , em função da diferença δ em erros padrão entre médias, com diferentes números de repetições (r) e tratamentos (k), considerando a distribuição lognormal, com parâmetros de posição e escala iguais a 0 e 1, sob H_1 e nível de significância de 0,05.

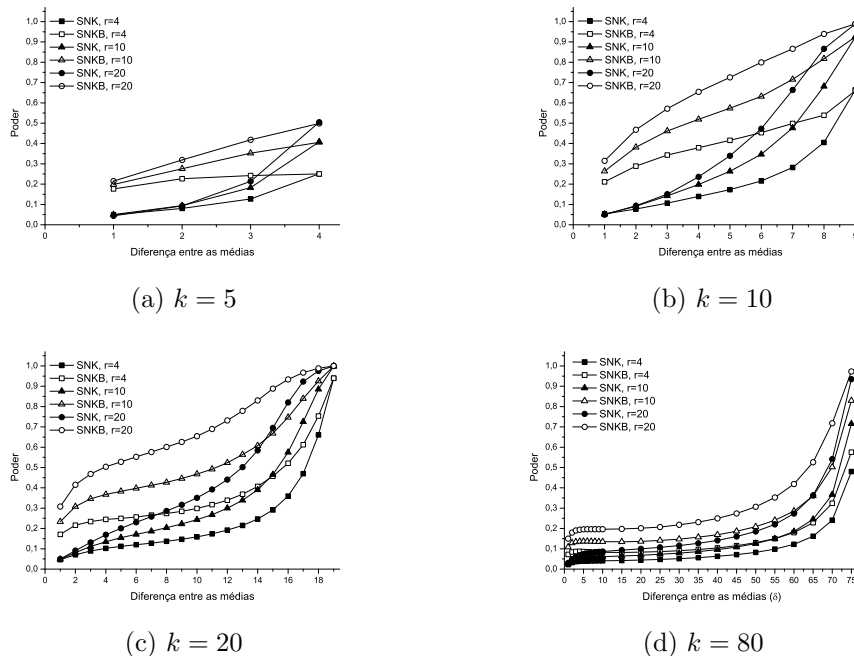


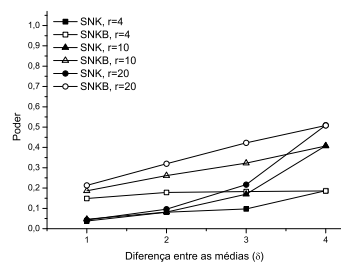
Figura 2 - Poder dos testes de Student-Newman-Keuls (SNK) e sua versão *bootstrap* (SNK_B), em função das diferenças entre médias δ , em função das diferenças entre médias δ e diferentes números de tratamentos k , considerando-se a distribuição lognormal (0,1), para, sob H_1 e $\alpha = 0,05$.

Ambos os testes apresentaram comportamento do poder diferente do apresentado sob distribuição normal. O número de repetições (r) afetou o poder dos testes indicando que, quanto maior o r , maior o poder. Além disso, para cada valor de r , o teste SNK_B apresentou maiores valores de poder do que o SNK. Ramos

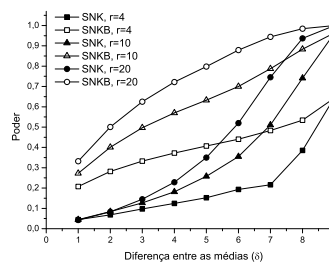
e Ferreira (2009) também observaram que houve efeito do número de repetições para todos os valores de k . Segundo os autores, isso se deve à heterogeneidade de variâncias. Como as médias e a variância não são independentes nas distribuições não normais, ao se determinar os parâmetros da distribuição para gerar outro tratamento, essas são alteradas. Portanto, quanto maior o número de repetições, menor a heterogeneidade de variâncias e maior o poder.

O poder do teste SNK_B , como na distribuição normal, é maior do que o poder do teste SNK e, à medida que a diferença entre as médias aumenta, os valores de poder de ambos os testes tendem a ser iguais. Como era de se esperar, à medida que a diferença entre médias consecutivas aumenta, a porcentagem de decisões corretas cresce rapidamente para valores de k maiores que 5. Para $k = 5$, o poder aumenta de forma mais lenta. Para valores de k iguais a 10 e 20, os valores do poder são mais altos mesmo para valores de δ pequenos.

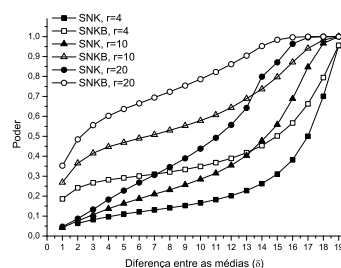
Na Figura 3 estão apresentados os gráficos de poder dos testes SNK e SNK_B , em função da diferença em erros padrão entre as médias (δ), com diferentes números de repetições (r) e tratamentos (k), considerando a distribuição exponencial, com parâmetro $\lambda = 0, 1$, sob H_1 e nível de significância de 0,05.



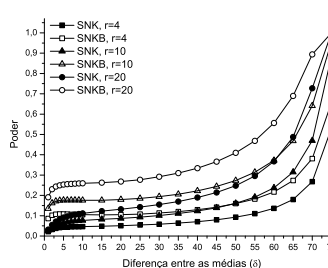
(a) $k = 5$



(b) $k = 10$



(c) $k = 20$



(d) $k = 80$

Figura 3 - Poder dos testes de Student-Newman-Keuls (SNK) e sua versão *bootstrap* (SNK_B), em função das diferenças entre médias δ , diferentes números de repetições r e para diferentes números de tratamentos k , considerando-se a distribuição exponencial, sob H_1 e $\alpha = 0,05$.

Observa-se na Figura 3 que os resultados obtidos foram semelhantes aos resultados obtidos na distribuição lognormal, porém, de uma forma geral, os valores de poder na exponencial foram um pouco mais altos. O aumento do poder dos testes também foi bastante influenciado pelo aumento no número de repetições e, para o teste SNK original, isso acontece principalmente quando a diferença entre médias passa a ser maior do que 2.

Como era de se esperar, à medida que a diferença entre médias aumenta, o poder dos testes também aumenta. O poder do teste SNK_B é superior ao poder do teste SNK original em todas as situações analisadas, porém, a taxa de aumento do poder no teste SNK é maior do que no teste SNK_B .

O comportamento sob H_1 foi o mesmo para as duas distribuições não normais porém, o poder de ambos os testes apresentaram valores menores para a distribuição lognormal para todos os valores de δ analisados. Para as duas distribuições verifica-se a mesma tendência observada na distribuição normal considerando a diferença entre as médias e número de tratamentos.

3.4 Poder sob H_0 parcial

Na Figura 4 estão apresentados os valores de poder dos testes SNK e SNK_B em função da diferença de δ erros padrão, número de repetições r e número de tratamentos k , sob normalidade e H_0 parcial e nível de significância de 0,05.

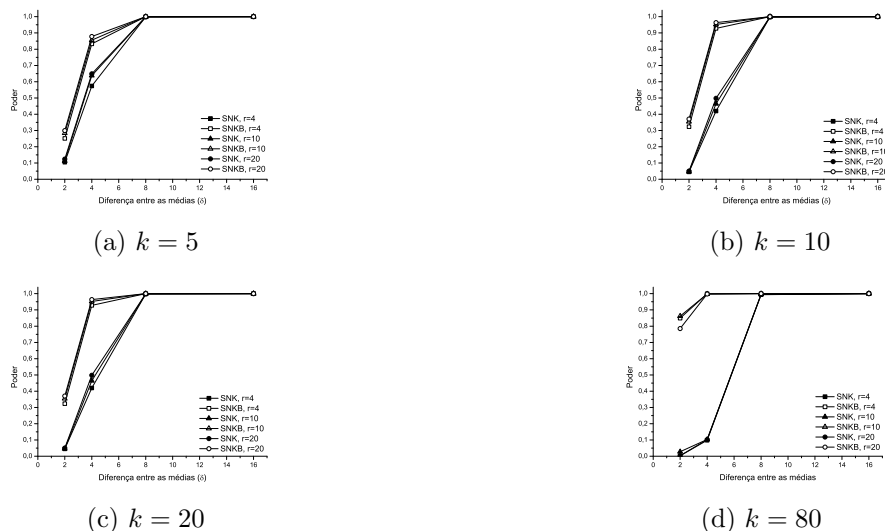


Figura 4 - Poder dos testes de Student-Newman-Keuls (SNK) e sua versão *bootstrap* (SNK_B), em função das diferenças entre médias δ , diferentes números de repetições r e para diferentes números de tratamentos k , considerando-se a distribuição normal, sob H_0 parcial e $\alpha = 0,05$.

Para valores de δ menores do que 4, os valores de poder do teste SNK foram bem menores do que os do SNK_B . Quando $\delta \geq 8$, os valores de poder se igualam. Ou seja, na situação de H_0 parcial, se as diferenças entre os grupos forem pequenas, o teste SNK_B discrimina melhor as diferenças do que o teste SNK.

Observa-se um aumento de poder à medida que a diferença δ entre as médias se torna maior. O mesmo comportamento foi observado por Ramos e Ferreira (2009) e Ramos e Vieira (2014).

3.5 Aplicação

Para avaliar o desempenho dos testes descritos neste trabalho foram utilizados dados reais de controles químico e mecânico de *Cerconota anonella* (Sepp.) (Lepidoptera: Oecophoridae) e de *Bephratelloides pomorum* (Fab.) (Hymenoptera: Eurytomidae), principais pragas da gravioleira, do trabalho realizado por Micheletti et al. (2001). O delineamento experimental utilizado foi o inteiramente casualizado, com 9 tratamentos e 20 repetições. Foi avaliada a variável peso e os tratamentos utilizados foram: 1- frutos sem proteção (testemunha); 2- saco de papel kraft (43cm de comprimento x 24cm de largura); 3- saco plástico comum fechado na extremidade inferior, apenas com alguns orifícios feitos para permitir o escoamento da água (49cm de comprimento x 28cm de largura); 4- saco plástico comum aberto na extremidade inferior (49cm de comprimento x 28cm de largura); 5- saco plástico perfurado (49cm de comprimento x 28cm de largura); 6- saco de papel impermeável nas duas faces (45cm de comprimento x 32cm de largura); 7- pulverização dos frutos semanal e localizada, com triflumuron 250g/kg, sendo a dosagem 100g/ha; 8- pulverizações com triflumuron 250g/kg, na dosagem 100g/ha, sendo posteriormente os frutos ensacados com saco plástico perfurado; 9- pulverização dos frutos semanal e localizada, com imidacloprid 700g/kg, sendo a dosagem 100g/ha. Os tratamentos foram aplicados em frutos de 4 a 6 cm de comprimento, repetidos 20 vezes, em plantas diferentes.

Os sacos foram presos aos ramos acima dos frutos, por meio de arame plastificado. No caso do tratamento com os inseticidas, considerou-se o intervalo de segurança de 10 dias que antecedem a colheita, para suspender a aplicação dos produtos em campo. Na ocasião da colheita, foram avaliadas o número de orifícios causados por *C. anonella* e por *B. pomorum*.

A característica peso dos frutos apresentou a média geral de 1,217 kg e coeficiente de variação (CV) igual a 41,21%. Na Tabela 6 são apresentados os pesos médios dos frutos de graviola, médias das 20 repetições de acordo com os respectivos tratamentos, ordenadas de forma decrescente.

A esses dados foram aplicadas as duas versões do teste de SNK, original e *bootstrap*, sendo adotado o nível nominal de 0,05 de significância, com o objetivo de ilustrar a aplicação de ambos os testes.

Para os dados apresentados na Tabela 6 foram obtidas as médias dos tratamentos, em que foi avaliada a variável peso de frutos de graviola em kg.

Tabela 6 - Peso médio de frutos colhidos de graviola, em kg, ordenados de forma decrescente. Sítio Aldeia Verde, Maceió-AL, outubro/1999 a fevereiro/2000

Tratamento	Peso
Saco de papel kraft	1,49
Saco de papel impermeável	1,40
Saco plástico aberto	1,37
Triflumuron + Saco plástico perfurado	1,37
Saco plástico fechado	1,36
Saco plástico perfurado	1,27
Testemunha	0,99
Imidacloprid	0,94
Triflumuron	0,76

Fonte: Micheletti et al. (2001).

As médias estão organizadas em ordem decrescente de magnitude: $\bar{Y}_{(1)} = 1,49$; $\bar{Y}_{(2)} = 1,40$; $\bar{Y}_{(3)} = 1,37$; $\bar{Y}_{(4)} = 1,37$; $\bar{Y}_{(5)} = 1,36$; $\bar{Y}_{(6)} = 1,27$; $\bar{Y}_{(7)} = 0,99$; $\bar{Y}_{(8)} = 0,94$; $\bar{Y}_{(9)} = 0,76$.

Na Tabela 7 estão apresentados os resultados obtidos pelos testes SNK e SNK_B .

Tabela 7 - Peso de frutos colhidos de graviola em kg. Médias de 20 repetições. Sítio Aldeia Verde, Maceió-AL, outubro/1999 a fevereiro 2000

Tratamento	Peso	SNK	SNK_B
1 - Saco de papel kraft	1,49	a	a
2 - Saco de papel impermeável	1,40	ab	a
3 - Saco plástico aberto	1,37	ab	a
4 - Triflumuron + Saco plástico perfurado	1,37	ab	a
5 - Saco plástico fechado	1,36	ab	a
6 - Saco plástico perfurado	1,27	ab	a
7 - Testemunha	0,99	bc	b
8 - Imidacloprid	1,94	bc	b
9 -Triflumuron	0,76	c	c

*Médias seguidas de mesma letra na coluna não diferem entre si ao nível de 5% de significância de acordo com os testes SNK e SNK_B .

Pelo teste SNK_B , o tratamento Triflumuron diferiu estatisticamente dos demais e, pelo teste SNK, o mesmo tratamento não diferiu dos testes Imidacloprid e testemunha. Os tratamentos saco plástico perfurado, saco plástico fechado, Triflumuron + saco plástico perfurado, saco plástico aberto, saco de papel

impermeável e de papel kraft não diferiram entre si pelos dois testes. De acordo com o teste SNK, a testemunha diferiu apenas dos tratamentos saco de papel kraft e Triflumuron e, pelo teste SNK_B , a testemunha diferiu de todos os tratamentos, exceto o Imidacloprid.

Nessa aplicação, pode-se observar que o teste SNK_B identificou mais diferenças significativas do que o teste SNK e isso se deve provavelmente a sua característica mais liberal. Os tempos computacionais gastos para aplicar os testes SNK e SNK_B para os dados da aplicação foram, respectivamente, 0,05s e 4,83s. Isso se deve ao fato que o teste SNK_B deve efetuar $B = 1.000$ reamostragens para retornar o resultado.

Conclusões

O desempenho do teste proposto foi considerado superior ao do teste SNK original sob H_0 completa e H_1 sob normalidade. Os testes SNK e SNK_B são exatos sob H_0 completa e normalidade. Sob H_0 completa e não normalidade, os testes SNK e SNK_B controlam as taxas de erro tipo I por experimento e são considerados exatos na maior parte dos casos simulados para $k = 5$ e $k = 10$, enquanto que, para $k = 20$ e $k = 80$, ambos os testes em alguns cenários são considerados liberais. Sob H_0 parcial, o teste SNK_B foi liberal em todos os casos simulados, enquanto que o teste SNK foi, em geral, conservador para $\delta \leq 2$ e liberal para os demais valores de δ . O poder do teste SNK_B é, na maioria das situações, superior ao do teste SNK sob H_1 e sob H_0 parcial. Os dois testes apresentaram certa robustez. Assim, em situações práticas, se as diferenças entre as médias dos tratamentos forem pequenas ($\delta \leq 2$), o teste SNK é mais indicado por controlar o erro tipo I e apresentar valores de poder satisfatórios. Nas demais situações, o teste SNK_B é mais recomendado, apesar de ambos serem liberais para $\delta \geq 4$, se a situação for de H_0 parcial.

Agradecimentos

Os autores agradecem Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro e aos pareceristas pelas sugestões que contribuíram muito para a melhoria do trabalho.

GONÇALVES, B. O.; RAMOS, P. S.; AVELAR, F. G. Test Student-Newman-Keuls bootstrap: proposal, evaluation and application productivity data of soursop. *Rev. Bras. Biom.*, São Paulo, v.33, n.4, p.445-470, 2015.

- **ABSTRACT:** *The aim of this work was to propose a modification of the SNK test for multiple comparisons using bootstrap resampling (SNK_B) and evaluate the performance of the two versions of the test. The performance was evaluated by experimentwise error rates and power using a Monte Carlo simulation study considering normal and non-normal residuals. Both tests were exact under H_0 and normality.*

Under H_0 and non-normality, the tests controlled experimentwise error rates and, therefore, they are considered exact in most simulated cases. Under H_0 partial, the SNK test B was liberal in all simulated scenarios, while the SNK test was conservative, in most situations, and liberal in some cases. In general, as the differences between the means increased the power also increased under partial H_0 and H_1 . Furthermore, the tests were applied to a real experiment designed to evaluate the chemical and mechanical controls of soursop pests in order to compare the results of both tests.

■ **KEYWORDS:** Multiple comparisons; resampling; Monte Carlo simulation; experiment-wise error rates; power.

Referências

- BIASE, N. G.; FERREIRA, D. F. Testes de igualdade e de comparações múltiplas para várias proporções binomiais independentes. *Revista Brasileira de Biometria*, v.29, n.4, p. 549–570, 2011.
- BOLFARINE, H.; SANDOVAL, M. C. *Introdução à inferência estatística*. Rio de Janeiro: Sociedade Brasileira de Matemática, 2001. 123p.
- BORGES, L. C.; FERREIRA, D. F. Poder e taxas de erro tipo I dos testes scott-knott, tukey e student-newman-keuls sob distribuições normal e não normais dos resíduos. *Revista de Matemática e Estatística*, v.21, p.67–83, 2003.
- CARMER, S.; SWANSON, N. R. An evaluation of ten pairwise multiple comparison procedures by monte carlo methods. *Journal of the American Statistical Association*, v.68, n.341, p.6–74, 1973.
- CALIŃSKI, T.; CORSTEN, L. C. A. Clustering means in anova by simultaneous testing. *Biometrics*, v.41, n.1, p.39–48, 1985.
- CARPENTER, J.; BITHELL, J. Bootstrap confidence intervals: When, which, what? A practical guide for medical statistician. *Statistics in Medicine*, v.19, p.1141–1164, 2000.
- CONAGIN, A.; BARBIN, D.; DEMETRIO, C. G. B. Modifications for the tukey test procedure and evaluation of the power and efficiency of multiple comparison procedures. *Sci. Agric.*, v.65, n.4, p.428–432, 2008.
- FERREIRA, E. B.; CAVALCANTI, P. P.; NOGUEIRA, D. A. ExpDes: An R Package for ANOVA and Experimental Designs. *Applied Mathematics*, v.5, p.2952–2958, 2014.
- FREUND, J. E. *Estatística aplicada: economia, administração e contabilidade*. 11. ed. Porto Alegre: Bookman, 2006. 536p.
- GIRARDI, L. H.; CARGNELUTTI FILHO, A.; STORCK, F. Erro tipo I e poder de cinco testes de comparações múltipla de médias. *Revista Brasileira de Biometria*, v.27, n.1, p.23–36, 2009.

- KEULS, M. The use of the studentized range in connection with an analysis of variance. *Euphytica*, v.1, p.112–122, 1952.
- KRISHNAMOORTHY, K.; LU, F; MATHEW, T. A parametric bootstrap approach for {ANOVA} with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis*, v.51, n.12, p.5731 - 5742, 2007.
- MACHADO, A. A. et al. Estatística experimental: uma abordagem fundamentada no planejamento e no uso de recursos computacionais. In: IN: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11., 2005, Londrina. *Anais...* Londrina: ISBN, 2005. 290p.
- MANLY, B. F. J. *Randomization, bootstrap and Monte Carlo methods in biology*. 2.ed. London: Chapman and Hall, 1998. 399p.
- MICHELETTI, S. M. F. B. et al. Controle de cerconota anonella (sepp.) (lep.: Oecophoridae) e de bephratelloides pomorum (fab.) (hym.: Eurytomidae) em frutos de graviola (annona muricata l.). *Revista Brasileira de Fruticultura*, v.23, n.3, 2001.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the theory of statistics*. 3. ed. New York: McGraw Hill, 1974. 577p.
- NEWMAN, D. The distribution of range in samples from a normal population, expressed in terms of independent estimate of a standard deviation. *Biometrika*, v.31, p.20–30, 1939.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>.
- RAMOS, P. S.; FERREIRA, D. F. Agrupamento de médias via bootstrap de populações normais e não-normais. *Revista Ceres*, v.56, n.1, p.140–149, 2009.
- RAMOS, P. S.; VIEIRA, M. T. Bootstrap multiple comparison procedure based on the f distribution. *Revista Brasileira de Biometria*, v.31, n.4, p.529–546, 2014.
- SILVA, E. C.; FERREIRA, D. F.; BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de monte carlo. *Ciência Agrotécnica*, v.23, p.687–696, 1999.
- SOUSA, V. A.; LIMA JUNIOR, M. A.; FERREIRA, L. R. C. Avaliação de testes estatísticos de comparações múltiplas de médias. *Revista Ceres*, v.59, n.3, p.350–354, 2012.
- STEEL, R. G. D.; TORRIE, J. H. *Principles and procedures of statistics*. 2. ed. New York: McGraw-Hill Book, 1980.

Recebido em 02.03.2015.

Aprovado após revisão em 05.09.2015.

ANEXO

Código R para aplicação dos testes SNK e SNK_B . Há exemplos de dados simulados e de dados reais.

```
library(gdata) #uso da função upperTriangle
library(ExpDes) #algumas funções deste pacote são utilizadas

# simular dados experimentais com média mu e variância comum e armazenar
# em uma matriz de 2 colunas com a primeira identificando os k tratamentos
# e segunda com os valores da variável resposta
# dois grupos de tratamentos consecutivos
simula.dados <- function(mu,sig2,k,r,kd=1){
  x <- matrix(0,r*k,2)
  i_ini <- 1
  mu0 <- mu
  k1 <- k %/% 2
  ind <- 0
  for (jj in 1:k){
    i_fin <- i_ini+r-1
    if ((jj>k1) & (ind==0))
    {
      mu0 <- mu0 + kd*(sig2/r)^0.5
      ind <- 1
    }
    x[i_ini:i_fin,2] <- rnorm(r, mu0, sig2^0.5)
    x[i_ini:i_fin,1] <- jj
    i_ini <- jj*r+1
  }
  list(trt=x[,1],y=x[,2])
}

#simular dados experimentais com média mu
#e reamostrar parametricamente
bootstrap <- function(mu,sig2,k,r){
  x <- matrix(0,r*k,2)
  i_ini <- 1
  mu0 <- mu
  for (jj in 1:k){
    i_fin <- i_ini+r-1
    x[i_ini:i_fin,2] <- rnorm(r, mu0, sig2^0.5)
    x[i_ini:i_fin,1] <- jj
    i_ini <- jj*r+1
  }
  list(trt=x[,1],y=x[,2])
}
```

```

#função que retorna os valores de S2, G.L. e o valor necessário para
#o cálculo do valor-p da amplitude estudentizada
valores.anava = function(dados,r){
  anava <- anova.dados(dados)
  S2 <- anava$"Mean Sq"[2]
  f <- anava$"Df"[2]
  raiz <- sqrt(S2/r)
  list(S2=S2,f=f,raiz=raiz)
}

#retorna um objeto com a análise de variância dos dados simulados
anova.dados <- function(dados){
  dados$trt <- as.factor(dados$trt)
  anovadados <- lm(dados$y ~ dados$trt)
  anovadad <- anova(anovadados)
  anovadad
}

#calcula médias dos tratamentos
anova.medias <- function(dados){
  medias <- tapply(dados$y,list(dados$trt),mean)
}

#função para calcular os valores-p de cada contraste
#as médias ordenadas devem ser passadas para a função
#obtem todos os valores-p (todos os contrastes)
valoresp <- function(medias,k,r,raiz,nu){
  valorp <- matrix(1,k,k)
  for(i in 1:(k-1)){
    for(j in k:(i+1)){
      contr <- abs(medias[i] - medias[j])
      R <- contr/raiz
      valorp[i,j] <- 1 - ptukey(R,(j-i+1),nu)
    }
  }
  valorp
}

#####BOOTSTRAP - obtenção da matriz quihb
qihb <- function(mu1,S21,k,r,B){
  comp <- sum(seq(k-1,1))
  q <- matrix(0,comp,B)
  for(b in 1:B)
  {
    dados.b <- bootstrap(mu1,S21,k,r)
    valores.b <- valores.anava(dados.b,r)
    medias.b <- anova.medias(dados.b)
    med.ord.b <- sort(medias.b)
    qib <- qi.f(med.ord.b,k,r,valores.b$raiz)
    qib.up <- upperTriangle(qib)
    #essa função toma os elementos da triangular superior por coluna, ou seja,
    #1-2, 1-3, 2-3, 1-4, 2-4, 3-4, 1-5, 2-5, 3-5, 4-5, nessa ordem
    q[,b] <- qib.up
  }
  q
}

```

```

#cálculo dos valores de qi para as amostras bootstrap
#q = c/raiz(QME/r), os dados ordenados de forma decrescentes
#são passados
qi.f <- function(dad,k,r,raiz.b){
  q <- matrix(1,k,k)
  for(i in 1:(k-1)){
    for(j in k:(i+1)){
      contr <- abs(dad[i] - dad[j])
      q[i,j] <- contr/raiz.b
    }
  }
  q
}

#função para calcular os valores-p bootstrap de cada contraste
#as médias ordenadas devem ser passadas para a função
#porque a estatística do teste é feita com médias em ordem
#obtem todos os valores-p (matriz kxk em que usamos a triangular
#superior)
valorespb <- function(medias,k,r,raiz,qihbb,B){
  valorpb <- matrix(1,k,k)
  cont <- 1 #contador para as linhas da matriz qihbb
  for(i in 2:k){
    for(j in 1:(i-1)){
      contr <- abs(medias[j] - medias[i])
      R <- contr/raiz
      valorpb[j,i] <- (length(qihbb[cont,R<qihbb[cont,]]))/B
      #dividir pelo n° de amostras bootstrap
      cont <- cont + 1
    }
  }
  valorpb
}

#função para executar as funções para simular os dados
#e retornar os tratamentos e valores da variável resposta
simula <- function(k,r,mu,sig2,alpha,nsim,B = 1000,delta){
  dados <- simula.dados(mu,sig2,k,r,delta)
  dados.d <- data.frame(dados$trt,dados$y)
  dados.d
}

#função para transformar o objeto com os tratamentos e médias
#já fornecidos em um objeto igual ao resultante de simula.dados
trans.fator <- function(dados){
  trt <- dados[,1]
  y <- dados[,2]
  list(trt=trt,y=y)
}

```

```

#função que recebe os dados e retorna
#as médias e os valores-p dos métodos SNK e SNKB
aplica <- function(dados,B = 1000,alpha = 0.05){
  dados <- trans.fator(dados)
  r <- table(dados$trt)[[1]]
  k <- length(unique(dados$trt))
  trats <- unique(dados$trt)
  valores <- valores.anava(dados,r)
  medias <- anova.medias(dados)
  med.ord <- sort(medias,decreasing=TRUE)
  nomes <- as.numeric(rownames(med.ord))
  valoresp.m <- valoresp(med.ord,k,r,valores$raiz,valores$f)
  elem <- upperTriangle(valoresp.m)
  valoresp.m <- diag(k)
  valoresp.m[upper.tri(valoresp.m, diag=F)] <- elem
  valoresp.m <- valoresp.m + t(valoresp.m) - diag(diag(valoresp.m))
  mu1 <- mean(dados$y)
  S21 <- var(dados$y)
  qihbb <- qihb(mu1,S21,k,r,B)
  valoresp.b <- valorespb(med.ord,k,r,valores$raiz,qihbb,B)
  elemb <- upperTriangle(valoresp.b)
  valoresp.b <- diag(k)
  valoresp.b[upper.tri(valoresp.b, diag=F)] <- elemb
  valoresp.b <- valoresp.b + t(valoresp.b) - diag(diag(valoresp.b))
  cat("\nTeste de Student-Newman-Keuls (SNK)")
  cat("\nGrupos Tratamentos Médias\n")
  saida <- order.SNK(trats,medias,valoresp.m,alpha)
  saida
  cat("-----\n")
  cat("\nTeste de Student-Newman-Keuls bootstrap (SNKB)")
  cat("\nGrupos Tratamentos Médias\n")
  saida <- order.SNK(trats,medias,valoresp.b,alpha)
  saida
  cat("-----\n")
}

```

```

#função inspirada na order.stat.SNK do pacote ExpDes,
#conforme explicado na Seção 2
order.SNK <- function(tratamentos,medias,valorp,alpha){
  n <- length(medias)
  z <- data.frame(tratamentos,medias)
  w <- z[order(z[, 2], decreasing = TRUE), ]
  M <- rep("", n)
  k <- 1
  k1 <- 0
  j <- 1
  i <- 1
  r <- 1
  cambio <- n
  cambio1 <- 0
  chequeo <- 0
  M[1] <- letters[k]
  while (j < n) {
    chequeo <- chequeo + 1
    if (chequeo > n)
      break
  }
}

```



```

for (i in j:n) {
  if (abs(j - i) == 0) {
    r <- 1
  }
  else {
    r <- abs(j - i)
  }
  s <- valorp[i,j] > alpha
  if (s) {
    if (lastC(M[i]) != letters[k])
      M[i] <- paste(M[i], letters[k], sep = "")
  }
  else {
    k <- k + 1
    cambio <- i
    cambio1 <- 0
    ja <- j
    for (jj in cambio:n) M[jj] <- paste(M[jj], " ",
                                       sep = "")
    M[cambio] <- paste(M[cambio], letters[k], sep = "")
    for (v in ja:cambio) {
      if (abs(v - cambio) == 0) {
        r <- 1
      }
      else {
        r <- abs(v - cambio)
      }
      if (valorp[v,cambio] <= alpha) {
        j <- j + 1
        cambio1 <- 1
      }
      else break
    }
    break
  }
}
if (cambio1 == 0)
  j <- j + 1
}
w <- data.frame(w, stat = M)
trt <- as.character(w$stratamentos)
means <- as.numeric(w$medias)
for (i in 1:n) {
  cat(M[i], "\t", trt[i], "\t", means[i], "\n")
}
output <- data.frame(trt, means, M)
return(output)
}

```

```

#Exemplo de dados simulados de uma H0 parcial:
#mu = 10, sig2 = 1, alpha = 0.05, B = 1000 reamostragens bootstrap
#k = 6 tratamentos, r = 4 repetições, delta = 8 de um DIC
#espera-se que dois grupos sejam formados, se o valor
#de delta for baixo, a diferença entre os dois grupos será pequena
#e os testes podem não encontrar as diferenças
#Para simular de uma H0 completa basta definir delta = 0
mu <- 10; sig2 <- 1; alpha <- 0.05
k <- 6; r <- 4; delta <- 8; B <- 1000
dadossim <- simula(k,r,mu,sig2,alpha,nsim,B,delta)
aplica(dadossim)

#Conjunto de dados real
#Ex.: conjunto de dados de Rhizobium trifolii (k=6, r=5)
#Fonte: STEEL; TORRIE, p. 141, 1980.
#dados - inserir valores da variavel resposta colocando na ordem
#todas as repetições de cada tratamento ou utilizar um arquivo externo
#no arquivo externo a primeira coluna deve ser a de tratamentos
# e a segunda deve conter os valores da variável resposta
#número de reamostragens bootstrap
k <- 6; r <- 5
(rhizobium <- matrix(c(rep(seq(1:k),each=r),
                      19.4,32.6,27,32.1,33,17.7,24.8,27.9,25.2,24.3,
                      17,19.4,9.1,11.9,15.8,20.7,21,20.5,18.8,18.6,
                      14.3,14.4,11.8,11.6,14.2,17.3,19.4,19.1,16.9,
                      20.8),r*k,2))
#aplicação dos testesvSNK e SNKB
aplica(rhizobium)

#exemplo retirado do pacote ExpDes
data(ex1)
ex1
attach(ex1)
crd(trat, ig, quali = T, sigF = 0.05)

#aplicação dos testesvSNK e SNKB
aplica(ex1)

```