

# Teste de comparações múltiplas *bootstrap* no pacote *ExpDes* para DBC e DQL

Patrícia de Siqueira Ramos<sup>1 2</sup>

Eric Batista Ferreira<sup>3</sup>

## 1 Introdução

A comparação de tratamentos qualitativos no contexto experimental conta com uma grande diversidade de testes disponíveis. Entretanto, a capacidade desses testes em controlar o nível nominal de significância e apresentar alto poder são muito variáveis. Como testes de comparações múltiplas baseados em computação intensiva têm se mostrado promissores por trabalhos recentes, a sua implementação e avaliação são de grande utilidade para pesquisadores que desejam rapidez e facilidade na execução das análises e interpretação dos resultados, além de poderem confiar nas taxas de erro praticadas.

Ramos e Ferreira (2009) propuseram uma alternativa *bootstrap* para o procedimento de comparações múltiplas de Caliński e Corsten (1985), uma extensão da amplitude estudentizada e baseado em análise de agrupamento, e avaliaram as versões original e *bootstrap* deste procedimento por meio de simulação Monte Carlo, considerando modelos de probabilidade normais e não-normais, para o delineamento inteiramente casualizado (DIC).

Considerando-se  $k$  tratamentos e  $r$  repetições, as  $rk$  observações amostrais são combinadas em uma única amostra em que são realizadas  $r$  reamostragens com reposição, formando novas amostras de cada população. A partir dessas amostras, novas médias são obtidas e a amplitude estudentizada é calculada por

$$q_b = \frac{\bar{X}_{(k)}^b - \bar{X}_{(1)}^b}{\sqrt{\frac{QME_b}{r}}}$$

, em que  $QME_b$  é o quadrado médio do resíduo da  $b$ -ésima amostra *bootstrap* e  $\bar{X}_{(1)}^b$  e  $\bar{X}_{(k)}^b$  são as médias amostrais ordenadas da primeira e  $k$ -ésima reamostragem de *bootstrap*, sendo que o sobrescrito  $b$  foi utilizado para indicar a  $b$ -ésima amostra de *bootstrap*.

O processo é repetido  $B = 1000$  vezes e o conjunto com todos os valores  $q_b$  obtidos ( $b = 1, 2, \dots, B$ ) são utilizados para calcular o valor- $p$  por

<sup>1</sup>ICEx - UNIFAL-MG. e-mail: [patricia.amos@unifal-mg.edu.br](mailto:patricia.amos@unifal-mg.edu.br).

<sup>2</sup>Agradecimento à FAPEMIG pelo apoio financeiro.

<sup>3</sup>ICEx - UNIFAL-MG. e-mail: [eric.ferreira@unifal-mg.edu.br](mailto:eric.ferreira@unifal-mg.edu.br).

$$\text{valor-}p = \frac{\sum_{b=1}^B I\left(\frac{R_\ell}{\sqrt{\frac{QME}{r}}} < q_b\right)}{B} \quad \ell = 1, 2, \dots, k-1,$$

em que  $I(x)$  é a função indicadora de  $x$ . O valor- $p$  é comparado com  $\alpha$  em cada ponto de fusão e o primeiro passo de fusão para o qual o valor- $p$  for menor do que o nível de significância  $\alpha$  é considerado o ponto de corte e os grupos de tratamentos foram obtidos.

Nesse trabalho, os autores afirmam que o teste *bootstrap* foi, em geral, considerado superior ao do teste original e robusto. Os dois testes foram exatos sob  $H_0$  completa e normalidade. O teste *bootstrap* sob não-normalidade e  $H_0$  completa controlou o erro tipo I por experimento e foi considerado robusto. O teste original sob não-normalidade e  $H_0$  completa foi conservador para valores pequenos de  $k$  e liberal para grandes valores de  $k$ .

Para analisar experimentos de forma simples no programa R (R CORE TEAM, 2013) foi criada a biblioteca de funções *Experimental Designs (ExpDes)*, um pacote destinado a análise de experimentos balanceados por meio modelos de análise de variância de efeito fixo (FERREIRA et al., 2013). Esse pacote também possui a sua versão em Português *ExpDes.pt*, e ambos se encontram disponíveis no CRAN: <http://www.vps.fmvz.usp.br/CRAN/web/packages/ExpDes/index.html>.

Após verificar a significância pelo teste  $F$  da análise de variância, o pacote permite a realização da comparação de múltiplas médias pelos testes de Tukey, SNK, Scott-Knott, Duncan,  $t$  (LSD) e Bonferroni (LSD protegido) se os tratamentos forem qualitativos ou o ajuste de modelos de regressão polinomial se os tratamentos forem quantitativos.

Embora tenha mostrado bons resultados, o teste *bootstrap* proposto por Ramos e Ferreira (2009) foi avaliado somente para o DIC, o que acaba restringindo o seu uso. Assim, os objetivos desse trabalho foram generalizar esse teste para DBC e DQL para uso no pacote *ExpDes*, já que estes autores o implementaram utilizando apenas o DIC, e avaliar seu comportamento após as modificações por meio de um estudo de simulação Monte Carlo.

## 2 Material e métodos

O programa R foi utilizado para implementar as funções. O teste *bootstrap* (CB) implementado no pacote *ExpDes* para o DIC foi generalizado para o DBC e DQL e seu comportamento foi testado após as modificações por meio de simulação Monte Carlo. Seu desempenho foi avaliado pelas taxas de erro tipo I por experimento e pelo poder. Foram consideradas as situações de  $H_0$  completa, em que as médias foram consideradas todas iguais e  $H_1$ , em que as médias eram todas diferentes.

Os dados foram simulados a partir do modelo probabilístico normal (um dos pressupostos da análise de variância) com média igual a 10 e desvio padrão 1. O processo de reamostragem, de acordo com cada tipo de delineamento, foi repetido  $B$  vezes. O valor- $p$  foi obtido para cada

situação e o mesmo critério definido para o teste de Ramos e Ferreira (2009) foi aplicado para obter os grupos de tratamentos homogêneos. Diversos valores de  $k$  para tratamentos foram considerados, além do número de repetições  $r$  e de blocos  $b$ ; e números de erros padrões  $\delta$  de diferenças entre médias consecutivas, já que o poder foi avaliado sob  $H_1$ , em que as médias foram todas diferentes. No DQL, o número de linhas e colunas correspondiam aos mesmos números de tratamentos utilizados, ou seja,  $k = 5, 10, 20$  e  $80$ . Em todos os casos, os níveis nominais de significância considerados foram de 1 e 5%.

### 3 Resultados e discussões

O teste CB foi implementado no pacote *ExpDes* para o DIC. Porém, percebeu-se que a sua implementação para DBC e DQL não seria feita de forma idêntica ao DIC. O tipo de reamostragem *bootstrap* adotada por Ramos e Ferreira (2009) foi a não-paramétrica, ou seja, as amostras *bootstrap* eram obtidas a partir de reamostragens da amostra de dados original.

O modelo estatístico de um DBC é

$$Y_{ij} = \mu + \tau_i + b_j + \varepsilon_{ij},$$

em que  $Y_{ij}$  é a observação feita na parcela referente ao tratamento  $i$  no bloco  $j$ ;  $\mu$  representa uma constante inerente a toda parcela;  $\tau_i$  representa o efeito do tratamento  $i$ ;  $b_j$  representa o efeito do bloco  $j$ ;  $\varepsilon_{ij}$  representa o erro experimental na parcela  $i, j$ .

De forma análoga, o modelo estatístico de um DQL é

$$Y_{ij} = \mu + \tau_i + c_j + l_m + \varepsilon_{jm(i)},$$

em que  $Y_{ij}$  é a observação feita na parcela da coluna  $j$  e na linha  $m$  que recebe o tratamento  $i$ ;  $c_j$  representa o efeito da coluna  $j$ ;  $l_m$  representa o efeito da linha  $m$ ;  $\varepsilon_{ij}$  representa o erro experimental na parcela  $i, j, m$ ;  $i, j, m = 1, \dots, k$ .

Para o DIC, os próprios dados simulados foram reamostrados  $B$  vezes. Para o DBC e DQL não foi efetuada a reamostragem dos dados e sim um outro tipo de resíduo, o  $\varepsilon^*$ , que não é o resíduo do modelo, mas sim um tipo de resíduo especial. Para o DBC:

$$\varepsilon_{ij\text{DBC}}^* = Y_{ij} - \bar{b}_j, \quad (1)$$

em que  $\bar{b}_j$  é a média das observações do bloco  $j$ . E, para o DQL:

$$\varepsilon_{jm(i)\text{DQL}}^* = Y_{ij} - \bar{c}_j + \bar{l}_m, \quad (2)$$

em que  $\bar{c}_j$  e  $\bar{l}_m$  são as médias das observações da coluna  $j$  e linha  $l$ , respectivamente.

Assim, esses resíduos especiais,  $\varepsilon^*$ , é que foram reamostrados em cada passo  $b$ . Dessa

forma, o efeito do controle local, seja bloco ou linha e coluna, é retirado.

Na Tabela 1 são apresentados os valores de erro tipo I obtidos com a aplicação dos testes de Caliński e Corsten (C) e sua versão *bootstrap* (CB) para o DIC e para o DBC a 1 e 5% de significância. Os valores para o DIC são usados como referência e foram obtidos por Ramos e Ferreira (2009). Os dois testes foram exatos em todas as situações.

Tabela 1: Taxas de erro por experimento (TPE), dos testes de Caliński e Corsten (C), sua versão *bootstrap* (CB) em função do número de repetições  $r$ , número de níveis  $k$  do fator e níveis nominais de significância  $\alpha = 0,05$  e  $\alpha = 0,01$ , sob  $H_0$  completa, considerando-se a distribuição normal (10, 1) e os delineamentos DIC e DBC.

$k$	$r$	DIC				DBC			
		C		CB		C		CB	
		0,05	0,01	0,05	0,01	0,05	0,01	0,05	0,01
5	4	0,049	0,009	0,053	0,010	0,047	0,008	0,053	0,012
5	10	0,047	0,009	0,046	0,012	0,041	0,007	0,048	0,010
5	20	0,057	0,018	0,058	0,016	0,046	0,007	0,047	0,009
10	4	0,055	0,009	0,057	0,011	0,050	0,010	0,059	0,012
10	10	0,042	0,007	0,042	0,007	0,053	0,010	0,053	0,016
10	20	0,044	0,010	0,046	0,012	0,042	0,007	0,046	0,007
20	4	0,047	0,006	0,050	0,008	0,057	0,011	0,059	0,014
20	10	0,048	0,010	0,054	0,013	0,055	0,010	0,060	0,012
20	20	0,054	0,009	0,054	0,010	0,048	0,007	0,048	0,008
80	4	0,044	0,010	0,051	0,015	0,046	0,003	0,044	0,004
80	10	0,048	0,008	0,050	0,009	0,057	0,011	0,055	0,013
80	20	0,041	0,007	0,042	0,011	0,054	0,013	0,059	0,014

Para o DQL as configurações foram um pouco diferentes já que esse tipo de delineamento exige que o número de tratamentos e repetições sejam iguais, como está expresso na Tabela 2. A única situação em que o teste CB não foi exato foi para  $k = 5$  e  $r = 5$  mas, mesmo assim, o valor de erro obtido não ficou muito além do limite superior, que era 0,070504.

Tabela 2: Taxas de erro por experimento (TPE), dos testes de Caliński e Corsten (C), sua versão *bootstrap* (CB) em função do número de repetições  $r$ , número de níveis  $k$  do fator e níveis nominais de significância  $\alpha = 0,05$  e  $\alpha = 0,01$ , sob  $H_0$  completa, considerando-se a distribuição normal (10, 1) e delineamento DQL.

$k$	$r$	C		CB	
		0,05	0,01	0,05	0,01
5	5	0,049	0,008	0,072**	0,018
10	10	0,051	0,013	0,053	0,018
20	20	0,043	0,005	0,042	0,008
80	80	0,045	0,012	0,050	0,013

\*\*Ultrapassou o LS do IC exato, com 99% de confiança para  $\alpha = 0,01(0,021276)$  e  $\alpha = 0,05(0,070504)$ .

++Não atingiu o LI do IC exato, com 99% de confiança para  $\alpha = 0,01(0,003727)$  e  $\alpha = 0,05(0,033927)$ .

Na Tabela 3 estão apresentados os valores de poder do teste CB para o DBC, em função do número de repetições  $r$ , número  $k$  de tratamentos e da diferença  $\delta$  em erros padrões entre médias, considerando a distribuição normal sob  $H_1$  e  $\alpha = 0,05$ .

Os valores de poder obtidos (Tabela 3) foram muito próximos aos obtidos por Ramos e Ferreira (2009) para o DIC, indicando que o poder se manteve do DIC para o DBC. Os valores de poder para o DQL com  $\alpha = 0,05$  se encontram na Tabela 4 e seus valores foram também muito próximos dos obtidos para o DIC, considerando-se o mesmo número de tratamentos  $k$ .

Tabela 3: Poder do teste de Caliński e Corsten *bootstrap* (CB), 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) e DBC.

$k$	$r$	Diferença real entre as médias ( $\delta$ )					
		$1 \sigma_{\bar{x}}$	$2 \sigma_{\bar{x}}$	$4 \sigma_{\bar{x}}$	$8 \sigma_{\bar{x}}$	$16 \sigma_{\bar{x}}$	$32 \sigma_{\bar{x}}$
5	4	0,1658	0,2913	0,5300	-	-	-
5	20	0,1948	0,3503	0,6350	-	-	-
10	4	0,2147	0,3766	0,7537	0,9980	-	-
10	20	0,2226	0,3935	0,7733	1,0000	-	-
20	4	0,2265	0,3871	0,7520	0,9973	1,0000	-
20	20	0,2277	0,3899	0,7504	0,9984	1,0000	-
80	4	0,2302	0,3838	0,7283	0,9960	1,0000	1,0000
80	20	0,2153	0,3628	0,6987	0,9932	1,0000	1,0000

Os resultados para  $\alpha = 0,01$  seguiram o mesmo padrão de resposta apresentado por  $\alpha = 0,05$  e não foram apresentados.

Tabela 4: Poder do teste de Caliński e Corsten *bootstrap* (CB), 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) e DQL.

$k$	$r$	Diferença real entre as médias ( $\delta$ )					
		$1 \sigma_{\bar{x}}$	$2 \sigma_{\bar{x}}$	$4 \sigma_{\bar{x}}$	$8 \sigma_{\bar{x}}$	$16 \sigma_{\bar{x}}$	$32 \sigma_{\bar{x}}$
5	5	0,2410	0,4527	0,7770	-	-	-
10	10	0,2211	0,3925	0,7640	0,9990	-	-
20	20	0,2309	0,3889	0,7518	0,9979	1,0000	-
80	80	0,2142	0,3604	0,6944	0,9927	1,0000	1,0000

## 4 Conclusões

De forma geral, o procedimento de se realizar o *bootstrap* nos resíduos permitiu que o teste fosse estendido para os delineamentos DBC e DQL mantendo as mesmas características apresentadas para DIC e permitindo seu uso em um número bem maior de experimentos.

## Referências

- [1] CALIŃSKI, T.; CORSTEN, L. C. A., Clustering means in ANOVA by simultaneous testing, **Biometrics**, v. 41, p. 39-48, 1985.
- [2] FERREIRA, E. B.; CAVALCANTI, P. P.; NOGUEIRA, D. A. **ExpDes**: Experimental Designs package. R package version 1.1.2. 2013.
- [3] R CORE TEAM. **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing, 2013.
- [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.