

Versão *bootstrap* paramétrico do teste de Student-Newman-Keuls

Bruna de Oliveira Gonçalves^{1 2}

Patrícia de Siqueira Ramos³

1 Introdução

Comumente, ao se analisar um conjunto de dados, o pesquisador precisa decidir se, em média, os tratamentos aplicados produziram resultados iguais ou, em caso contrário, qual tratamento produziu melhores resultados. Uma análise de variância visa fundamentalmente verificar se existe uma diferença significativa entre as médias dos tratamentos, mas não indica quais médias são diferentes entre si. Para isso, existem os testes de comparações múltiplas (MONTGOMERY, 2001).

Segundo Machado et al. (2005), há muitos testes de comparações múltiplas e eles se diferem quanto ao controle do erro tipo I e em relação ao poder. Alguns testes, como o *t* de Student e Duncan, possuem elevadas taxas de erro tipo I por experimento, sendo liberais, enquanto outros, como Scheffé e Tukey, possuem taxas de erro tipo I por experimento inferiores ao nível nominal de significância e são considerados conservadores.

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

O *bootstrap* é um método computacional desenvolvido recentemente e é usado principalmente na obtenção de estimativas de parâmetros (MANLY, 1998). A idéia básica de *bootstrap*, na ausência de qualquer conhecimento sobre a população, 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.

Em estudos de desempenho de testes estatísticos, devido à dificuldade de se obterem analiticamente informações sobre as taxas de erro tipo I e poder, a simulação Monte Carlo é uma

¹ICEX - UNIFAL-MG. e-mail: brunag25_@hotmail.com.br

²Agradecimento à FAPEMIG pelo apoio financeiro.

³ICEX - UNIFAL-MG. e-mail: patricia.ramos@unifal-mg.edu.br

alternativa viável para comparar os testes de comparações múltiplas e vários trabalhos já utilizaram a simulação para comparar testes de comparações múltiplas.

Segundo Machado et al. (2005) e Ramos e Ferreira (2009), o teste SNK é um teste que controla as taxas de erro tipo I por experimento sob H_0 mas se torna liberal sob H_0 parcial. Além disso, seu poder é superior ao poder dos testes de Tukey, t protegido de Bonferroni e Scheffé. Portanto, é um teste com boas qualidades mas que podem ser melhoradas com o uso do *bootstrap*.

Assim, o objetivo desse trabalho foi propor uma versão utilizando *bootstrap* paramétrico do teste de comparações múltiplas SNK, comparar suas taxas de erro tipo I e poder com o teste SNK original e com outros testes com boas propriedades, tais como Scott-Knott e os testes de Caliński & Corsten, avaliados por Ramos e Ferreira (2009).

2 Material e métodos

Os testes de Student-Newman-Keuls original (SNK) e sua versão *bootstrap* (SNK_B) tiveram seu desempenho avaliado por meio de simulação Monte Carlo. Para avaliar os desempenhos desses testes foram delineadas simulações Monte Carlo sob H_0 , para mensurar as taxas de erro tipo I por experimento (TPE), e sob H_1 , para mensurar o poder. Todas as rotinas necessárias para a implementação e avaliação dos testes estão sendo realizadas utilizando o programa R em sua versão 3.0.2 (R Development Core Team, 2013).

2.1 Teste Student-Newman-Keuls

Para aplicação do teste SNK são seguidos os passos descritos:

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;
 - 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- $p < \alpha$, 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 é obtido por

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

em que p é o número de médias abrangidas pelo contraste, QME é o quadrado médio do resíduo da análise de variância e ν são os graus de liberdade associados ao QME .

2.2 Teste Student-Newman-Keuls *bootstrap*

O teste SNK_B difere do teste SNK no cálculo do valor- p para decidir se as médias abrangidas pelo contraste em questão podem ser consideradas iguais ou diferentes, pois será obtida uma distribuição da amplitude estudentizada q para cada reamostragem. Para efetuar a reamostragem *bootstrap*, novas rk observações são obtidas a partir da distribuição normal com média $\mu = 10$ e desvio padrão $\sigma = 1$ para compor novas amostras de cada tratamento. Calculam-se as médias das amostras e um valor q_b é obtido por meio de

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

em que QME_b é o quadrado médio do resíduo da b -ésima amostra *bootstrap* e $\bar{Y}_{(1)}^b$ e $\bar{Y}_{(k)}^b$ são a maior e a menor média obtidas na b -ésima amostra.

Este processo foi repetido $B = 1.000$ vezes e o conjunto com todos os valores q_b obtidos foram utilizado para obter os valores- p por:

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

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

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 obtenção das letras que distinguem médias iguais e diferentes.

2.3 Simulações

Foram consideradas $N = 1.000$ simulações de k tratamentos qualitativos não estruturados e número de repetições r . Os valores de k utilizados foram $k = 5; 10; 20$ e 80 tratamentos e valores de $r = 4; 10$ e 20 repetições e o nível nominal de significância considerado é 5% . Essas combinações de valores tentam ilustrar o que acontece em situações reais, com poucos e muitos tratamentos e com números de repetições variados. Foi adotado o modelo de probabilidade normal, mas outros modelos, como o lognormal e o exponencial, serão considerados posteriormente para avaliar a robustez dos testes em situações adversas.

3 Resultados e discussões

Na Tabela 1 são apresentadas as taxas de erro tipo I por experimento (TPE) sob normalidade em função de r e k para H_0 completa dos testes SNK original (SNK), versão *bootstrap* do teste (SNK_B), e os resultados obtidos por Ramos e Ferreira (2009) sobre as taxas de erro tipo I por

experimento dos testes de Caliński & Corsten (C) a sua versão *bootstrap* (CB) e o teste de Scott-knott (SK) também sob normalidade. As taxas observadas foram confrontadas com os intervalos de 99% de confiança para proporções considerando o nível nominal de 0,05.

Tabela 1: Taxas de erro por experimento (TPE), dos testes SNK original (SNK), SNK *Bootstrap* (SNK_B), teste de Caliński & Corsten (C), sua versão *bootstrap* (CB) e teste de Scott-knott (SK) sob H_0 completa, considerando-se a distribuição normal (10, 1) e nível de significância 0,05.

Nº Tratamentos (k)	Repetições (r)	SNK	SNK _B	C	CB	SK
5	4	0,045	0,043	0,049	0,053	0,051
	10	0,053	0,053	0,047	0,046	0,067
	20	0,052	0,051	0,057	0,058	0,059
10	4	0,052	0,051	0,055	0,057	0,043
	10	0,066	0,067	0,042	0,042	0,062
	20	0,054	0,055	0,044	0,046	0,051
20	4	0,047	0,050	0,047	0,050	0,038
	10	0,051	0,051	0,048	0,054	0,052
	20	0,057	0,056	0,054	0,054	0,056
80	4	0,047	0,053	0,044	0,041	0,036
	10	0,043	0,046	0,048	0,050	0,052
	20	0,050	0,048	0,041	0,042	0,050

Na Tabela 2 são apresentadas o poder dos testes SNK e SNK_B sob normalidade, em função de r e k e da diferença δ em erros padrões entre médias sob H_1 e $\alpha = 0,05$.

Tabela 2: Poder dos testes SNK original (SNK), SNK *Bootstrap* (SNK_B), em função do número de níveis k do fator, número de repetições r para o nível nominal de significância $\alpha = 0,05$, sob H_1 completa, considerando-se a distribuição normal (10, 1).

k	r	Teste	Diferença real entre as médias (δ)					
			1 $\sigma_{\bar{x}}$	2 $\sigma_{\bar{x}}$	4 $\sigma_{\bar{x}}$	8 $\sigma_{\bar{x}}$	16 $\sigma_{\bar{x}}$	32 $\sigma_{\bar{x}}$
5	4	SNK	0,0532	0,1360	0,5500	-	-	-
		SNK _B	0,2280	0,31366	0,3810	-	-	-
5	20	SNK	0,0465	0,1310	0,6470	-	-	-
		SNK _B	0,6405	0,8847	0,9940	-	-	-
10	4	SNK	0,0369	0,0967	0,4996	0,9985	-	-
		SNK _B	0,2744	0,4177	0,6293	0,8160	-	-
10	20	SNK	0,0328	0,0985	0,5733	1,0000	-	-
		SNK _B	0,6404	0,8932	0,9975	1,0000	-	-
20	4	SNK	0,0318	0,0955	0,5288	0,9994	1,0000	-
		SNK _B	0,3041	0,4956	0,7741	0,9722	0,9995	-
20	20	SNK	0,0321	0,0965	0,5517	0,9999	1,0000	-
		SNK _B	0,5230	0,7954	0,9859	1,000	1,0000	-
80	4	SNK	0,0277	0,08826	0,5492	0,9999	1,0000	1,0000
		SNK _B	0,3243	0,5473	0,8615	0,9979	1,0000	1,0000
80	20	SNK	0,0268	0,0877	0,5578	0,9999	1,0000	1,0000
		SNK _B	0,6501	0,9016	0,9980	1,0000	1,0000	1,0000

Em todas as situações (diferentes r e k), todos os testes apresentaram valores de taxas de

erro tipo I próximos do valor do nível nominal correspondente de 0,05 (TABELA 1). Assim, sob normalidade, não houve nenhuma tendência dos testes SNK original e SNK_B serem liberais ou conservadores, assim como os testes de Caliński & Corsten, sua versão *bootstrap* e o teste Scott-knott.

Os valores do poder, nas situações simuladas, aumentaram com o aumento de erros padrões (δ). O poder do teste SNK_B , é em geral, superior ao do poder do teste SNK mas, para valores baixos de r , o poder do SNK_B é um pouco menor que o poder do SNK quando a diferença δ entre as médias é grande (TABELA 2).

Os resultados obtidos foram sob H_0 completa e H_1 , porém os programas de simulação estão sendo implementados e os resultados ainda serão obtidos sob H_0 parcial e considerando distribuições não-normais.

4 Conclusões

Os testes de Student-Newman-Keuls (SNK) original e *Bootstrap* são exatos sob H_0 completa e normalidade. O poder do teste proposto superou o poder do teste original sob H_1 na maioria das situações.

Referências

- [1] MACHADO, A. A.; DEMÉTRIO, C. G. B.; FERREIRA, D. F.; SILVA, J. G. C. Estatística experimental: uma abordagem fundamentada no planejamento e no uso de recursos computacionais. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11., Londrina, 2005. Curso, Londrina: ISBN, 2005. 290 p.
- [2] MANLY, B. F. J. **Randomization, bootstrap and Monte Carlo methods in biology**. 2. ed. London: Chapman-Hall, 1998. 399 p.
- [3] MONTGOMERY, D. E. **Design and Analysis of Experiments**. 5. ed. New York: John Wiley and Sons, 2001.
- [4] 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.
- [5] R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org>>. Acesso em: 25 out. 2013.