

TESTES DE HIPÓTESES *BOOTSTRAP* PARA DUAS AMOSTRAS OBTIDAS A PARTIR DE CONJUNTOS ORDENADOS

Alexandre Souza MAGRO¹
Maria Cecília Mendes BARRETO¹

- RESUMO: Diversos trabalhos envolvendo o delineamento por amostras em conjuntos ordenados mostram sua eficiência em relação à amostragem aleatória simples em diferentes procedimentos de inferência estatística quando a mensuração da variável de interesse é dispendiosa. Bohn e Wolfe (1990, 1992) desenvolveram uma modificação da estatística de Mann-Whitney-Wilcoxon para a comparação de duas populações quando o delineamento amostral é por conjuntos ordenados. Por outro lado, a necessidade de se usar amostras de tamanho reduzido decorrentes do custo de mensuração enseja a utilização de métodos *bootstrap* quando métodos assintóticos podem não ser recomendados. Este trabalho apresenta uma proposta original para a implementação de reamostragem *bootstrap* em testes de comparação dos parâmetros de locação de duas distribuições de probabilidade utilizando-se amostras obtidas por conjuntos ordenados. O tamanho e o poder desses testes baseados em reamostragem *bootstrap* são avaliados por meio de experimentos Monte Carlo usando as estatísticas de Mann-Whitney-Wilcoxon, a diferença entre médias e a diferença entre medianas amostrais como estatística de teste. Esses resultados são praticamente equivalentes em tamanho e poder do teste da estatística modificada de Mann-Whitney-Wilcoxon de Bohn e Wolfe (1990; 1992). Dessa maneira o teste baseado em reamostragem *bootstrap* usando a estatística modificada de Mann-Whitney-Wilcoxon pode ser usado com êxito como uma alternativa ao teste assintótico de Bohn e Wolfe (1990; 1992).
- PALAVRAS-CHAVE: Amostragem por conjuntos ordenados; *ranked set sampling*; *bootstrap*; testes de hipóteses para duas amostras independentes; estatística modificada de Mann-Whitney-Wilcoxon.

1 Introdução

Em diversos estudos amostrais, a mensuração de variáveis pode ser demasiadamente dispendiosa, comprometendo sua viabilidade caso o tamanho amostral seja grande. Principalmente para essas situações, os procedimentos eficientes de amostragem são de grande valia. Em alguns casos, os indivíduos da população podem ser ordenados em relação à variável de interesse antes de sua efetiva mensuração. Por exemplo, a ordenação dos indivíduos em relação à variável de interesse pode ser feita por de uma variável secundária fortemente correlacionada com a variável de interesse, mas de custo de

¹Departamento de Estatística, Universidade Federal de São Carlos – UFSCar, CEP 13565-905, São Carlos, SP, Brasil, E-mail: cbarreto@power.ufscar.br.

mensuração mais baixo, ou mesmo pela de inspeção visual. Quando isso é possível, o esquema de amostragem por conjuntos ordenados (ACO), ou *ranked set sampling*, proporciona estimadores mais eficientes e testes mais poderosos que os que seriam obtidos se fossem empregadas amostras aleatórias simples (McIntyre, 1952; Barnett, 1999).

Os diversos estudos apresentados na literatura sobre ACO avaliam estimadores sob esse delineamento e fazem comparações com estimadores baseados em amostragem aleatória simples. Na estimação da média populacional, Dell e Clutter (1972), estudaram a eficiência relativa do estimador média amostral considerando diferentes tamanhos de amostras e distribuições e verificaram a superioridade da ACO, mesmo quando a ordenação não é perfeita. Kaur et al. (1997) apresentam um estimador da média populacional que se mostrou mais eficiente que a média amostral quando a população sob estudo é assimétrica. Estimadores lineares ótimos baseados em ACO para os parâmetros de distribuições da família locação e escala foram desenvolvidos por Stokes (1995) e Barnett e Moore (1997) e para os parâmetros de um modelo de regressão linear simples por Barreto e Barnett (1999). Barnett (1999) apresenta um estudo sobre a eficiência de diversos estimadores da média populacional e conclui que os estimadores lineares ótimos apresentam melhores resultados.

Trabalhos que abordam o desempenho de procedimentos baseados em ACO para estimação intervalar são encontrados em Chen (2000), Cesário e Barreto (2003), Raminelli (2003), Taconeli e Barreto (2003) e Garcia (2004).

Testes de hipóteses que levam em consideração ACO são abordados por Hettmansperger (1995), Koti e Babu (1996), Barabesi (1998) e Öztürk e Wolfe (2000), que apresentam extensões do teste do sinal, e em Bohn e Wolfe (1990, 1992, 1994) que desenvolveram uma modificação na estatística de Mann-Whitney-Wilcoxon (MWW) para o caso de comparação de duas distribuições em função do parâmetro de locação.

Por outro lado, métodos computacionalmente intensivos, entre eles, o de reamostragem, têm sido utilizados para superar dificuldades analíticas encontradas no desenvolvimento de procedimentos para inferência estatística em geral. Em particular, o *bootstrap* (por exemplo, Efron e Tibishirani, 1993; Davison e Hinkley, 1997) tem sido bastante discutido como um método computacionalmente intensivo válido para diversas situações inferenciais, além de ser facilmente implementado.

Neste artigo discute-se a aplicação do método *bootstrap* para testes de hipóteses utilizando amostras obtidas por conjuntos ordenados. Os objetivos específicos do trabalho são (i) apresentar a adaptação da versão não-paramétrica do método *bootstrap* para o esquema de ACO, desenvolvido por Magro (2003) e não encontrado em outro texto da revisão bibliográfica aqui feita, (ii) comentar sua implementação na realização de testes de hipóteses envolvendo a diferença nos parâmetros de locação de duas populações, (iii) discutir o experimento Monte Carlo realizado para avaliar o poder do teste *bootstrap* para algumas configurações amostrais e distribuições e (iv) confrontar o poder do teste *bootstrap* ao poder do teste não-paramétrico desenvolvido por Bohn e Wolfe (1990; 1992).

Na primeira seção, discute-se o esquema de ACO e apresentam-se alguns resultados básicos para o estimador média amostral e uma revisão bibliográfica sobre procedimentos de inferência estatística. Em seguida, são introduzidos alguns conceitos sobre testes de hipóteses estatísticas e feita uma apresentação do estudo conduzido por Bohn e Wolfe (1990; 1992), que abordam o desempenho da estatística modificada de MWW em ACO.

Nas duas seções seguintes discute-se o método *bootstrap*, justificando-se seu uso para inferência estatística, explicando-se como ele pode ser utilizado para a realização de testes de hipóteses estatísticas envolvendo duas populações e estendendo-o para a situação em que se dispõe de amostras obtidas a partir de conjuntos ordenados. Por fim, apresentam-se e discutem-se os resultados do experimento Monte Carlo utilizado para avaliar o poder dos testes estudados e as conclusões extraídas.

2 Amostragem por conjuntos ordenados

Motivado pela necessidade de superar a dificuldade de coletar amostras de tamanho pequeno de pastagem para estimar com precisão determinadas quantidades de interesse, McIntyre (1952) desenvolveu a ACO com o objetivo de mensurar uma variável X em uma amostra de $n=km$ elementos. O esquema de ACO pode ser descrito em três passos. O primeiro passo consiste em selecionar aleatoriamente k^2m indivíduos I_{jie} sendo $j=1, \dots, m$, $i=1, \dots, k$, $e=1, \dots, k$, que vão formar km conjuntos, denotados por i_j , de k elementos cada. No segundo passo, em cada conjunto i_j , os k indivíduos são ordenados de acordo com alguma variável fortemente correlacionada com a variável X ou por meio de julgamento. Nesse passo, os elementos são apenas ordenados e *não efetivamente mensurados*. A mensuração dos indivíduos ocorre no terceiro passo. Assim, em cada um dos conjuntos i_j é mensurado apenas o indivíduo que ocupa a posição ou posto i , $I_{ji(i)}$, que passa agora a ser denotado por $X_{ji(i)}$, enfatizando o fato de que só nesse momento a variável de interesse é efetivamente mensurada. Desse modo, a amostra resultante, formada por $X_{11(1)}, X_{12(2)}, \dots, X_{1k(k)}, X_{21(1)}, X_{22(2)}, \dots, X_{2k(k)}, \dots, X_{m1(1)}, X_{m2(2)}, \dots, X_{mk(k)}$, é denominada amostra por conjuntos ordenados.

O objetivo básico desse esquema de amostragem é favorecer o espalhamento dos valores amostrais entre os possíveis valores da variável de interesse. Outra característica importante dos elementos de uma amostra obtida por conjuntos ordenados é que os km indivíduos mensurados são selecionados de amostras independentes de tamanho k , correspondendo a m repetições de cada um dos k postos de ordenação. Como cada uma das m replicações amostrais do i -ésimo posto, $i = 1, \dots, k$, é selecionada de conjuntos independentes, e formam uma amostra aleatória simples do i -ésimo posto, entre k postos, de X .

De acordo com Barnett (1999), para um número km fixo de elementos a serem mensurados, quanto maior é o número de postos de ordenação (k) em relação ao número de replicações de cada posto (m) mais informativa é a amostra e melhores são as propriedades dos métodos de inferência estatística produzidos. Entretanto, se o método de ordenação se basear em julgamento, o maior número de postos dificulta comparações e favorece erros na ordenação. Conseqüentemente, as propriedades dos métodos de inferência são afetadas de alguma maneira, equivalendo, na pior das hipóteses, à amostra aleatória simples (por exemplo, Takahasi e Wakimoto, 1968). Na situação em que a ordenação é feita por meio de uma variável auxiliar, quanto maior for o coeficiente de correlação entre a variável de interesse e a variável auxiliar, melhores resultados são obtidos a partir de amostras de conjuntos ordenados (Patil et al., 1993; Stokes, 1995; Yu e Lam, 1997).

Uma suposição básica adotada no presente trabalho é que a ordenação dos indivíduos em cada conjunto é realizada corretamente, ou seja, os valores mensurados correspondem às estatísticas de ordem.

Seja X_1, \dots, X_k uma amostra aleatória de uma variável X com função densidade $f(x)$ e função de distribuição $F(x)$, mostra-se que (por exemplo, David, 1981) a função de distribuição da i -ésima estatística de ordem é dada por

$$F_{(i)}(x) = P(X_{(i)} \leq x) = \sum_{j=i}^k \binom{k}{j} F^j(x) [1 - F(x)]^{k-j}$$

e a relação entre as $F_{(i)}(x)$, $i = 1, \dots, k$, e $F(x)$ é dada por

$$F(x) = \frac{1}{k} \sum_{i=1}^k F_{(i)}(x) \quad (1)$$

Esses dois resultados fundamentam o desenvolvimento do procedimento *bootstrap* proposto na seção 5.1.

3 Teste de Mann-Whitney-Wilcoxon para duas amostras de conjuntos ordenados

Bohn e Wolfe (1990; 1992) desenvolveram um teste para a diferença entre os parâmetros de locação de duas distribuições de probabilidades sob o delineamento de amostras por conjuntos ordenados. Usando uma estatística análoga à estatística de MWW para amostras aleatórias simples, a estatística modificada de MWW é uma função de uma estatística da classe U . Mostra-se que a maior parte das estatísticas dessa classe corresponde a uma soma de elementos identicamente distribuídos mas não independentes. Portanto, a forma clássica do teorema central do limite não pode ser diretamente aplicada. Em Magro (2003) está uma demonstração da normalidade assintótica de estimadores da classe U baseada em um método desenvolvido por Hoeffding (1948).

Sejam X_{1ij}, \dots, X_{kij} , $i = 1, \dots, k$ e $j = 1, \dots, m$, mk amostras aleatórias simples de tamanho k mutuamente independentes de uma variável aleatória contínua com função densidade $f(x)$ e função de distribuição $F(x)$. Sejam Y_{1st}, \dots, Y_{qst} , $s = 1, \dots, q$ e $t = 1, \dots, n$, nq amostras aleatórias simples de tamanho q mutuamente independentes de uma segunda variável aleatória contínua com função densidade $g(y) = f(y - \Delta)$ e função de distribuição $G(y) = F(y - \Delta)$, onde Δ é a diferença entre os parâmetros de locação das distribuições de X e de Y . Assume-se que todos os elementos do conjunto de mk^2 X 's são independentes dos elementos do conjunto de nq^2 Y 's.

A amostra de conjuntos ordenados de mk observações de X é dada por $\mathbf{X}_1 = (X_{11(1)}, \dots, X_{1k(k)}), \dots, \mathbf{X}_m = (X_{m1(1)}, \dots, X_{mk(k)})$, e a de nq observações de Y é dada por $\mathbf{Y}_1 = (Y_{11(1)}, \dots, Y_{1q(q)}), \dots, \mathbf{Y}_n = (Y_{n1(1)}, \dots, Y_{nq(q)})$, obtidas de acordo com a seção 2.

Sejam $f_{(j)}(x)$ a função densidade de $X_{j(i)}$ ($j = 1, \dots, m$; $i = 1, \dots, k$) e $g_{(s)}(y)$ a função densidade de $Y_{ts(s)}$, ($t = 1, \dots, n$; $s = 1, \dots, q$). Dada a independência mútua entre as mk^2 observações de X , $\mathbf{X}_1, \dots, \mathbf{X}_m$ são uma amostra aleatória de tamanho m da distribuição conjunta com função densidade conjunta $h(\mathbf{x}) = \prod_{i=1}^k f_{(i)}(x_{i(i)})$. Da mesma forma, as

nq^2 observações de $Y, \mathbf{Y}_1, \dots, \mathbf{Y}_n$ são uma amostra aleatória de tamanho n da distribuição conjunta com função densidade conjunta $p(\mathbf{y}) = \prod_{s=1}^q g_{(s)}(y_{s(s)})$. Além disso, $(\mathbf{X}_1, \dots, \mathbf{X}_m)$ e $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ são também independentes. Supondo-se ordenação perfeita na seleção de estatísticas de ordem para as amostras de conjuntos ordenados, as funções densidade $f_{(i)}(x)$ e $g_{(s)}(y)$ são as densidades marginais das estatísticas de ordem de forma

$$f_{(i)}(x) = \frac{k!}{(i-1)!(k-i)!} [F(x)]^{i-1} [1-F(x)]^{k-i} f(x) \quad (2)$$

e

$$g_{(s)}(y) = \frac{q!}{(s-1)!(q-s)!} [F(y-\Delta)]^{s-1} [1-F(y-\Delta)]^{q-s} f(y-\Delta), \quad (3)$$

e suas integrais, dadas por $F_{m,k}(x)$ e $G_{n,q}(x)$, correspondem a uma função distribuição acumulada (David, 1981).

Para testar a hipótese $H_0: \Delta = 0$ contra $H_1: \Delta > 0$ Bohn e Wolfe (1990, 1992) elaboraram uma estatística análoga à estatística de MWW para ACO que é definida como

$$\begin{aligned} U_{RSS} &= mnkq \int_{-\infty}^{\infty} F_{m,k}(x) dG_{n,q}(x) = \\ &= \sum_{s=1}^q \sum_{t=1}^n \sum_{i=1}^k \sum_{j=1}^m \Psi(Y_{ts(s)} - X_{ji(i)}) = \\ &= (\# \text{ de } X\text{'s} \leq Y\text{'s} \text{ nas amostras de conjuntos ordenados}) \end{aligned} \quad (4)$$

Rejeita-se H_0 para valores grandes de U_{RSS} .

Sob H_0 , as probabilidades dos vários arranjos das $mk + nq$ observações combinadas não dependem da forma da distribuição subjacente comum $F \equiv G$ se ela for contínua, tal como o teste para uma amostra aleatória simples. Entretanto, os $mk + nq$ arranjos não são igualmente prováveis devido ao fato de que os elementos não são igualmente distribuídos, pois são estatísticas de ordem.

Bohn and Wolfe (1990, 1992) mostraram que $\gamma = \sum_{i=1}^k \sum_{s=1}^q P(X_{li(i)} < Y_{1s(s)})$ pode ser estimado sem vício pela estatística

$$U(\mathbf{X}_1, \dots, \mathbf{X}_m; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = \frac{1}{mn} \sum_{s=1}^q \sum_{t=1}^n \sum_{i=1}^k \sum_{j=1}^m \Psi(Y_{TS(s)} - X_{ji(i)}) = \frac{U_{RSS}}{mn}, \quad (5)$$

com esperança e variância dadas respectivamente por

$$E(U_{RSS}) = E \left[\sum_{s=1}^q \sum_{t=1}^n \sum_{i=1}^k \sum_{j=1}^m \Psi(Y_{ts(s)} - X_{ji(i)}) \right] = mn\gamma$$

$$\begin{aligned} Var(U_{RSS}) &= Var \left[\sum_{s=1}^q \sum_{t=1}^n \sum_{i=1}^k \sum_{j=1}^m \Psi(Y_{ts(s)} - X_{ji(i)}) \right] = \\ &= mn[(n-1)\zeta_{1,0} + (m-1)\zeta_{0,1} + \zeta_{1,1}], \end{aligned}$$

onde

$$\begin{aligned} \zeta_{1,0} &= Cov(v(\mathbf{X}_1, \mathbf{Y}_1), v(\mathbf{X}_1, \mathbf{Y}_2)), \\ \zeta_{0,1} &= Cov(v(\mathbf{X}_1, \mathbf{Y}_1), v(\mathbf{X}_2, \mathbf{Y}_1)) \text{ e} \\ \zeta_{1,1} &= Var(v(\mathbf{X}_1, \mathbf{Y}_1)) \end{aligned}$$

As expressões de $\zeta_{1,0}$, $\zeta_{0,1}$ e $\zeta_{1,1}$ são bastante intrincadas e estão em Bohn and Wolfe (1990; 1992).

Ao desenvolver as propriedades assintóticas da forma padronizada de U_{RSS} , Bohn e Wolfe (1990; 1992) consideraram os tamanhos k e q dos conjuntos fixos, uma vez que as dificuldades de ordenação crescem com o aumento do tamanho dos conjuntos. Sejam, então, $N = m + n$ e $\lambda = \lim_{N \rightarrow \infty} (m/N)$. Se $0 < \lambda < 1$ e $\zeta_{1,0}/\lambda + \zeta_{0,1}/(1-\lambda) > 0$, então, $\sqrt{N} (U_{RSS}/mn - \gamma) = \sqrt{N} (U_{RSS} - E[U_{RSS}])/mn$ tem uma distribuição normal assintótica, no sentido de $\min(m, n) \rightarrow \infty$, com média 0 e variância:

$$\sigma_{\infty}^2 = \frac{\zeta_{1,0}}{\lambda} + \frac{\zeta_{0,1}}{(1-\lambda)}, \quad (6)$$

Bohn e Wolfe (1990) mostram também que sob H_0 a distribuição de U_{RSS} é simétrica em torno de $E_{H_0}(U_{RSS}) = mnkq \frac{1}{2}$ e, conseqüentemente, $\sqrt{N} (U_{RSS}/mn - kq/2)$ tem uma distribuição assintótica normal com média 0 e variância dada em (6).

Para conduzir o teste de hipótese de $H_0: \Delta = 0$ baseado na estatística U_{RSS} , Bohn e Wolfe (1990) compararam valores críticos encontrados por aproximação assintótica e pela distribuição exata. Como U_{RSS} tem distribuição livre de parâmetros sob H_0 , para se obter a função de distribuição de U_{RSS} para quaisquer m , n , q e k basta avaliar U_{RSS} para cada possível arranjo ordenado das $mk + nq$ observações combinadas $X_{j1(1)}, \dots, X_{jk(k)}$, para $j = 1, \dots, m$, e $Y_{t1(1)}, \dots, Y_{tq(q)}$, para $t = 1, \dots, n$, calcular as probabilidades de cada arranjo e tabular os resultados. Como os arranjos não são igualmente prováveis esses cálculos são relativamente trabalhosos.

Os valores críticos para o teste proposto também poderiam ser obtidos pela distribuição assintótica da estatística U_{RSS} padronizada. Para avaliar a precisão dessa aproximação normal, Bohn e Wolfe (1990) simularam a distribuição nula de U_{RSS} para

$k = q = 2$ e $m = n = 3, 6$ e 12 e verificaram que a aproximação é bastante boa para $m = n = 6$ ou 12 e não tão ruim para $m = n = 3$.

Bohn e Wolfe (1990; 1992) ainda compararam o poder do teste baseado na estatística U_{RSS} com o poder do teste baseado na estatística de MWW para amostras aleatórias simples. Eles constataram que o teste baseado na estatística U_{RSS} , com um número especificado de observações mensuradas é consideravelmente mais poderoso que o teste correspondente para amostras aleatórias simples com o mesmo número de observações.

Na próxima seção discute-se o método de reamostragem *bootstrap* e em seguida são apresentados a proposta de um teste *bootstrap* para duas amostras obtidas por conjuntos ordenados e um estudo por simulação de seu poder e discutidos os resultados obtidos.

4 O método *Bootstrap*

O *bootstrap* (Efron, 1979) é um método de reamostragem computacionalmente intensivo para inferência estatística, originalmente desenvolvido para fornecer medidas de vício e variabilidade de estimativas. Posteriormente ele foi estendido para a construção de intervalos de confiança, testes de hipóteses e situações inferenciais mais complexas (Efron e Tibishrani, 1993). Sua vantagem está no fato de não ser necessário o conhecimento da distribuição da variável do estimador do parâmetro de interesse ou não depender de aproximações assintóticas, algumas vezes inadequadas para o tamanho da amostra que se dispõe. Para algumas situações, os resultados obtidos pelos métodos *bootstrap* são melhores do que os que seriam obtidos por aproximações assintóticas. Na sua execução, o esforço analítico é, em geral, menor do que o necessário para se obter a distribuição exata ou aproximada do estimador de interesse.

O problema aqui abordado envolve duas variáveis aleatórias, X e Y , com distribuições de probabilidade respectivamente $F(x)$ e $G(y) = F(y - \Delta)$. A hipótese nula a ser testada é $H_0: \Delta = 0$ e a hipótese alternativa é $H_1: \Delta > 0$. Seja $t(\mathbf{x}, \mathbf{y})$ o valor de uma estatística de teste t para uma amostra aleatória simples observada, $(\mathbf{x}, \mathbf{y}) = \mathbf{z}$. Sejam Z^* uma variável aleatória com distribuição J_0 especificada pela hipótese nula e \mathbf{z}^* uma amostra observada de Z^* . Seja $t_{J_0, \alpha}$ o percentil $(1 - \alpha)$ da função de distribuição D_{J_0} de $t(\mathbf{z}^*) = t_{J_0}$. Então, o teste de hipótese Υ envolvido é do tipo: Rejeite-se $H_0: \Delta = 0$ em favor de $H_1: \Delta > 0$ com probabilidade $\varphi_{\Upsilon}(\mathbf{x}, \mathbf{y})$, onde $\varphi_{\Upsilon}(\mathbf{x}, \mathbf{y})$ é a função crítica definida por

$$\varphi_{\Upsilon}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{se } (\mathbf{x}, \mathbf{y}) \in \{(\mathbf{x}, \mathbf{y}) : t(\mathbf{x}, \mathbf{y}) > t_{J_0, \alpha} = D_{J_0}^{-1}(1 - \alpha)\} \\ 0, & \text{caso contrário} \end{cases}, \quad (7)$$

O método *bootstrap* é útil nesse caso para fornecer uma aproximação à distribuição de t sob a hipótese nula (Hall e Wilson, 1991; Efron e Tibishirani, 1993). A preocupação básica ao se utilizar o *bootstrap*, além da escolha da estatística de teste, é formular um

mecanismo de reamostragem que leve à obtenção da distribuição da estatística de teste sob a hipótese nula.

Em testes de hipóteses para duas amostras, a distribuição da estatística de teste sob a hipótese nula pode ser obtida pelo do seguinte procedimento. Primeiro denota-se por \mathbf{z} a união das amostras aleatórias simples $\mathbf{x} = (x_1, x_2, \dots, x_m)$ de X e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ de Y e define-se a função de distribuição empírica de \mathbf{z} como \hat{J}_0 . Sob H_0 , \hat{J}_0 é uma estimativa não-paramétrica da distribuição comum J_0 que originaria tanto \mathbf{x} quanto \mathbf{y} . Então, substitui-se J_0 por \hat{J}_0 em (7), obtendo-se a função crítica

$$\varphi_\gamma(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{se } (\mathbf{x}, \mathbf{y}) \in \{(\mathbf{x}, \mathbf{y}) : t(\mathbf{x}, \mathbf{y}) > t_{J_0, \alpha} = D_{J_0}^{-1}(1 - \alpha)\} \\ 0 & \text{caso contrário} \end{cases},$$

onde $t_{\hat{J}_0, \alpha}$ é a estimativa *plug-in* do ponto que define a região crítica do teste. O método *bootstrap* é então empregado para fornecer uma aproximação à estimativa *plug-in* $t_{\hat{J}_0, \alpha}$ segundo o Algoritmo 1.

Algoritmo 1

- 1) Denota-se por \mathbf{z} a união das amostras observadas $\mathbf{x} = (x_1, x_2, \dots, x_m)$ de X e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ de Y e por \hat{J}_0 a distribuição empírica de \mathbf{z} . Sob H_0 , \hat{J}_0 proporciona uma estimativa não-paramétrica da distribuição comum que originou tanto \mathbf{x} quanto \mathbf{y} .
- 2) Substitui-se J_0 por \hat{J}_0 em (7) e obtêm-se B amostras *bootstrap*, denotadas por \mathbf{z}_b^* , $b = 1, \dots, B$, de tamanho $n + m$ de \hat{J}_0 .
- 3) Em cada amostra *bootstrap* \mathbf{z}_b^* , chamam-se as m primeiras observações de \mathbf{x}_b^* e as n restantes de \mathbf{y}_b^* , $b = 1, \dots, B$.
- 4) Avalia-se $t(\cdot)$ em cada amostra *bootstrap*: $t(\mathbf{z}_b^*) = t(\mathbf{x}_b^*, \mathbf{y}_b^*)$, $b = 1, \dots, B$.
- 5) Rejeita-se H_0 se $t(\mathbf{z}) = t(\mathbf{x}, \mathbf{y}) > \hat{t}_{\hat{J}_0, \alpha} = \hat{D}_{\hat{J}_0}^{-1}(1 - \alpha)$, onde $\hat{t}_{\hat{J}_0, \alpha}$ é o percentil amostral $(1 - \alpha)$ de $\{t(\mathbf{z}_1^*), t(\mathbf{z}_2^*), \dots, t(\mathbf{z}_B^*)\}$, ou seja, $\hat{t}_{\hat{J}_0, \alpha}$ é tal que

$$\hat{P}(t_{\hat{J}_0} > t_{\hat{J}_0, \alpha}) = \frac{\#\{t(\mathbf{z}_b^*) > \hat{t}_{\hat{J}_0, \alpha}\}}{B} = \alpha, \quad (8)$$

5 Testes de hipóteses *bootstrap* para duas amostras obtidas por meio de conjuntos ordenados

O algoritmo para a realização de testes *bootstrap* pressupõe que as observações formam uma amostra aleatória simples. Nesta seção apresenta-se uma modificação do algoritmo discutido anteriormente e desenvolvido por Magro (2003) para aplicação em amostras obtidas a partir de conjuntos ordenados. Introduce-se também as estatísticas de teste a serem utilizadas no experimento Monte Carlo.

5.1 O algoritmo *bootstrap* para amostras obtidas por meio de conjuntos ordenados

Pela seção 2, uma amostra de tamanho km obtida a partir de conjuntos ordenados em k postos de ordenação pode ser vista como a união de amostras aleatórias simples de tamanho m de cada um dos k postos, já que as observações em cada posto são independentes e identicamente distribuídas. As observações de um posto são independentes das observações dos outros postos, mas não têm a mesma distribuição. Um procedimento *bootstrap* para amostras obtidas por meio de conjuntos ordenados deve levar em conta esses aspectos.

A nova proposta, elaborada por Magro (2003), modifica o algoritmo *bootstrap* de forma que o processo de reamostragem seja feito independentemente nas amostras de cada posto, onde os elementos são independentes e identicamente distribuídos. Para isso, é necessário estender a notação introduzida na Seção 4.

Sejam as amostras aleatórias simples de m elementos de cada um dos k postos considerados de X , $\mathbf{X}_{(1)}, \dots, \mathbf{X}_{(k)}$, onde $\mathbf{X}_{(i)} = (X_{1i(i)}, \dots, X_{mi(i)})$, para $i = 1, \dots, k$. Sejam também as amostras aleatórias simples de n elementos de cada um dos q postos considerados de Y , $\mathbf{Y}_{(1)}, \dots, \mathbf{Y}_{(q)}$, onde $\mathbf{Y}_{(s)} = (Y_{1s(s)}, \dots, Y_{ns(s)})$, para $s = 1, \dots, q$.

Para $q = k$, a idéia básica da reamostragem *bootstrap* para o teste de duas amostras obtidas a partir de conjuntos ordenados aqui desenvolvido é concatenar para cada posto as amostras de X e de Y e realizar o procedimento de reamostragem em cada posto.

O procedimento envolve diversos passos. Primeiro, denotam-se por $\mathbf{z}_{(1)}, \dots, \mathbf{z}_{(k)}$ as amostras observadas combinadas $(\mathbf{x}_{(1)}, \mathbf{y}_{(1)}), \dots, (\mathbf{x}_{(k)}, \mathbf{y}_{(k)})$ e por $\hat{J}_{(1)0}, \dots, \hat{J}_{(k)0}$ suas respectivas funções de distribuição empíricas. Sob H_0 , $\hat{J}_{(i)0}$ proporciona uma estimativa não-paramétrica da distribuição comum $J_{(i)0}$ que dá origem a $\mathbf{x}_{(i)}$ e $\mathbf{y}_{(i)}$, para $i = 1, \dots, k$. A seguir, substitui-se J_0 , a distribuição de probabilidades que, sob a hipótese nula, gera tanto as observações de X quanto as de Y , dada em (7), pela estimativa \hat{J}_0 dada pela média das funções de distribuição empírica dos k postos, segundo a propriedade enunciada em (1), ou seja, substitui-se J_0 por $\hat{J}_0 = \frac{1}{k} \sum_{i=1}^k \hat{J}_{(i)0}$. Com isso, obtém-se a função crítica

$$\varphi'_y(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{se } (\mathbf{x}, \mathbf{y}) \in \left\{ (\mathbf{x}, \mathbf{y}) : t(\mathbf{x}, \mathbf{y}) > t_{\frac{1}{k} \sum_{i=1}^k \hat{J}_{(i)0}, \alpha} = D_{\frac{1}{k} \sum_{i=1}^k \hat{J}_{(i)0}}^{-1}(1 - \alpha) \right\}, \\ 0, & \text{caso contrário} \end{cases}$$

onde $t_{\frac{1}{k} \sum_{i=1}^k \hat{J}_{(i)0}, \alpha}$ é a estimativa *plug-in* do ponto que define a região crítica do teste. O método *bootstrap* é então empregado para fornecer uma aproximação à estimativa *plug-in* $t_{\frac{1}{k} \sum_{i=1}^k \hat{J}_{(i)0}, \alpha}$. Desse modo, o algoritmo *bootstrap* para duas ACO pode ser formulado por:

Algoritmo 2

- 1) Denotam-se por $\mathbf{z}_{(1)}, \dots, \mathbf{z}_{(k)}$ as amostras observadas combinadas $(\mathbf{x}_{(1)}, \mathbf{y}_{(1)}), \dots, (\mathbf{x}_{(k)}, \mathbf{y}_{(k)})$ e por $\hat{J}_{(1)0}, \dots, \hat{J}_{(k)0}$ suas respectivas distribuições empíricas. Sob H_0 , $\hat{J}_{(i)0}$ proporciona uma estimativa não-paramétrica da distribuição comum que originaria tanto $\mathbf{x}_{(i)}$ quanto $\mathbf{y}_{(i)}$, para $i = 1, \dots, k$.
- 2) Substitui-se J_0 por $\hat{J}_0 = \frac{1}{k} \sum_{i=1}^k \hat{J}_{(i)0}$ em (8) e, para cada $i = 1, \dots, k$, obtêm-se B amostras *bootstrap*, denotadas por $\mathbf{z}_{(i)b}^*$, $b = 1, \dots, B$, de tamanho $n + m$ de $\hat{J}_{(i)0}$.
- 3) Para cada $i = 1, \dots, k$, chamam-se as m primeiras observações de $\mathbf{z}_{(i)b}^*$ de $\mathbf{x}_{(i)b}^*$ e as n restantes de $\mathbf{y}_{(i)b}^*$, $b = 1, \dots, B$.
- 4) Para $b = 1, \dots, B$, combinam-se as amostras $\mathbf{x}_{(1)b}^*, \dots, \mathbf{x}_{(k)b}^*$ em uma única amostra \mathbf{x}_b^* e as amostras $\mathbf{y}_{(1)b}^*, \dots, \mathbf{y}_{(k)b}^*$ em uma única amostra \mathbf{y}_b^* e avalia-se $t(\cdot)$ em cada amostra *bootstrap*: $t(\mathbf{z}_b^*) = t(\mathbf{x}_b^*, \mathbf{y}_b^*)$,
- 5) Rejeita-se H_0 se $t(\mathbf{z}) = t(\mathbf{x}, \mathbf{y}) > \hat{t}_{\hat{J}_0, \alpha} = \hat{D}_{\hat{J}_0}^{-1}(1 - \alpha)$, onde $\hat{t}_{\hat{J}_0, \alpha}$ é o percentil amostral $(1 - \alpha)$ de $\{t(\mathbf{z}_1^*), t(\mathbf{z}_2^*), \dots, t(\mathbf{z}_B^*)\}$, ou seja, $\hat{t}_{\hat{J}_0, \alpha}$ é tal que

$$\hat{P}(t_{\hat{J}_0} > t_{\hat{J}_0, \alpha}) = \frac{\#\{t(\mathbf{z}_b^*) > \hat{t}_{\hat{J}_0, \alpha}\}}{B} = \alpha. \quad (9)$$

Se $q \neq k$ não é possível utilizar a extensão do *bootstrap* apresentada acima para testar hipóteses para duas ACO. Isso ocorre porque não seria possível concatenar as amostras de X e de Y em todos os postos, impossibilitando a obtenção de funções de distribuição empíricas para cada posto que sigam a hipótese nula.

5.2 Estatísticas de teste analisadas

Diferentes estatísticas são analisadas por Magro (2003) para avaliar o comportamento do teste baseado em reamostragem *bootstrap* para testar a hipótese de que a diferença entre os parâmetros de locação das funções de distribuição de X e de Y é nula.

A primeira é a estatística U_{RSS} definida em (4), que conta o número de vezes em que cada elemento da amostra de Y é maior ou igual que cada elemento de X . No algoritmo 2 as amostras de cada posto i , $i = 1, \dots, k$, de X e de Y são combinadas em uma única amostra $\mathbf{z}_{(i)}$ e de cada uma são obtidas amostras com reposição $\mathbf{z}_{(i)b}^*$, $b = 1, \dots, B$. Em seguida, para cada $i = 1, \dots, k$, chamam-se as m primeiras observações de $\mathbf{z}_{(i)b}^*$ de $\mathbf{X}_{(i)b}^*$ e as n restantes de $\mathbf{Y}_{(i)b}^*$, $b = 1, \dots, B$. Assim, as amostras $\mathbf{X}_{(i)b}^*$ e $\mathbf{Y}_{(i)b}^*$ podem conter elementos de mesmo valor, aqui chamados de "empates", devido à amostragem com reposição.

Para cada empate na replicação *bootstrap* da amostra, a definição de U_{RSS} dada em (4) adiciona uma unidade à replicação *bootstrap* correspondente da estatística. Esse fato causa um aumento no valor de cada replicação *bootstrap* da estatística e toda a sua distribuição *bootstrap* é deslocada, assim como o percentil $\hat{t}_{j_0, \alpha}$ que define a região crítica do teste, para a direita. Portanto a probabilidade de rejeição da hipótese nula é sistematicamente reduzida. Se U_{RSS} fosse definida como o número de vezes em que cada elemento da amostra de Y é maior que cada elemento de X , ou seja, se os empates fossem desconsiderados, um efeito contrário ao relatado anteriormente ocorreria: a distribuição *bootstrap* é deslocada para a esquerda aumentando sistematicamente tanto o poder quanto o tamanho do teste. Esse fato foi observado nos resultados de Guo (1999), para amostras aleatórias simples.

Para lidar com essa situação decorrente do processo de reamostragem, adicionou-se a cada elemento da amostra de Y um distúrbio seguindo distribuição Normal com média 0 e variância 10^{-4} antes de se extrair cada amostra *bootstrap*. Com esse procedimento, os empates foram evitados sem que a magnitude das observações fosse substancialmente afetada.

As duas outras estatísticas de teste avaliadas, denotadas por T_{mdn} e T_{md} , baseiam-se em diferenças entre medidas amostrais de posição. T_{mdn} é a diferença entre as medianas de Y e de X . T_{md} é a diferença entre as médias de Y e de X . Sob H_0 , os valores observados de ambas estatísticas devem ser próximos a zero e sob H_1 devem assumir valor maiores que zero.

As expressões (8) e (9) especificam que, para o nível de significância fixado, a hipótese nula deve ser rejeitada sempre que o valor da estatística de teste calculado na amostra original for maior que o percentil amostral $(1 - \alpha)$ da distribuição *bootstrap* da estatística. Além do mais as distribuições das estatísticas baseadas na estatística de MWW são discretas (por exemplo, a estatística U_{RSS} assume valores inteiros de 0 a 36 para configurações com seis elementos em cada amostra) e as demais estatísticas, pelo uso do método *bootstrap*, assumem um número limitado de valores para configurações com número pequeno de elementos. Portanto, o valor das estatísticas calculado a partir amostras originais pode coincidir com o valor do percentil amostral $(1 - \alpha)$ das

distribuições *bootstrap* com uma probabilidade não desprezível, afetando a obtenção do tamanho do teste desejado.

Para favorecer a obtenção do tamanho do teste desejado, adota-se um teste aleatorizado definido pela seguinte função crítica

$$\varphi''_r(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{se } (\mathbf{x}, \mathbf{y}) \in \{(\mathbf{x}, \mathbf{y}) : t(\mathbf{x}, \mathbf{y}) > t_{j_0, \alpha} = D_{j_0}^{-1}(1 - \alpha)\} \\ p & \text{se } (\mathbf{x}, \mathbf{y}) \in \{(\mathbf{x}, \mathbf{y}) : t(\mathbf{x}, \mathbf{y}) = t_{j_0, \alpha} = D_{j_0}^{-1}(1 - \alpha)\} \\ 0 & \text{c.c.} \end{cases}$$

onde p é tal que

$$E[\varphi''_r(\mathbf{x}, \mathbf{y}) | H_0] = 1 \cdot P(t_{j_0} > t_{j_0, \alpha}) + p \cdot P(t_{j_0} = t_{j_0, \alpha}) = \alpha,$$

sendo estimado por \hat{p} , obtido de forma que

$$\hat{E}[\varphi''_r(\mathbf{x}, \mathbf{y}) | H_0] = 1 \cdot \frac{\#\{t(\mathbf{z}_b^*) > \hat{t}_{j_0, \alpha}\}}{B} + \hat{p} \cdot \frac{\#\{t(\mathbf{z}_b^*) = \hat{t}_{j_0, \alpha}\}}{B} = \alpha.$$

A seguir é descrito o experimento Monte Carlo implementado por Magro (2003) para avaliar as funções poder para os testes de hipóteses *bootstrap* realizados a partir de cada uma das estatísticas de teste apresentadas acima.

6 Experimento Monte Carlo para estudo do poder do teste *bootstrap*

Ao estudar o desempenho do método de reamostragem *bootstrap* para o teste de hipótese de duas amostras obtidas por meio de conjuntos ordenados e compará-lo com o teste de Bohn e Wolfe (1990, 1992), Magro (2003) analisou a função poder por um experimento de simulação Monte Carlo para as mesmas configurações por eles estudadas. A simulação foi implementada pela linguagem R (Ihaka e Gentleman, 1996).

As observações de X foram geradas segundo as distribuições Uniforme (0, 1), Cauchy (0, 1), Normal (0, 1) e Exponencial (1). As observações de Y foram geradas por essas mesmas distribuições, adicionando-se a cada um dos valores gerados o valor do parâmetro de deslocamento Δ . Para a distribuição Uniforme, variou-se Δ entre 0 e 0,5, inclusive, em incrementos de 0,1. Para as distribuições Exponencial, Cauchy e Normal, variou-se Δ entre 0 e 1, inclusive, em incrementos de 0,1.

As configurações amostrais consideradas, dadas em termos de (m, n, k, q) , foram: (6, 6, 1, 1), (3, 3, 2, 2), (2, 2, 3, 3), (12, 12, 1, 1), (6, 6, 2, 2), (4, 4, 3, 3), (24, 24, 1, 1), (12, 12, 2, 2), (8, 8, 3, 3). As configurações (6, 6, 1, 1), (12, 12, 1, 1), (24, 24, 1, 1) são amostras aleatórias simples, que podem ser consideradas casos particulares de amostras obtidas de conjuntos ordenados com apenas um posto amostrado. A inclusão dessas configurações amostrais foi feita para se avaliar o efeito sobre o poder do teste resultante da utilização de amostras de conjuntos ordenados em vez de amostras aleatórias simples com o mesmo número de elementos.

Para cada uma das configurações amostrais, distribuições e valores de Δ foram geradas aleatoriamente 1.000 amostras de X e de Y . De cada uma dessas amostras foram obtidas 999 amostras *bootstrap* balanceadas. Deve ser salientado que, para as configurações de amostras de conjuntos ordenados, as amostras *bootstrap* foram obtidas separadamente de cada posto amostrado. O poder do teste para cada estatística foi calculado, de acordo com o método Monte Carlo, como a proporção de rejeições da hipótese nula entre os 1.000 testes correspondentes às 1.000 amostras de X e de Y simuladas, ou seja, $\pi(\Delta) = R_{H_0}/1000$, onde R_{H_0} é o número de rejeições da hipótese nula.

O nível de significância de todos os testes foi especificado como 0,05. Por fim, o tamanho do teste foi calculado como a proporção de rejeições da hipótese nula para as amostras simuladas de acordo com $\Delta = 0$, isto é, quando a hipótese nula é verdadeira.

As estimativas do valor da função poder associada a cada valor do parâmetro de deslocamento Δ , a cada uma das configurações amostrais e a cada uma das estatísticas de teste estudadas estão apresentadas graficamente na Figura 1 para a distribuição Exponencial, na Figura 2 para a distribuição Normal, na Figura 3 para a distribuição Uniforme e na Figura 4 para a distribuição Cauchy.

As estimativas das funções poder estão apresentadas em gráficos de linhas e as estimativas dos tamanhos dos testes estão apresentadas em gráficos de barras. Para todas as distribuições e estatísticas de teste, dado cada tamanho de amostra, observa-se que a função poder aumenta conforme aumenta o número de postos amostrados. Entretanto, ocorrem algumas discrepâncias entre o tamanho e o nível de significância, fixado em 0,05, de alguns dos testes.

O experimento Monte Carlo aqui realizado, envolvendo 1.000 repetições de um teste com tamanho igual a 0,05, pode ser visto como um experimento Binomial com 1.000 ensaios de Bernoulli, em que sucesso é definido como a rejeição da hipótese nula quando ela é correta e a probabilidade de sucesso é igual a 0,05. Assim, se um teste tiver tamanho igual a 0,05, em um experimento Monte Carlo envolvendo 1.000 repetições do teste seriam esperados com 90% de probabilidade de 39 a 62 repetições rejeitando a hipótese nula quando ela é correta. Portanto, tamanhos de teste entre 0,039 e 0,062 obtidos por simulação Monte Carlo fornecem evidências significativas de que o tamanho real do teste não é muito diferente do nível de significância fixado em 0,05.

A diferença entre medianas é a estatística que apresenta maiores discrepâncias entre o tamanho e o nível de significância. Essa estatística apresenta tamanho entre 0,039 e 0,062 para a maior parte das configurações amostrais para a distribuição Exponencial. Para a distribuição Cauchy, essa estatística parece proporcionar um teste *bootstrap* com tamanho abaixo dos 0,05 esperados para 11 das 12 configurações amostrais. Para as demais distribuições, o único padrão evidente é que esse teste parece apresentar tamanho superior a 0,05 apenas para a configuração amostral (2,2,3,3).

A diferença entre médias proporcionou um teste em que o tamanho parece ser sempre igual ou maior que 0,05. O tamanho do teste excedeu 0,05 em geral para as configurações com 6 elementos e para algumas configurações com 12 elementos. Portanto, para tamanhos amostrais tão pequenos quanto esses, um teste *bootstrap* envolvendo a diferença entre médias poderá levar a erro Tipo I com maior probabilidade do que o esperado. Para todas as configurações com 24 elementos esse teste apresentou tamanho entre 0,039 e 0,062 e, em geral, bastante próximo a 0,05.

O teste *bootstrap* usando a estatística U é o que apresentou tamanho mais próximo a 0,05 para a maior parte das configurações amostrais e distribuições. Para a distribuição Normal (Figura 2), esse teste apresentou tamanho bastante próximo a 0,05 para todas as configurações amostrais estudadas. Para as distribuições Exponencial (Figura 1) e Uniforme (Figura 2) o nível de significância fixado em 0,05 parece ter sido excedido apenas em configurações com 6 elementos envolvendo ACO, a se julgar pelos tamanhos de teste observados acima do intervalo $[0,039, 0,062]$. Para a distribuição Cauchy (Figura 4), isso ocorreu apenas para a configuração (6,6,2,2).

O teste *bootstrap* envolvendo a estatística U é o que apresentou poder em níveis mais próximos ao teste de Bohn & Wolfe para todas, exceto duas, configurações amostrais envolvendo todas as distribuições estudadas. Para as configurações (2,2,3,3) para as distribuições Exponencial e Cauchy esse teste apresentou uma função poder ligeiramente superior à do teste daqueles autores.

As discrepâncias no tamanho dos testes observadas dificultam as comparações em termos do poder dos testes. Entretanto, alguns padrões podem ser claramente identificados. A diferença entre medianas, além de proporcionar testes com tamanho em geral abaixo de 0,05, parece apresentar, algumas vezes, poder abaixo do poder do teste estudado por Bohn & Wolfe. Apenas para as configurações com 24 elementos e a configuração (3,3,2,2) para a distribuição Cauchy o teste apresentou poder em níveis bastante próximos aos obtidos por aqueles autores. Note-se que para as configurações com 24 elementos envolvendo a distribuição Cauchy, tanto o teste de Bohn & Wolfe quanto os testes *bootstrap* aqui estudados apresentam funções de poder praticamente coincidentes, exceto para $\Delta = 0$. Para as demais distribuições, toda a função poder associada a essa estatística está posicionada abaixo da função poder do teste de Bohn & Wolfe.

A diferença entre médias apresentou padrões distintos entre as distribuições estudadas. Para a distribuição Exponencial, essa estatística proporciona um teste com poder evidentemente abaixo do teste de Bohn & Wolfe e em níveis muito próximos ao teste envolvendo a diferença entre medianas. Para a distribuição Uniforme, o teste envolvendo essa estatística é notavelmente mais poderoso que o teste estudado por aqueles autores para as configurações com 6 elementos e ligeiramente mais poderoso para as configurações com maior número de elementos. Para a distribuição Normal esse efeito é também notado para as configurações envolvendo 6 e 12 elementos.

Conclusões

Uma primeira conclusão básica extraída dos resultados do experimento de simulação para amostragem por conjuntos ordenados é que testes de reamostragem *bootstrap* usando a estatística modificada MWW para a diferença entre médias podem ser empregados com êxito como uma alternativa ao teste proposto por Bohn e Wolfe (1990; 1992). O teste de reamostragem *bootstrap* envolvendo a diferença entre medianas, entretanto, parece não ser um procedimento vantajoso ao teste estudado por aqueles autores.

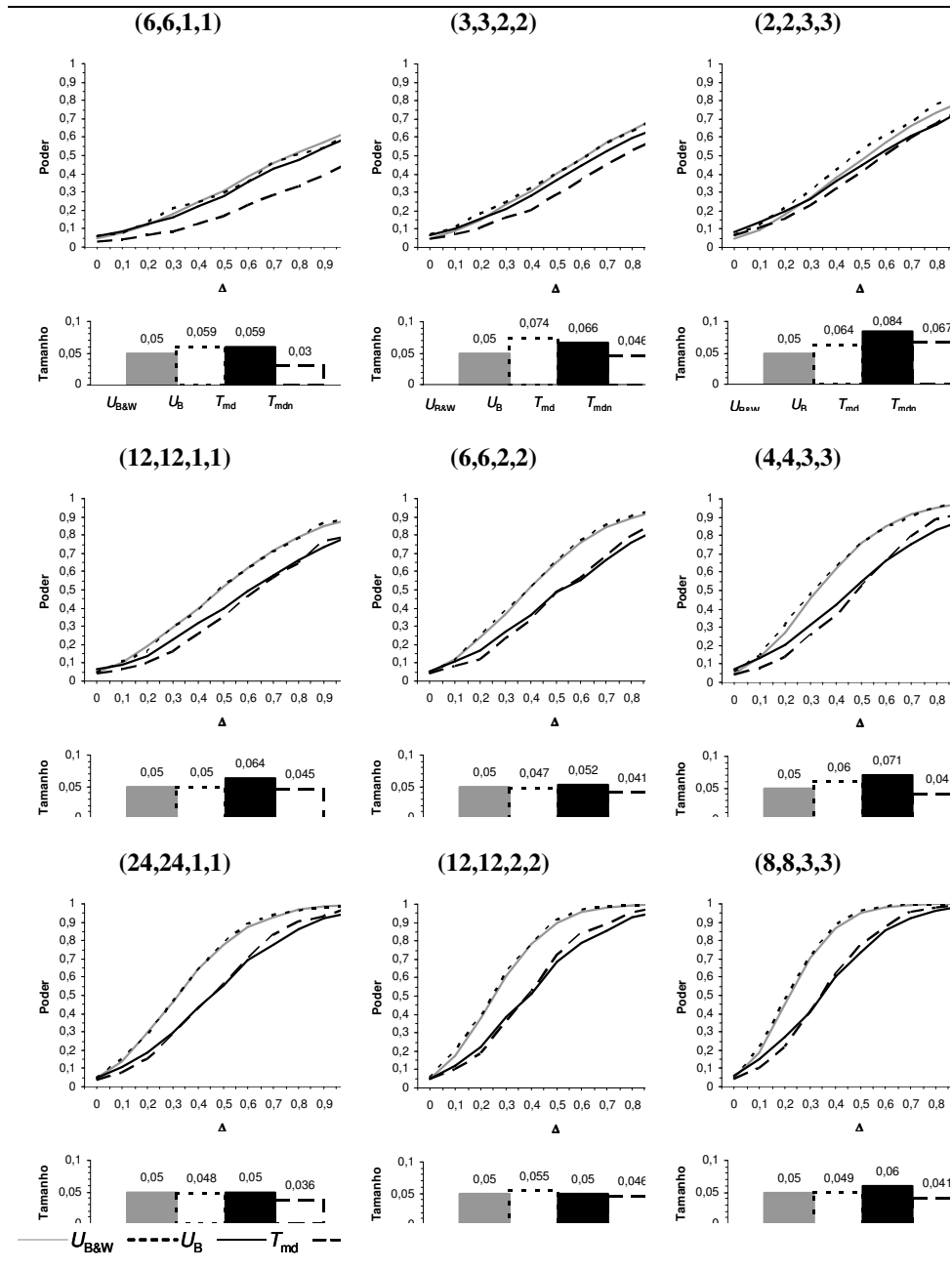


Figura 1 - Gráficos das estimativas das funções poder para cada uma das estatísticas de teste analisadas e alguns valores de Δ para cada configuração amostral (m,n,k,q) para a Distribuição Exponencial.

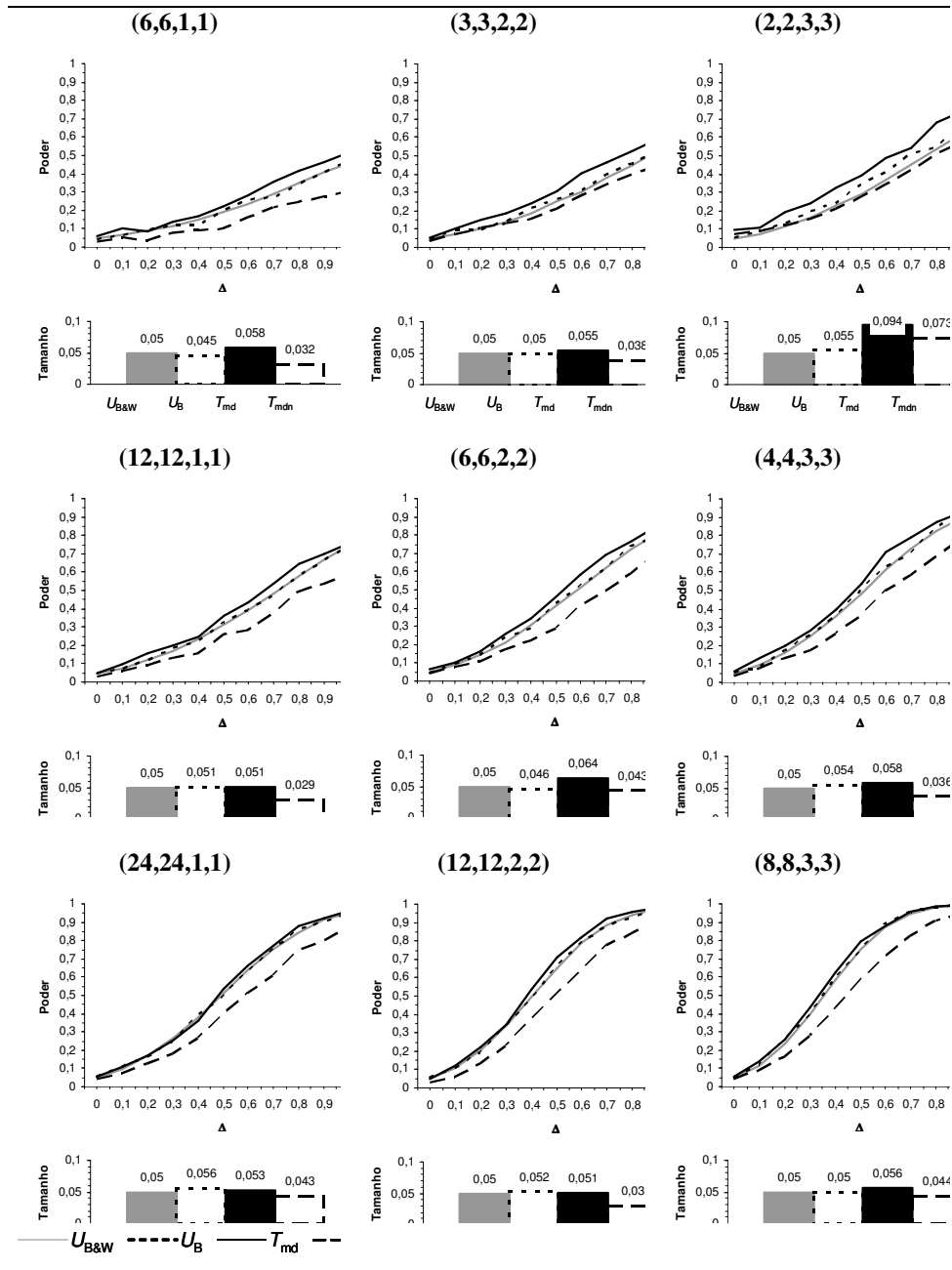


Figura 2 - Gráficos das estimativas das funções poder para cada uma das estatísticas de teste analisadas e alguns valores de Δ para cada configuração amostral (m,n,k,q) para a Distribuição Normal.

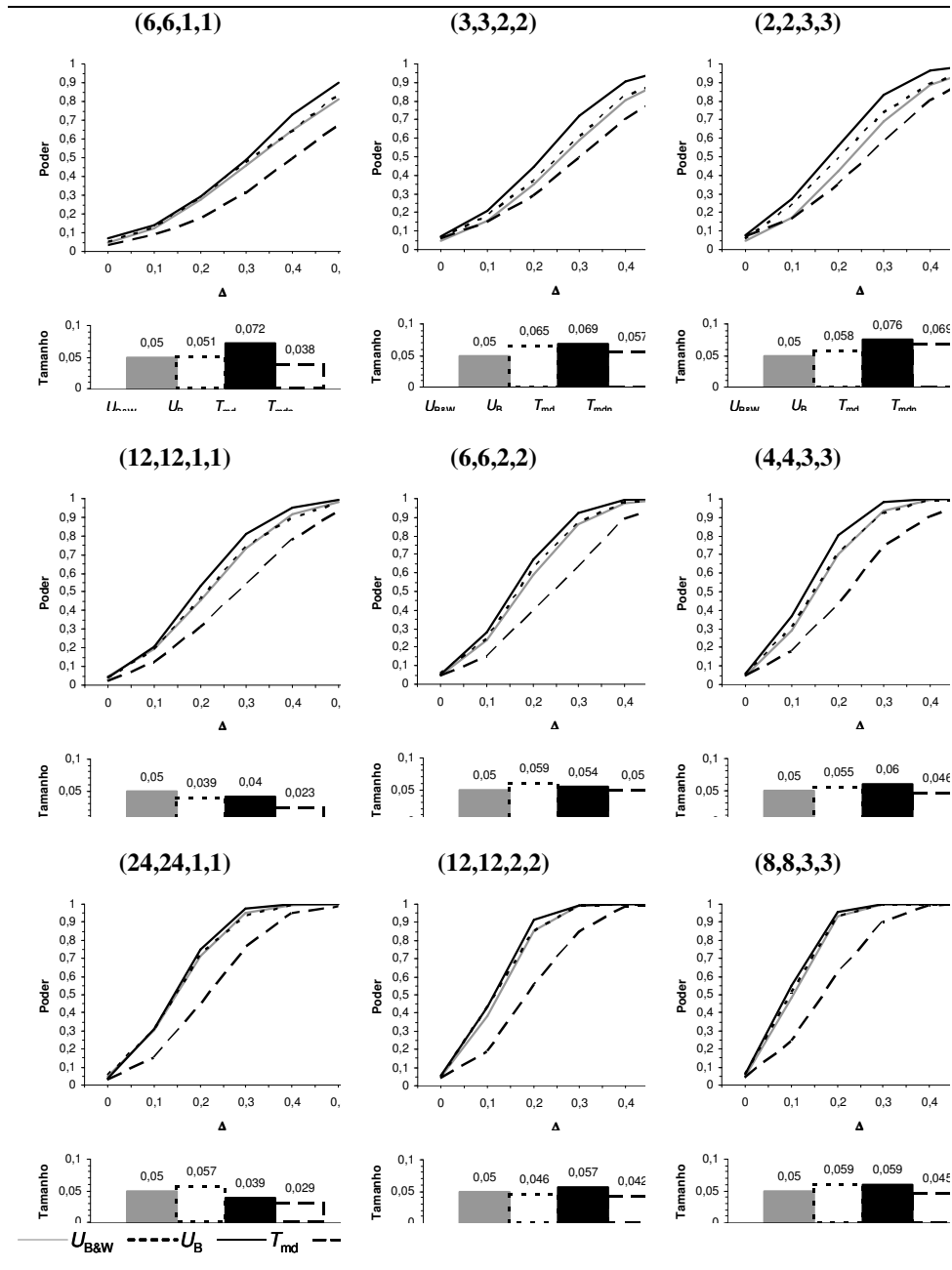


Figura 3 - Gráficos das estimativas das funções poder para cada uma das estatísticas de teste analisadas e alguns valores de Δ para cada configuração amostral (m, n, k, q) para a Distribuição Uniforme.

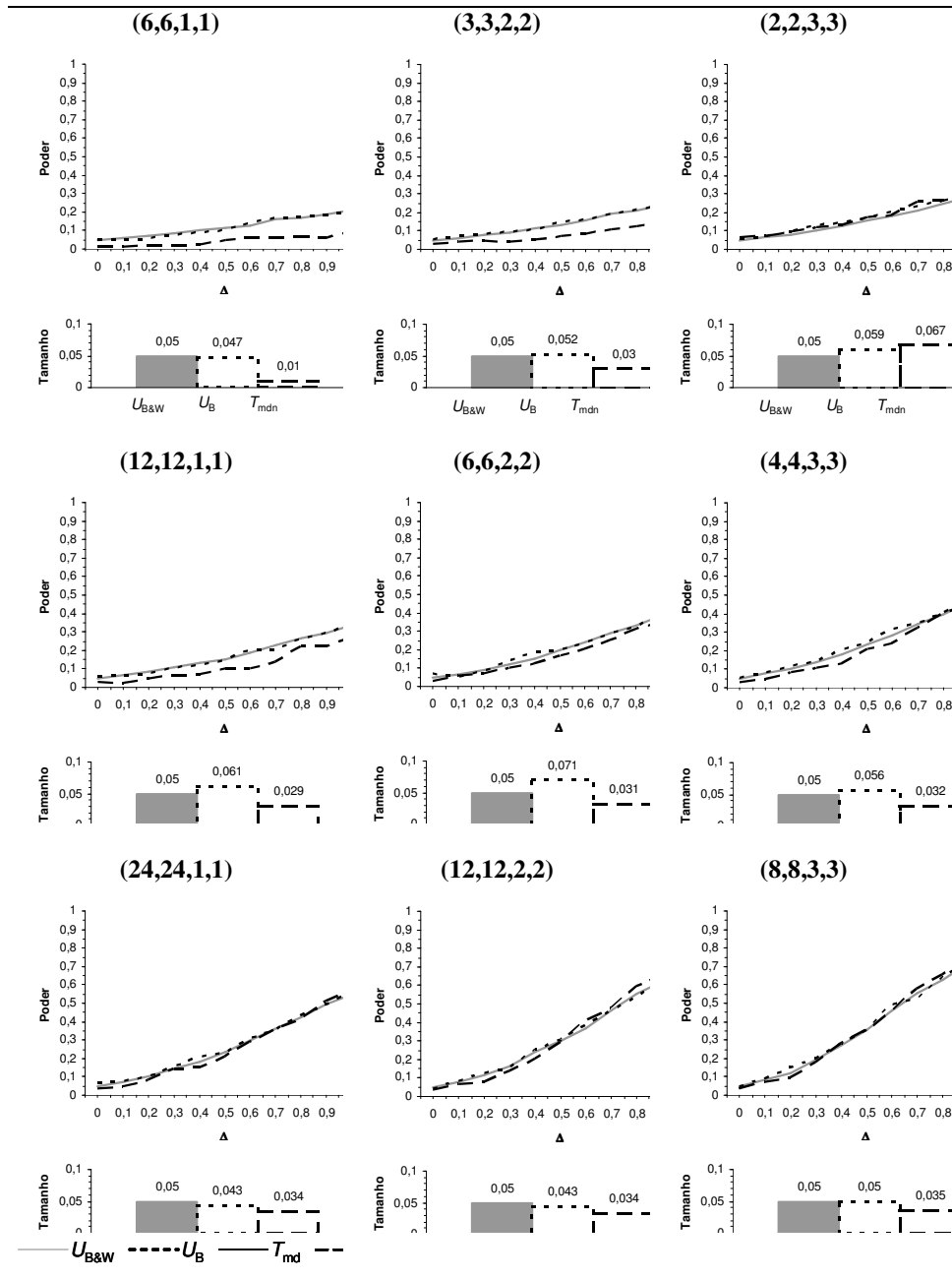


Figura 4 - Gráficos das estimativas das funções poder para cada uma das estatísticas de teste analisadas e alguns valores de Δ para cada configuração amostral (m,n,k,q) para a Distribuição Cauchy.

Considerando-se a estatística de MWW e a diferença entre médias para amostras com pequeno número de elementos, o tamanho do teste apresenta algumas discrepâncias em relação ao tamanho nominal. Como essas discrepâncias não são grandes, em uma situação prática esse efeito pode ser pouco relevante. Além disso a alternativa analítica exigiria esforços eventualmente inviáveis, dada a infactibilidade de se tabelar valores críticos para o teste para todas as possíveis combinações de número de postos amostrados e número de observações em cada posto no esquema de ACO.

Os testes *bootstrap* usando a estatística de MWW apresentaram propriedades de tamanho e poder de teste praticamente iguais ao teste estudado por Bohn & Wolfe para todas as distribuições. Se houver evidências de que as distribuições das variáveis de interesse são normais, um teste *bootstrap* utilizando-se a diferença entre médias como estatística de teste parece proporcionar resultados em geral melhores que os obtidos por Bohn & Wolfe para amostras com menos do que 24 elementos. Para a distribuição Uniforme, o teste parece ser sempre mais poderoso que o teste proposto por aqueles autores, mas haverá menor controle sobre o tamanho do teste se as amostras forem pequenas. Para a distribuição Exponencial esse teste apresentou desempenho significativamente inferior ao teste de Bohn & Wolfe, fato que sugere que o teste *bootstrap* usando a diferença entre médias pode ser desvantajoso para distribuições assimétricas. Para a distribuição Cauchy, as únicas estatísticas recomendadas para um teste *bootstrap*, entre as analisadas, são a estatística de MWW e, se os tamanhos amostrais forem grandes, a diferença entre medianas.

Agradecimentos

Alexandre de Souza Magro contou com o apoio da Capes durante parte do período de desenvolvimento deste trabalho.

MAGRO, A. S.; BARRETO, M. C. M. Bootstrap tests for two independent samples under ranked set sampling. *Rev. Mat. Est.*, São Paulo, v.24, n.1, p.7-27, 2006.

- *ABSTRACT: Ranked set sampling is more efficient than single random sampling for several statistical procedures when the measurement of the main variable is costly or difficult to obtain. Bohn and Wolfe (1990, 1992) presented a modified statistic of the Mann-Whitney-Wilcoxon test to compare two independent samples under a ranked set sample procedure. This paper considers an original implementation of the bootstrap resampling method to compare the location parameters of two independent samples under the ranked set design. A Monte Carlo study is conducted to validate the size and the power of the bootstrap test using the modified statistic of Mann-Whitney-Wilcoxon and the difference of sample means and the difference of sample medians as the statistic test. Our results are quite equivalent in size and power to those of the modified statistic of Bohn and Wolfe. The bootstrap tests for the modified statistic of Mann-Whitney-Wilcoxon can be used as a successful alternative to Bohn and Wolfe's asymptotic test.*
- *KEYWORDS: Ranked set sampling; bootstrap; modified statistic of Mann-Whitney-Wilcoxon test; tests for two independent samples.*

Referências

- BARABESI, L. The computation of the distribution of the sign test statistic for ranked-set sampling. *Commun. Stat. Simul. Comput.*, New York, v.27, p.833-42, 1998.
- BARNETT, V. Ranked set sample design for environmental investigations. *Environ. Ecol. Stat.*, Norwell, v.6, p.59-74, 1999.
- BARNETT, V. MOORE, K. Best linear unbiased estimates in raked-set sampling with particular reference to imperfect ordering. *J. Appl. Stat.*, Abingdon, v.24, p.697-710, 1997.
- BARRETO, M. C. M.; BARNETT, V. Best linear unbiased estimates for the simple linear regression model using ranked-set sampling. *Environ. Ecol. Stat.*, Norwell, v.6, p.119-33, 1999.
- BOHN, L. L.; WOLFE, D. A. *Nonparametric two-sample procedures for ranked-set samples data*. Ohio: Ohio State University, Dept. of Statistics, 1990. 25p. (Technical Report, 447)
- BOHN, L. L.; WOLFE, D. A. Nonparametric two-sample procedures for ranked-set samples data. *J. Am. Stat. Assoc.*, New York, v.81, n.418, p.552-61, 1992.
- BOHN, L. L.; WOLFE, D. A. The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic. *J. Am. Stat. Assoc.*, New York, v.89, p.168-76, 1994.
- CESÁRIO, L. C.; BARRETO, M. C. M. Um estudo sobre o desempenho de intervalos de confiança *bootstrap* para a média de uma distribuição normal usando amostragem por conjuntos ordenados perfeitamente. *Rev. Mat. Estat.*, São Paulo, v.21, p.7-20, 2003.
- CHEN, Z. On ranked-set sample percentiles and their applications. *J. Stat. Plan. Infer.*, Amsterdam, v.83, p.125-35, 2000.
- DAVID, H. A. *Order statistics*. 2nd. ed., New York: John Wiley, 1981. 360p.
- DAVISON, A. C.; HINKLEY, D. V. *Bootstrap methods and their application*. Cambridge: Cambridge University Press, 1997. 582p.
- DELL, T. R.; CLUTTER, J. L. Ranked set sampling theory with order statistics background. *Biometrics*, Washington, v.28, p.545-55, 1972.
- EFRON, B. *Bootstrap methods: another look at the jackknife*. *Ann. Stat.*, Washington, v.7, p.1-26, 1979.
- EFRON, B. TIBSHIRANI, R. *An introduction to the bootstrap*. Nova York: Chapman & Hall, 1993. 436p.
- GARCIA, C. A. *Precisão relativa e intervalos de confiança bootstrap para a variância populacional de uma distribuição normal baseados na amostragem por conjuntos ordenados*, 2004, 127f. Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, São Carlos, 2004.
- GUO, J. H. A nonparametric test for the parallelism of two first-order autoregressive processes. *Aust. New Zeal. J. Stat.*, Oxford, v.41, p.59-65, 1999.

- HALL, P.; WILSON, S. R. Two guidelines for bootstrap hypothesis testing. *Biometrics*, Washington, v.47, p.757-62, 1991.
- HETTMANSPERGER, T. P. The ranked-set sample sign test. *Nonparametric Stat.*, Londres, v.4, p.263-70, 1995.
- HOEFFDING, W. A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.*, An Arbor, v.19, p.293-325, 1948.
- IHAKA, R.; GENTLEMAN, R. R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, Alexandria, v.5, n.3, p.299-314, 1996.
- KAUR, A.; PATIL, G.P.; TAILLIE, C. Unequal allocation models for ranked set sampling with skew distributions. *Biometrics*, Washington, v.53, p.123-32, 1997.
- KOTI, K. M.; BABU, G. J. Sign test for ranked-set sampling. *Commun. Stat. - Theory Methods*, New York, v.27, p.1617-30, 1996.
- MAGRO, A. S. *Testes de hipóteses bootstrap para duas amostras obtidas a partir de conjuntos ordenados*. 2003. 67f. Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, São Carlos, 2003.
- McINTYRE, G. A. A method of unbiased selective sampling, using ranked sets. *Aust. J. Agric. Res.*, Collingwood, v.3, p.385-90, 1952.
- ÖZTÜRK, Ö.; WOLFE, D. A. Alternative ranked set sampling protocols for the sign test. *Stat. Probab. Lett.*, Amsterdam, v.47, p.15-23, 2000.
- PATIL, G. P.; SINHA, A. K., TAILLIE, C. Relative precision of ranked set sampling: comparison with the regression estimator. *Environmetrics*, Chester, v.4, p.399-412, 1993.
- RAMINELLI, J. A. *Procedimentos de estimação intervalar de um quantil populacional em amostragem por conjuntos ordenados*, 2003. 81f. Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, 2003.
- STOKES, S.L. Parametric ranked set sampling. *Ann. Inst. Stat. Math.*, Tokio, v.47, p.465-82, 1995.
- TACONELI, C. A.; BARRETO, M. C. M. Intervalos de confiança para a média populacional usando amostragem por conjuntos ordenados. *Rev. Mat. Estat.*, São Paulo, v.21, p.41-66, 2003.
- TAKAHASI, K.; WAKIMOTO, K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Stat. Math.*, Tokio, v.20, p.1-31, 1968.
- YU, P. L. H.; LAM, K. Regression estimator in ranked set sampling. *Biometrics*, Washington, v.53, p.1070-80, 1997.

Recebido em 16.11.2004.

Aprovado após revisão em 10.12.2005.