

ESTIMAÇÃO INTERVALAR VIA *BOOTSTRAP*

Edson Zangiacomi MARTINEZ¹
Francisco LOUZADA-NETO²

- RESUMO: Em inferência estatística, tem-se interesse na quantificação do erro cometido ao se estimar um parâmetro de interesse θ através de $\hat{\theta}$. Uma estratégia usual para a busca de medidas de incerteza, que expressem este erro, é a estimação do erro padrão de $\hat{\theta}$. Entretanto, métodos analíticos para a obtenção destas medidas nem sempre são disponíveis, ou constituem processos altamente complexos, enquanto métodos assintóticos, nos quais a construção de intervalos de confiança é baseada, dependem de aproximações nem sempre alcançadas. Neste contexto, o método *bootstrap* constitui uma eficiente alternativa, fornecendo estimativas do erro padrão de $\hat{\theta}$ livres de complexidades algébricas e possibilitando a obtenção de intervalos de confiança sem a necessidade de pressupostos sobre a distribuição do estimador. Neste artigo, apresentamos de forma concisa a obtenção do erro padrão de $\hat{\theta}$ através do método *bootstrap*, bem como uma discussão dos diferentes tipos de intervalos de confiança *bootstrap* descritos na literatura. Uma aplicação em dados de um estudo real em oncologia ilustra a utilização do método.
- PALAVRAS-CHAVE: *Bootstrap*, Erro Padrão, Intervalos de Confiança, Técnicas de Simulação

1 Centro de Atenção Integral à Saúde da Mulher, CAISM/UNICAMP, CEP 13083-970, Campinas – SP.

2 Departamento de Estatística, Universidade Federal de São Carlos, Caixa Postal 676, CEP 13565-905, São Carlos – SP.

1 Introdução

Neste artigo, o *bootstrap* e os diferentes tipos de intervalos de confiança construídos através deste método, descritos pela literatura, são apresentados de forma concisa.

Uma breve revisão sobre estimação intervalar e motivações para a utilização de métodos de simulação para a obtenção do erro padrão são apresentadas na Seção 2. O estimador *bootstrap* do erro padrão é descrito na Seção 3, enquanto a Seção 4 descreve uma outra ferramenta empregada para estimar esta medida, o método *jackknife*. Comparações entre estes métodos são consideradas na Seção 5. Diferentes tipos de intervalos de confiança *bootstrap* são apresentados na Seção 6. Na Seção 7, a metodologia é ilustrada através de uma aplicação em dados de um estudo real em oncologia. O artigo é finalizado com alguns comentários, na Seção 8.

2 Uma breve revisão sobre estimação intervalar

Ao usuário de um processo de inferência estatística, uma simples estimativa pontual de um parâmetro θ de seu interesse pode não ser suficiente para fornecer evidências que, de fato, auxiliem em suas deduções. São necessárias também medidas da precisão desta estimativa, que possibilitem ao pesquisador, frente aos enunciados conjecturais de seu estudo, tecer conclusões baseadas em suas observações. As medidas de precisão aqui descritas são o erro padrão e intervalos de confiança.

O erro padrão tem sua definição no desvio padrão de uma estatística, e, conseqüentemente, traduz-se em uma medida da incerteza da estimação baseada nesta estatística. Intervalos de confiança também fornecem uma avaliação do erro que se comete na estimação pontual, sob um outro ponto de vista. De posse de uma amostra aleatória $\mathbf{x} = (x_1, x_2, \dots, x_n)$ composta por n observações, de uma distribuição F , tais intervalos são escritos na forma

$$(L_I(\mathbf{x}); L_S(\mathbf{x})) \quad (1)$$

onde $L_I(\mathbf{x})$ e $L_S(\mathbf{x})$ são chamados, respectivamente, limites inferior e superior do intervalo. Sendo X uma variável aleatória e \mathbf{x} uma obser-

vação amostral de X , a interpretação de (1) torna-se clara na compreensão da expressão

$$P_F [L_I(X) \leq \theta \leq L_S(X)] = 1 - \alpha, \quad 0 \leq \alpha \leq 1, \quad (2)$$

ou seja, dentre um grande número de possíveis realizações de um experimento que objetiva gerar uma estimativa $\hat{\theta}$ de θ , todas considerando amostras de tamanho n , de cada 100, $100(1-\alpha)$ produzirão intervalos de confiança que vão conter θ , o verdadeiro valor do parâmetro.

Ao usuário de inferência estatística, compete o entendimento de que a amostra obtida em seu experimento é apenas uma dentre várias ou infinitas possibilidades. A estimativa $\hat{\theta}$ de θ obtida em sua amostra, poderia ser diferente em alguma outra amostra obtida da mesma população, através do mesmo processo de amostragem. Realizado o experimento uma única vez, o pesquisador tem em mãos apenas uma destas possíveis estimativas $\hat{\theta}$, e daí, naturalmente, surge a indagação: Qual a confiança que podemos depositar nesta estimativa? A quantidade $1-\alpha$ é dita coeficiente de confiança, e auxilia o pesquisador a responder a esta questão. O lado esquerdo de (2) é chamado de probabilidade de cobertura, onde $0 \leq \alpha \leq 1$.

Geralmente, intervalos de confiança são construídos com áreas iguais sob as caudas, ou seja, satisfazendo a

$$P_F [\theta \leq L_I(X)] = \frac{\alpha}{2} \quad \text{e} \quad P_F [\theta \geq L_S(X)] = \frac{\alpha}{2}. \quad (3)$$

Verifica-se que (3) implica (2), mas nem sempre o contrário é verdadeiro.

Intervalos de confiança como o descrito em (2) são chamados exatos, quando a probabilidade em questão é igual a $1-\alpha$, e não uma aproximação. Entretanto, nem sempre é possível construir intervalos com esta propriedade. Os intervalos de confiança mais utilizados são os chamados aproximados, que satisfazem a

$$P_F [L_I(X) \leq \theta \leq L_S(X)] \approx 1 - \alpha. \quad (4)$$

A ferramenta essencial para a construção de intervalos como o descrito em (4) é a teoria assintótica. Sua principal limitação é o fato do pesquisador necessitar de uma amostra grande o suficiente para garantir a validade das aproximações em questão, caso contrário, a eficácia do método está seriamente comprometida.

Intervalos de confiança exatos muitas vezes são construídos através de soluções analíticas nem sempre simples, enquanto intervalos aproximados, como foi dito, dependem de aproximações assintóticas nem sempre alcançadas. Uma ferramenta alternativa, eficiente não só para a construção de intervalos de confiança mas também para estabelecer erros padrão de estimadores de interesse, são os métodos computacionalmente intensivos. Livre de complexidades analíticas, surge neste âmbito o *bootstrap*.

O método *bootstrap* foi proposto por Efron (1979), como uma ferramenta para estimar o erro padrão da estimativa de um parâmetro. Se comparado a outras técnicas estatísticas, o método teve uma gênese um tanto tardia, devido à sua dependência do uso de computadores. Os progressos da informática experimentados nas últimas décadas do século XX possibilitaram a popularização do uso do computador e incrementaram o surgimento e acesso a *softwares* matemáticos e estatísticos. Conseqüentemente, as aplicações de métodos *bootstrap* nas mais diferentes áreas da estatística se intensificaram.

3 O estimador *bootstrap* do erro padrão

Seja uma amostra aleatória baseada em n observações independentes x_1, x_2, \dots, x_n . O erro padrão de uma média \bar{x} baseada nesta amostra é estimado pela expressão,

$$\hat{ep}(\bar{x}) = \sqrt{\frac{s^2}{n}}, \quad (5)$$

onde

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (6)$$

é o estimador não viciado da variância. Nota-se que o erro padrão não é uma estimativa de uma quantidade pertinente a uma população, mas uma medida da incerteza da média amostral vista como uma estimativa da média populacional (Altman, 1991, p.160). A expressão (5) deixa claro que a magnitude desta incerteza diminui conforme o tamanho da amostra, n , aumenta.

Vamos agora generalizar este conceito. Seja X uma variável aleatória com distribuição F , sendo sua esperança denotada por μ_F e sua variância denotada por σ_F^2 . Usaremos a notação $X \sim (\mu_F, \sigma_F^2)$ e escreveremos $\hat{F}_n \rightarrow (x_1, x_2, \dots, x_n)$ para indicar que $\mathbf{x} = (x_1, x_2, \dots, x_n)$ é uma amostra aleatória de tamanho n obtida de uma população com função de probabilidade F . A média \bar{X} é também uma variável aleatória e tem esperança μ_F e variância σ_F^2/n , ou seja, $\bar{X} \sim (\mu_F, \sigma_F^2/n)$. Nota-se então que X e \bar{X} têm a mesma esperança, entretanto, a variância de \bar{X} é n^{-1} vezes a variância de X . Define-se então o erro padrão de \bar{x} como a raiz quadrada da variância de \bar{X} , ou seja,

$$ep_F(\bar{x}) = \sqrt{\text{var}_F(\bar{X})} = \frac{\sigma_F}{\sqrt{n}}. \quad (7)$$

A expressão (5) fornece um estimador para (7). Entretanto, nem todos os estimadores têm expressões de fácil manejo para seu respectivo erro padrão, como (5). Isto significa que o trabalho de encontrar medidas de precisão para outros estimadores $\hat{\theta} = s(X)$, que não a média, pode ser algo bastante complicado.

Sob certas condições, o teorema central do limite diz que a distribuição de \bar{X} é aproximadamente normal quando n é grande (Leite & Singer, 1990, p.77), ou seja,

$$\bar{X} \sim N(\mu_F; \sigma_F^2/n), \quad (8)$$

e, conseqüentemente,

$$\frac{\bar{X} - \mu_F}{\sigma_F/\sqrt{n}} \sim N(0; 1). \quad (9)$$

Usando uma tabela da distribuição normal padrão, de (9) podemos escrever, por exemplo, a probabilidade

$$P\left(|\bar{X} - \mu_F| < \frac{2\sigma_F}{\sqrt{n}}\right) \doteq 0,954, \quad (10)$$

onde o erro padrão mostra-se uma útil medida da precisão da estimativa da média. A probabilidade da distância entre a média amostral e a populacional ser de pelo menos dois erros padrão é aproximadamente 0,954. Verifica-se assim que, quanto menor o erro padrão da média, a expressão (10) melhor sugere uma “proximidade” entre as médias amostral e populacional.

Uma vantagem do *bootstrap* é que esta técnica não depende inteiramente do teorema central do limite, já que, em suas aplicações, medidas de precisão são obtidas diretamente dos dados (Efron & Tibshirani, 1993, p.40).

Observada uma amostra aleatória de tamanho n , oriunda de uma distribuição F , define-se uma função distribuição empírica \hat{F}_n como uma distribuição discreta, que atribui probabilidade n^{-1} a cada valor x_i , $i = 1, \dots, n$. Uma amostra *bootstrap* $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ é obtida reamostrando aleatoriamente n vezes, com reposição, as observações $\mathbf{x} = (x_1, x_2, \dots, x_n)$, onde verifica-se que $\hat{F}_n \rightarrow (x_1^*, x_2^*, \dots, x_n^*)$.

Pelo teorema de Glivenko-Cantelli (Rohatgi, 1976, p.300), \hat{F}_n converge uniformemente para F , isto é, para qualquer $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\sup_{-\infty < t < \infty} |\hat{F}_n(t) - F(t)| > \varepsilon\right) = 0, \quad (11)$$

o que justifica o uso de \hat{F}_n .

Para um t fixo, podemos escrever \hat{F}_n na forma

$$\hat{F}_n(t) = n^{-1} \sum_{j=1}^n \mathbf{I}_{\{X_j \leq t\}}, \quad (12)$$

com esperança e variância

$$\mathbf{E}(\hat{F}_n(t)) = n^{-1} \sum_{j=1}^n \mathbf{E} \mathbf{I}_{\{X_j \leq t\}} \stackrel{iid}{=} P(X_1 \leq t) = F(t) \quad (13)$$

e

$$\mathbf{V}(\hat{F}_n(t)) = \frac{F(t) - (F(t))^2}{n}, \quad (14)$$

respectivamente. Assim, mostra-se que para um t fixo, $\hat{F}_n(t)$ é um estimador não viciado de $F(t)$, e, a medida que n cresce, a variância de $\hat{F}_n(t)$ decresce.

Selecionadas B amostras *bootstrap*, $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$, de forma independente, estima-se θ em cada uma destas amostras através de $\hat{\theta}^{*b} = s(\mathbf{x}^{*b})$, $b=1, 2, \dots, B$. Uma expressão para o estimador *bootstrap* do erro padrão da estatística $\hat{\theta}$ é dada por

$$\hat{ep}_{boot} = \sqrt{\frac{\sum_{b=1}^B (s(\mathbf{x}^{*b}) - s(\cdot))^2}{B-1}}, \quad (15)$$

onde

$$s(\cdot) = \frac{\sum_{b=1}^B s(\mathbf{x}^{*b})}{B}. \quad (16)$$

Efron & Tibshirani (1993, p.47) chamam de estimador *bootstrap* ideal de $ep_F(\hat{\theta})$ o limite de \hat{ep}_{boot} quando B vai para o infinito, ou seja,

$$\lim_{B \rightarrow \infty} \hat{ep}_{boot} = ep_{\hat{F}}(\hat{\theta}^*). \quad (17)$$

O estimador *bootstrap* ideal e sua aproximação (15) são chamados estimadores *bootstrap* não paramétricos, já que se baseiam em \hat{F} , um estimador não paramétrico de F . Um estimador *bootstrap* paramétrico do erro padrão é baseado em um estimador \hat{F} de F derivado de um modelo paramétrico. Por exemplo, ao invés de estimarmos F pela função distribuição empírica \hat{F} , podemos assumir que a população tem distribuição normal.

4 *Jackknife*: um outro método para estimar o erro padrão

O *jackknife* é um método computacionalmente intensivo pioneiro para a estimação de tendenciosidades e erros padrão, introduzido por volta de 1950, por Maurice Quenouille, e mais tarde aprimorado por Tukey (Miller, 1974). Observada uma amostra de tamanho n , o método *jackknife* consiste em gerar outras n amostras de tamanho $n-1$, sendo a i -ésima nova amostra ($i = 1, \dots, n$) composta pelas observações que compõem a amostra original, com a exceção da i -ésima observação. Assim, sendo $\mathbf{x} = (x_1, x_2, \dots, x_n)$ a amostra observada, são geradas as amostras $\mathbf{x}_{(1)} = (x_2, x_3, \dots, x_n)$, $\mathbf{x}_{(2)} = (x_1, x_3, \dots, x_n)$, $\mathbf{x}_{(3)} = (x_1, x_2, x_4, \dots, x_n)$, ..., $\mathbf{x}_{(n-1)} = (x_1, x_2, \dots, x_{n-2}, x_n)$, $\mathbf{x}_{(n)} = (x_1, x_2, \dots, x_{n-1})$. Em cada uma destas amostras geradas, estima-se $\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$, $i=1,2,\dots,n$, as replicações *jackknife* da estatística $s(\cdot)$ de interesse. Da amostra original \mathbf{x} , estima-se $\hat{\theta} = s(\mathbf{x})$. Sendo

$$\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^n \hat{\theta}_{(i)}}{n}, \quad (18)$$

o estimador *jackknife* da tendenciosidade é definido por

$$\hat{tend}_{JACK} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}) \quad (19)$$

e o estimador *jackknife* do erro padrão é dado por

$$\hat{ep}_{JACK} = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2} . \quad (20)$$

Considerando, por exemplo, $\hat{\theta} = s(\mathbf{x}) = \bar{x}$, obtemos a média amostral, sem considerar a i -ésima observação, dada pela expressão

$$\bar{x}_{(i)} = \frac{n\bar{x} - x_i}{n-1} = \frac{\sum_{j \neq i} x_j}{n-1}, \quad i = 1, \dots, n . \quad (21)$$

Nota-se que a média $\bar{x}_{(\cdot)}$ das n médias dadas por (21) é igual à média amostral, ou seja, conforme (18),

$$\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^n \bar{x}_{(i)}}{n} = \frac{\sum_{i=1}^n (n\bar{x} - x_i)}{n(n-1)} = \bar{x} . \quad (22)$$

E, segundo (20), o estimador *jackknife* do erro padrão de $\hat{\theta} = \bar{x}$, é dado por

$$\hat{ep}_{JACK}(\bar{x}) = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\bar{x}_{(i)} - \bar{x})^2} . \quad (23)$$

Substituindo (21) em (23) e desenvolvendo o resultado encontrado, verifica-se que a expressão (23) é igual a

$$\hat{ep}_{JACK}(\bar{x}) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}} , \quad (24)$$

que é idêntica à expressão (5), o que ilustra uma motivação para o termo $(n-1)/n$, que aparece em (20). Nota-se que este termo pode ser interpretado como uma correção, para que a expressão (23) seja um estimador não viciado de θ .

Seja, por exemplo, uma amostra aleatória tamanho $n = 10$, sendo observados os valores expostos na Tabela 1.

Tabela 1 – Amostra aleatória, $n = 10$.

555	558	576	578	580
605	635	651	661	666

Poderão assim ser estimadas dez médias $\bar{x}_{(i)}$ conforme (21), apresentadas na Tabela 2. A média amostral é de 606,5, cujo erro padrão é estimado em 13,644. Uma estimativa *jackknife* deste erro padrão é dada por

$$\hat{ep}_{JACK} = \sqrt{\frac{9}{10} 206,846} = 13,644,$$

ou seja, igual ao erro padrão da média estimado por (5).

Tabela 2 – Construção do erro padrão *jackknife* da média.

i	$x_{(i)}$	$\bar{x}_{(i)}$	$(\bar{x}_{(i)} - \bar{x}_{(.)})^2$
1	555	612,22	32,744
2	558	611,89	29,040
3	576	609,89	11,485
4	578	609,67	10,028
5	580	609,44	8,670
6	605	606,67	0,028
7	635	603,33	10,028
8	651	601,56	24,448
9	661	600,44	36,670
10	666	599,89	43,707
$\bar{x} = 606,5$		$\hat{\theta}_{(.)} = 606,5$	soma = 206,846

5 *Jackknife* e *bootstrap*: uma comparação

O *jackknife* pode ser visto como uma aproximação do *bootstrap* conforme demonstrado por Efron & Tibshirani (1993, p.287). Seja, por exemplo, $\hat{\theta} = s(\mathbf{x}) = \bar{x}$, novamente. Quando $B \rightarrow \infty$, através da lei fraca dos grandes números mostra-se que

$$\hat{ep}_{boot}(\bar{x}) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n^2}}. \quad (25)$$

Vejam, empiricamente, no que se traduz esta afirmação. A Figura 1 mostra estimativas *bootstrap* para o erro padrão da média amostral dos dados da Tabela 1. Foram considerados valores de B entre 5 e 3.500, com incrementos de 5 unidades. Conforme B cresce, observa-se uma maior estabilização da variabilidade das estimativas do erro padrão, que se aproximam do valor estimado por (25).

Generalizando o exemplo anterior, seja uma estatística linear

$$\hat{\theta} = s(\mathbf{x}) = \mu + \frac{1}{n} \sum_{i=1}^n \alpha(x_i), \quad (26)$$

onde μ é uma constante e $\alpha(\cdot)$ é uma função.

A média amostral é um caso particular de (26), onde $\mu = 0$ e $\alpha(x_i) = x_i$. De acordo com a notação $\alpha_i = \alpha(x_i)$, quando $B \rightarrow \infty$ a estimativa *bootstrap* do erro padrão de $\hat{\theta} = s(\mathbf{x})$ definido pela expressão (26) é dada por

$$\hat{ep}_{boot}(s(\mathbf{x})) = \sqrt{\frac{\sum_{i=1}^n (\alpha_i - \bar{\alpha})^2}{n^2}} \quad (27)$$

e a estimativa *jackknife* do erro padrão deste estimador é dada por

$$\hat{ep}_{JACK}(s(\mathbf{x})) = \sqrt{\frac{\sum_{i=1}^n (\alpha_i - \bar{\alpha})^2}{n(n-1)}}. \quad (28)$$

Nota-se que as expressões (27) e (28), e, conseqüentemente, (24) e (25) são iguais, a não ser pelo fator de correção $\sqrt{(n-1)/n}$, que desaparece quando $n \rightarrow \infty$.

6 Intervalos de confiança baseados no método *bootstrap*

Baseado em seu erro padrão estimado \hat{ep} , usualmente escrevemos intervalos de confiança para um parâmetro θ de interesse, na forma

$$\hat{\theta} \pm z_{\alpha} \hat{ep}, \quad (29)$$

onde z_{α} é o 100α -ésimo percentil de uma distribuição normal padrão.

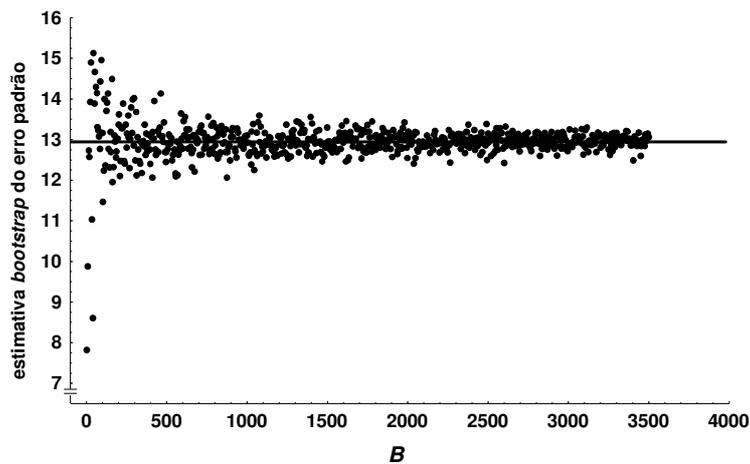


FIGURA 1 – Estimativas *bootstrap* do erro padrão da média amostral \bar{x} , onde B varia de 5 a 3.500, com incremento de 5 unidades. A linha horizontal indica a estimativa do erro padrão dada por (25), ou seja, $\hat{ep}_{boot}(\bar{x}) = 12,9439$.

Dado que, por construção, $L_I(\mathbf{x}) \leq \hat{\theta} \leq L_S(\mathbf{x})$, intervalos de confiança como (29) são ditos simétricos, pois neste caso, $\hat{\theta} - L_I(\mathbf{x})$ é igual a $L_S(\mathbf{x}) - \hat{\theta}$, ou seja, a estimativa dada por $\hat{\theta}$ é o ponto médio entre $L_I(\mathbf{x})$ e $L_S(\mathbf{x})$, segundo a notação introduzida na expressão (1). Assim, diz-se que um intervalo de confiança é simétrico quando a quantidade

$$S(\mathbf{x}) = \frac{|L_I(\mathbf{x}) - \hat{\theta}|}{|L_S(\mathbf{x}) - \hat{\theta}|} \quad (30)$$

é igual a 1. Denominaremos $S(\mathbf{x})$ de índice de simetria de um intervalo de confiança.

A motivação de (29) está na propriedade de $\hat{\theta}$ ter, assintoticamente, distribuição normal, ou seja, $\hat{\theta} \sim N(\theta; \hat{ep}^2)$. Desta propriedade, segue que

$$Z = \frac{\hat{\theta} - \theta}{\hat{ep}} \sim N(0; 1), \quad (31)$$

e daí podemos escrever a probabilidade

$$P_F[\theta \in (\hat{\theta} - z_{1-\alpha/2} \hat{ep}; \hat{\theta} + z_{1-\alpha/2} \hat{ep})] = 1 - \alpha \quad (32)$$

sendo os dados amostrados procedentes de uma distribuição F , $F \rightarrow X = (X_1, X_2, \dots, X_n)$.

De (32), definimos a expressão

$$(\hat{\theta} - z_{1-\alpha/2} \hat{ep}; \hat{\theta} + z_{1-\alpha/2} \hat{ep}) \quad (33)$$

como o intervalo de confiança padrão com probabilidade de cobertura $1 - \alpha$, ou nível de confiança $100(1 - \alpha)\%$.

Quando $\hat{\theta}$ é a média amostral, Gosset mostrou que a expressão (31) tem uma melhor aproximação em

$$Z = \frac{\hat{\theta} - \theta}{\hat{ep}} \sim t_{n-1}, \quad (34)$$

onde t_{n-1} se refere à distribuição t de Student com $n-1$ graus de liberdade. Baseado neste fato, o intervalo de confiança descrito em (33) seria melhor escrito

$$(\hat{\theta} - t_{(1-\alpha/2), n-1} \hat{ep}; \hat{\theta} + t_{(1-\alpha/2), n-1} \hat{ep}), \quad (35)$$

sendo $t_{(\alpha), n-1}$ o α -ésimo percentil da distribuição t com $n-1$ graus de liberdade.

Intervalos de confiança da forma (33) ou (35) são aproximados, dado que a probabilidade de cobertura não será exatamente igual a $1 - \alpha$, conforme seria o desejado. Segundo Efron & Tibshirani (1993, p.154), intervalos *bootstrap* também são aproximados, entretanto, oferecem melhor aproximação que os intervalos de confiança padrão.

Podemos obter intervalos de confiança *bootstrap* para θ de diversas maneiras, como através do método percentil ou do método percentil t . O método percentil tem por base unicamente os quantis e outras medidas da distribuição empírica \hat{F}_n do estimador $\hat{\theta}$ (Souza, 1998, p.452), enquanto o método percentil t , que será descrito adiante, gene-

realiza o intervalo de confiança produzido por (35), permitindo escrever intervalos da forma

$$(\hat{\theta} - t_1 \hat{e}p; \hat{\theta} - t_2 \hat{e}p), \quad (36)$$

onde as quantidades t_1 e t_2 , obtidas através da estimação de \hat{F}_n , substituem $t_{(1-\alpha/2), n-1}$ nas expressões que geram os limites inferior e superior.

6.1 Propriedades desejáveis de um intervalo de confiança

Uma propriedade desejada para intervalos de confiança é a da preservação de amplitude. Seja um parâmetro θ , que assume valores em um conjunto Θ , de forma que $\Theta = [a; b] \subset \mathfrak{R}$, $a \leq b$. Diz-se que um intervalo de confiança para θ possui a propriedade de preservação de amplitude se seus limites inferior e superior estão contidos em Θ (Silva, 1995, p.46). Por exemplo, desde que o coeficiente de correlação ρ sempre assume valores no intervalo $\Theta = [-1; 1]$, deseja-se que um intervalo de confiança para ρ construído através de uma amostra aleatória tenha limite inferior não menor que -1 e limite superior não maior que 1 .

Outra propriedade desejável é a acurácia. Seja, por simplicidade, um intervalo de confiança unilateral para um parâmetro de interesse θ , cujo único limite de confiança é dado por $\hat{\theta}_{(\alpha)}$, onde α é a probabilidade de cobertura deste intervalo. Assim,

$$P[\theta \leq \hat{\theta}_{(\alpha)}] \approx \alpha. \quad (37)$$

Um limite de confiança $\hat{\theta}_{(\alpha)}$ é dito acurado de primeira ordem se

$$P[\theta \leq \hat{\theta}_{(\alpha)}] = \alpha + O(n^{-1/2}), \quad (38)$$

e, é dito acurado de segunda ordem se

$$P[\theta \leq \hat{\theta}_{(\alpha)}] = \alpha + O(n^{-1}). \quad (39)$$

Em outras palavras, se $\hat{\theta}_{(\alpha)}$ é acurado de primeira ordem, a probabilidade de cobertura $P[\theta \leq \hat{\theta}_{(\alpha)}]$ será igual a α a não ser por

um termo cuja ordem de magnitude é, no máximo, igual à de $n^{-1/2}$, para todo n suficientemente grande.

Por outro lado, quando $\hat{\theta}_{EXATO(\alpha)}$ é um limite de confiança exato, satisfazendo a

$$P[\theta \leq \hat{\theta}_{EXATO(\alpha)}] = \alpha, \quad (40)$$

diz-se que $\hat{\theta}_{(\alpha)}$ é correto de primeira ordem se

$$\hat{\theta}_{(\alpha)} = \hat{\theta}_{EXATO(\alpha)} + O_p(n^{-1}), \quad (41)$$

e correto de segunda ordem se

$$\hat{\theta}_{(\alpha)} = \hat{\theta}_{EXATO(\alpha)} + O_p(n^{-3/2}). \quad (42)$$

Se $X_n = O_p(a_n)$ e $Y_n = O_p(b_n)$, onde $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são duas seqüências de números reais (ou variáveis aleatórias), então $X_n Y_n = O_p(a_n b_n)$, conforme demonstrado em Leite & Singer (1990, p.41). Assim, as expressões (41) e (42) equivalem a dizer que $\hat{\theta}_{(\alpha)}$ é correto de primeira ordem se

$$\hat{\theta}_{(\alpha)} = \hat{\theta}_{EXATO(\alpha)} + O_p(n^{-1/2}) \hat{\sigma}, \quad (43)$$

e correto de segunda ordem se

$$\hat{\theta}_{(\alpha)} = \hat{\theta}_{EXATO(\alpha)} + O_p(n^{-1}) \hat{\sigma}, \quad (44)$$

dado que $\hat{\sigma}$, um plausível estimador do erro padrão de $\hat{\theta}$, é usualmente de ordem $n^{-1/2}$.

Estes conceitos, acurácia e corretividade, são descritos por Efron & Tibshirani (1993, p.321). Segundo estes autores, corretividade em dada ordem implica acurácia naquela mesma ordem.

Maiores detalhes sobre as definições das ordens de magnitude $O(\cdot)$ e $O_p(\cdot)$ encontram-se no Apêndice 1.

6.2 Intervalo *bootstrap* padrão

Seja $\hat{ep}_{boot}(\hat{\theta})$, $\hat{\theta} = s(x)$, a estimativa *bootstrap* do erro padrão de $\hat{\theta}$, definida em (15), o intervalo de confiança *bootstrap* padrão para θ , com probabilidade de cobertura de aproximadamente $1-\alpha$, é dado por

$$\hat{\theta} \pm z_{1-\alpha/2} \hat{ep}_{boot}(\hat{\theta}). \quad (45)$$

A maior vantagem deste método reside na sua simplicidade algébrica de encontrar um intervalo de confiança para θ . Entretanto, sendo a equação (45) uma consequência de

$$Z = \frac{\hat{\theta} - \theta}{\hat{ep}_{boot}(\hat{\theta})} \sim N(0; 1), \quad (46)$$

o fato desta aproximação assintótica estar condicionada ao tamanho da amostra n implica que o intervalo assim encontrado pode ser impreciso. Além disto, características como assimetria e tendenciosidades podem estar presentes na distribuição de Z , podendo prejudicar o desempenho do intervalo de confiança. Uma solução para contornar este problema é buscar uma transformação monótona $g(\cdot)$, tal que, segundo Silva(1995, p.54), a distribuição de $\hat{\phi} = g(\hat{\theta})$ esteja mais próxima da distribuição normal do que a distribuição de $\hat{\theta}$. De qualquer forma, o intervalo de confiança *bootstrap* padrão nem sempre satisfaz à propriedade de preservação de amplitude.

Tratando-se de amostras pequenas, pelo menos no caso em que $\hat{\theta} = s(\mathbf{x}) = \bar{x}$ (Silva, 1995, p.57), uma melhor aproximação para (46) encontra-se na distribuição t de Student, ou seja,

$$T = \frac{\hat{\theta} - \theta}{\hat{ep}_{boot}(\hat{\theta})} \sim t_{n-1}. \quad (47)$$

6.3 O método *t-bootstrap*

O procedimento *t-bootstrap*, também conhecido como método pivotal, segundo Efron & Tibshirani (1993, p.161) é uma generalização do usual método t de Student, sendo particularmente aplicável a estatísticas de locação, como a média amostral, a mediana, ou percentis amostrais. Estes autores citam que, pelo menos em sua forma tradicional, o método *t-bootstrap* não é bom para a construção de intervalos para outras estatísticas, como por exemplo, o coeficiente de correlação. Nestes casos, é necessário o uso de transformações.

Enquanto em (34) Z tem distribuição t de Student, a essência do método está na estimação da distribuição de Z diretamente dos dados amostrados. Geradas B amostras *bootstrap* $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$, independentes, estima-se em cada uma destas amostras $\hat{\theta}^{*b} = s(\mathbf{x}^{*b})$, $b=1,2,\dots,B$, e daí encontra-se

$$Z^{*b} = \frac{\hat{\theta}^{*b} - \hat{\theta}}{\hat{ep}(\hat{\theta}^{*b})}, \quad (48)$$

onde $\hat{ep}(\hat{\theta}^{*b})$ é o erro padrão estimado de $\hat{\theta}^{*b}$, $b=1,2,\dots,B$ (Efron & Tibshirani, 1993, p.160). O α -ésimo percentil de Z^{*b} é estimado por $\hat{t}_{(\alpha)}$ tal que

$$\frac{\sum_{b=1}^B \mathbf{I}\{Z^{*b} \leq \hat{t}_{(\alpha)}\}}{B} = \alpha. \quad (49)$$

O intervalo de confiança construído por este método é então

$$(\hat{\theta} - \hat{t}_{(1-\alpha)} \hat{ep}; \hat{\theta} - \hat{t}_{(\alpha)} \hat{ep}). \quad (50)$$

Nota-se que o lado direito da expressão (48), interpretada como uma quantidade pivotal aproximada, contém em seu denominador um estimador do desvio padrão de $\hat{\theta}^*$ para a amostra \mathbf{x}^{*b} . Isto significa que, quando não estiver disponível um plausível estimador $\hat{ep}(\hat{\theta}^{*b})$, um novo esquema de reamostragem é necessário para estimar-se este erro padrão, para cada uma das B amostras *bootstrap*. Desta forma, o número de amostras *bootstrap* necessárias para a construção do intervalo de confiança *t-bootstrap* é dado pelo produto de B pelo número de novas amostras *bootstrap* requerida para cada estimação de $\hat{ep}(\hat{\theta}^{*b})$, o que, em muitos casos, pode representar um custo computacional muito grande.

Karlsson & Löthgren (2000) chamam este processo de *double bootstrap*, e propõem um novo estimador de $\hat{ep}(\hat{\theta}^{*b})$ que elimina o trabalho de extrair-se novas amostras *bootstrap* além das B amostras iniciais. Através de simulações, os autores mostram que o intervalo de confiança proposto por eles é similar ao intervalo *t-bootstrap*, em sua

cobertura. Entretanto, o trabalho destes autores não será discutido aqui com detalhes.

6.4 O método percentil

Encontradas replicações \mathbf{x}^* de $\mathbf{x} = (x_1, x_2, \dots, x_n)$, e, conseqüentemente estimadas replicações *bootstrap* da estatística de interesse $\hat{\theta}^* = s(\mathbf{x}^*)$, o intervalo de confiança de probabilidade de cobertura $1-\alpha$ construído pelo método percentil é obtido pelos $(\alpha/2)$ -ésimo e $(1-\alpha/2)$ -ésimo percentis de \hat{G} , a função distribuição acumulada de $\hat{\theta}^*$ (Efron, 1982, p.78). Uma expressão para tal intervalo é dada por

$$(L_I(\mathbf{x}^*); L_S(\mathbf{x}^*)) = (\hat{G}^{-1}_{(\alpha/2)}; \hat{G}^{-1}_{(1-\alpha/2)}) . \quad (51)$$

Como $\hat{G}^{-1}(\alpha) = \hat{\theta}^*_{(\alpha)}$, o 100α -ésimo percentil de $\hat{\theta}^*$, esta expressão pode ser reescrita da forma (Efron & Tibshirani, 1993, p.170)

$$(L_I(\mathbf{x}^*); L_S(\mathbf{x}^*)) = (\hat{\theta}^*_{(\alpha/2)}; \hat{\theta}^*_{(1-\alpha/2)}) . \quad (52)$$

Desta forma, este intervalo consiste na porção central de tamanho $1-\alpha$ da distribuição de $\hat{\theta}^*$.

As expressões (51) e (52), segundo Efron & Tibshirani (1993, p.170), referem-se à situação ideal, onde o número de replicações *bootstrap* é infinita. Na prática, utiliza-se um número finito B de replicações, onde são obtidas as amostras $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$, e, a partir destas, $\hat{\theta}^{*b} = s(\mathbf{x}^{*b})$, $b=1,2,\dots,B$. O intervalo de confiança será então obtido por

$$(L_I(\mathbf{x}^*); L_S(\mathbf{x}^*)) \approx (\hat{\theta}^{*B}_{(\alpha/2)}; \hat{\theta}^{*B}_{(1-\alpha/2)}) , \quad (53)$$

sendo $\hat{\theta}^{*B}_{(\alpha/2)}$ o $100\alpha/2$ -ésimo percentil dos valores empíricos $\hat{\theta}^{*b}$, ou seja, o $B\alpha/2$ -ésimo valor de uma ordenação das B replicações de $\hat{\theta}^{*b}$. Se $B\alpha/2$ não for um número inteiro, Efron & Tibshirani (1993, p.170) recomendam tomar $k = \text{int}((B+1)\alpha/2)$, o maior inteiro menor ou igual a $(B+1)\alpha/2$, e daí os quantis empíricos $\alpha/2$ e $1-\alpha/2$ são obtidos pelo k -ésimo e $(B+1-k)$ -ésimo maiores valores de $\hat{\theta}^{*b}$.

Uma boa propriedade do intervalo de confiança construído pelo método percentil é a invariância a transformações monótonas. Seja o

parâmetro $\phi = g(\theta)$, estimado pela estatística $\hat{\phi} = g(\hat{\theta})$, onde $g(\cdot)$ é uma função monótona. O intervalo de confiança para ϕ será dado por

$$(L_I(\mathbf{x}^*); L_S(\mathbf{x}^*)) = (g(\hat{\theta}_{(a/2)}^*); g(\hat{\theta}_{(1-a/2)}^*)) \quad (54)$$

se $g(\cdot)$ é uma função monótona crescente, e

$$(L_I(\mathbf{x}^*); L_S(\mathbf{x}^*)) = (g(\hat{\theta}_{(1-a/2)}^*); g(\hat{\theta}_{(a/2)}^*)) \quad (55)$$

se $g(\cdot)$ é uma função monótona decrescente.

As figuras 2 e 3 mostram, respectivamente, os limites inferior e superior para intervalos de confiança 90% para a média dos dados da Tabela 1, estimados segundo o método percentil, para um número B de amostras *bootstrap*, onde B varia de 100 a 7.000. De 100 a 2.500, o incremento é de 5 unidades, e de 2.500 a 7.000, usa-se incremento de 50 unidades.

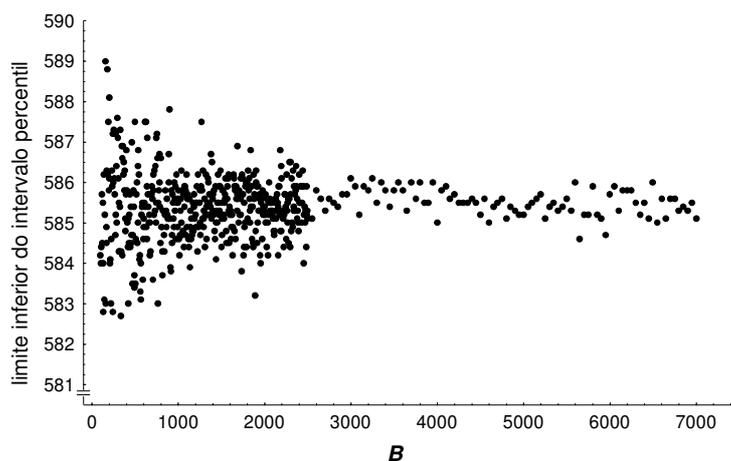


FIGURA 2 – Limite inferior do intervalo de confiança 90% percentil para a média μ dos dados da Tabela 1, em função do número de amostras *bootstrap* B , onde B varia de 100 a 7.000.

6.5 O método BCa

O método BCa também utiliza os percentis da distribuição *bootstrap* para a construção dos intervalos de confiança para parâmetros de

interesse. Descrito por Efron (1987), DiCiccio & Tibshirani (1987) e Efron & Tibshirani (1986 e 1993, p.184-8), este método utiliza percentis que dependem de duas constantes, \hat{z}_0 e \hat{a} , denominadas correção para tendência e aceleração, respectivamente. Do nome original destas quantidades, *bias-corrected* e *acceleration*, vem a abreviatura BC_a .

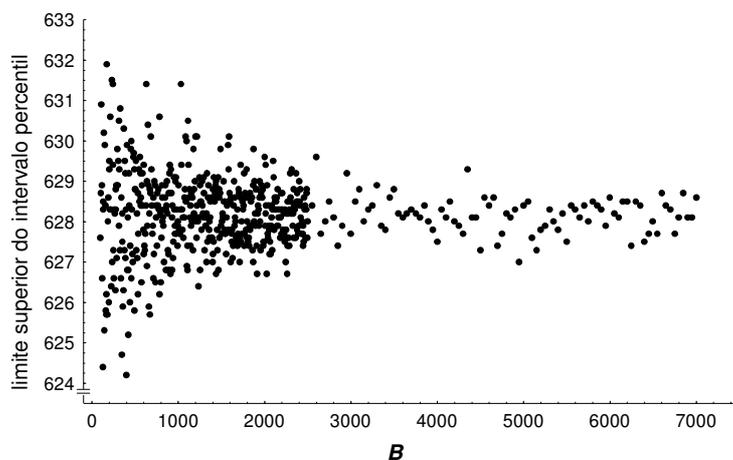


FIGURA 3 – Limite superior do intervalo de confiança 90% percentil para a média μ dos dados da Tabela 1, em função do número B de amostras *bootstrap*, onde B varia de 100 a 7.000.

Sendo $\hat{\theta}_{(\alpha_i)}^*$, $i = 1, 2$, o $100\alpha_i$ -ésimo percentil das B estimativas da estatística de interesse, geradas por B amostras *bootstrap*, o intervalo de confiança BC_a , de desejada probabilidade de cobertura $1-\alpha$, é dado por

$$BC_a : \left(\hat{\theta}_{L_1} ; \hat{\theta}_{L_5} \right) = \left(\hat{\theta}_{(\alpha_1)}^* ; \hat{\theta}_{(\alpha_2)}^* \right), \quad (56)$$

onde

$$\alpha_1 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})} \right) \quad \alpha_2 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})} \right). \quad (57)$$

Em (57), $\Phi(\cdot)$ é a função distribuição acumulada de uma normal padrão e $z_{(\alpha)}$ o 100α -ésimo percentil de uma distribuição normal padrão. Se \hat{a} e \hat{z}_0 são iguais a zero, (56) é similar a (52).

A constante de correção para tendência, \hat{z}_0 , é obtida da proporção de replicações *bootstrap* cujas estimativas $\hat{\theta}^{*b}$ são menores que a estimativa original $\hat{\theta}$, da seguinte maneira:

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\sum_{b=1}^B \mathbf{I}_{\{\hat{\theta}^{*b} < \hat{\theta}\}}}{B} \right). \quad (58)$$

Assim, segundo Efron & Tibshirani (1993, p.186), \hat{z}_0 representa a magnitude da tendenciosidade mediana de $\hat{\theta}^*$, ou seja, a discrepância entre a mediana de $\hat{\theta}^*$ e $\hat{\theta}$, em uma escala normalizada. Obteremos \hat{z}_0 igual a zero se exatamente a metade das medidas de $\hat{\theta}^{*b}$ forem menores ou iguais a $\hat{\theta}$.

Dentre as várias possibilidades de se obter \hat{a} , uma forma é dada em termos dos valores *jackknife* de $\hat{\theta} = s(\mathbf{x})$,

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6 \left[\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2 \right]^{3/2}}, \quad (59)$$

onde $\hat{\theta}_{(\cdot)}$ é dado por (18) e $\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$, $i=1,2,\dots,n$.

Intervalos de confiança usuais, baseados na teoria da máxima verossimilhança (*MV*), possuem a forma $\hat{\theta} \pm \hat{\sigma} z_{(\alpha)}$, sendo $\hat{\theta}$ o estimador de *MV* de θ , $\hat{\sigma}$ uma estimativa de seu desvio padrão e $z_{(\alpha)}$ o α -ésimo percentil de uma distribuição normal padrão. Tais intervalos se baseiam na aproximação assintótica

$$\frac{\hat{\theta} - \theta}{\hat{\sigma}} \sim N(0; 1), \quad (60)$$

onde $\hat{\sigma}$ é fixo. A motivação de construir intervalos de confiança pelo método BC_a está na busca de um melhor desempenho nesta convergência à normalidade e na procura de uma constância para σ , onde, através de alguma transformação monótona g , da forma $\phi = g(\theta)$ e $\hat{\phi} = g(\hat{\theta})$, produz-se

$$\frac{\hat{\phi} - \phi}{\tau} \sim N(-z_0 \sigma_\phi; \sigma_\phi^2), \quad (61)$$

onde τ é uma constante que expressa o erro padrão de $\hat{\phi}$ (Efron, 1987). Ainda em (61), temos que $\sigma_\phi = 1 + a\phi$, ou seja, o método BC_a admite que o desvio padrão de $\hat{\phi}$ tenha uma variação linear conforme ϕ , a uma taxa expressa por a . Em outras palavras, Silva (1995, p.91) explica que a constante a mede a taxa de mudança do desvio padrão de $\hat{\phi}$ sobre a escala da distribuição normal padrão, o que motiva a denominação “constante de aceleração”. A construção de um intervalo de confiança BC_a não exige o conhecimento de $g(\theta)$, o que torna este método atraente (DiCiccio & Tibshirani, 1987).

Uma vantagem do método BC_a , apontada por Efron & Tibshirani (1993, p.187), está no fato dele gerar limites de confiança acurados de segunda ordem. Isto significa que, para um intervalo de confiança BC_a , dado por $(\hat{\theta}_{L_I}; \hat{\theta}_{L_S})$, com probabilidade de cobertura $1 - \alpha$, observa-se

$$P(\theta < \hat{\theta}_{L_I}) \approx \frac{\alpha}{2} + \frac{c_I}{n} \quad \text{e} \quad P(\theta > \hat{\theta}_{L_S}) \approx \frac{\alpha}{2} + \frac{c_S}{n}, \quad (62)$$

para c_I e c_S constantes. Estas probabilidades são, portanto, iguais a α a não ser por um termo que se aproxima de zero à medida que n cresce, a uma razão n^{-1} .

Intervalos gerados pelo método *t-bootstrap* também são acurados de segunda ordem, entretanto, não possuem a propriedade de invariância a transformações monótonas, que os intervalos gerados pelo método BC_a possuem.

O método percentil, além de possuir a propriedade de invariância a transformações monótonas, gera limites de confiança acurados de primeira ordem, ou seja, com uma maior ordem de magnitude.

O coeficiente de aceleração \hat{a} , para os dados da Tabela 1, é estimado em 0,0125. Para a média destes dados, intervalos de confiança de 90% foram gerados através do método BC_a , sendo os limites inferior e superior mostrados nas Figuras 4 e 5, respectivamente, segundo valores de B . Nota-se nestas Figuras que as estimativas dos limites inferior e superior do intervalo convergem a um determinado ponto, conforme B cresce. Neste estudo, B variou de 100 a 6.800, com incremento de 5 unidades.

A Figura 6 mostra as estimativas de \hat{z}_0 utilizadas na construção destes intervalos, segundo B . Nota-se que, à medida que B cresce, \hat{z}_0

tende a convergir para um valor maior que zero, o que, associado ao fato de $\hat{\alpha}$ ser diferente de zero, sugere que as correções para tendência e aceleração proporcionada pelo método BC_a sejam importantes.

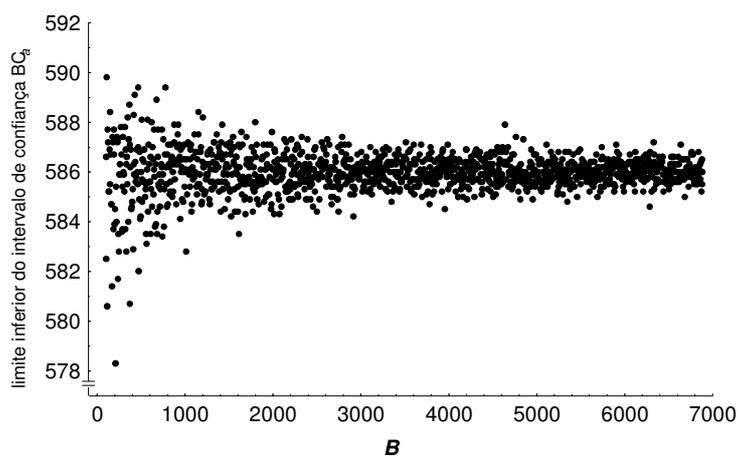


FIGURA 4 – Limite inferior do intervalo de confiança 90% BC_a para a média μ dos dados da Tabela 1, em função do número de amostras *bootstrap* B , onde B varia de 100 a 6.800, com incremento de 5 unidades.

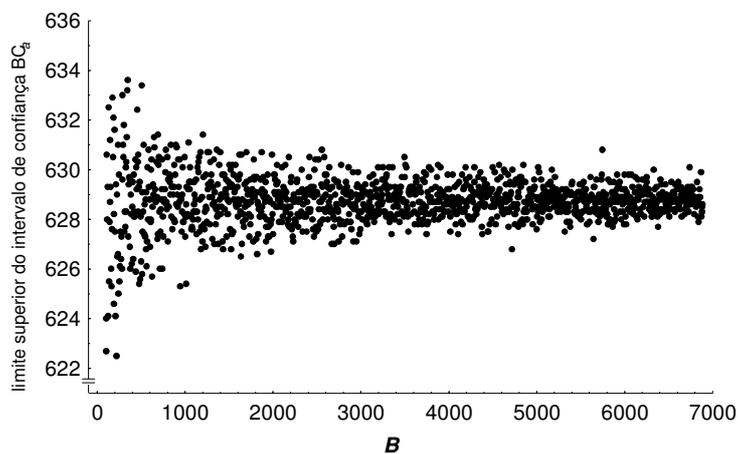


FIGURA 5 – Limite superior do intervalo de confiança 90% BC_a para a média μ dos dados da Tabela 1, em função do número de amostras *bootstrap* B , onde B varia de 100 a 6.800, com incremento de 5 unidades.

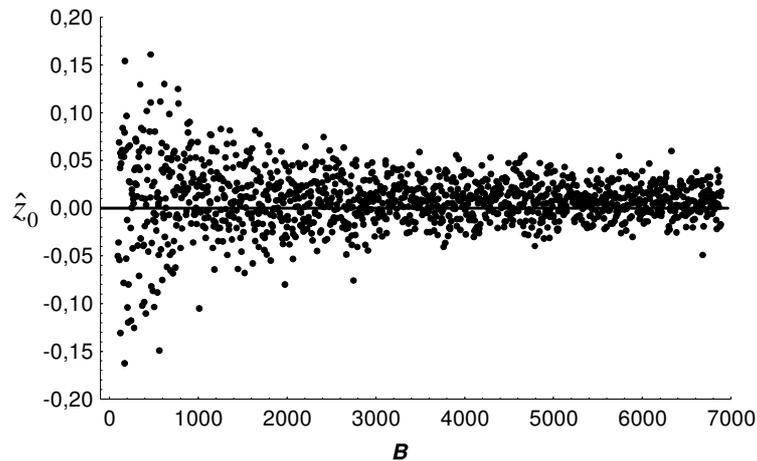


FIGURA 6 – Estimativas de \hat{z}_0 utilizadas na construção de intervalos de confiança 90% para a média dos dados da Tabela 1, segundo o número de amostras *bootstrap* B , onde B varia de 100 a 6.800, com incremento de 5 unidades. \hat{z}_0 é dado por (58).

7 Um exemplo numérico

Em oncologia, define-se um marcador como qualquer substância que pode constituir um sinal da presença e do desenvolvimento de um tumor (Farante & Chacón, 1990). O CA125 (*Cancer Antigen 125*) é um marcador tumoral usado em larga escala no diagnóstico e seguimento do câncer de ovário. Pacientes que apresentam concentrações séricas elevadas de CA125 (mensurado em unidades/ml) no sangue periférico teriam uma chance maior de serem portadoras de um tumor maligno (Ward & Shepherd, 1986).

Em um estudo, dentre 158 pacientes portadoras de tumores pélvicos, um exame histopatológico indicou que 67 destas pacientes eram portadoras de tumores malignos, enquanto outras 91 pacientes eram portadoras de processos benignos (Torres, 1998). A Tabela 3 mostra a média amostral para as concentrações séricas de CA125 (em ln unidades/ml), para as amostras de pacientes portadoras de processos benignos e portadores de tumores malignos, o erro padrão destas médias medido pela equação (7) e segundo o método *bootstrap*, e

intervalos de confiança 90% para as médias, estimados segundo os diversos métodos apresentados. Nas estimativas dos intervalos de confiança *bootstrap*, foram utilizadas $B = 10.000$ amostras. O uso do logaritmo da concentração sérica do CA125, e não sua mensuração original em unidades/ml, é justificado pelo fato da distribuição das observações possuir uma grande assimetria.

Como esperado, a média amostral \bar{x}_2 para a concentração sérica de CA125 mensurada na amostra composta por pacientes com tumores malignos é maior que aquela mensurada na amostra composta por pacientes portadores de processos benignos, \bar{x}_1 .

A Figura 7 mostra *box plots* comparativos para as concentrações séricas de CA125, evidenciando que nas amostra das mulheres com tumores malignos estão presentes algumas observações atípicas (“*outliers*”).

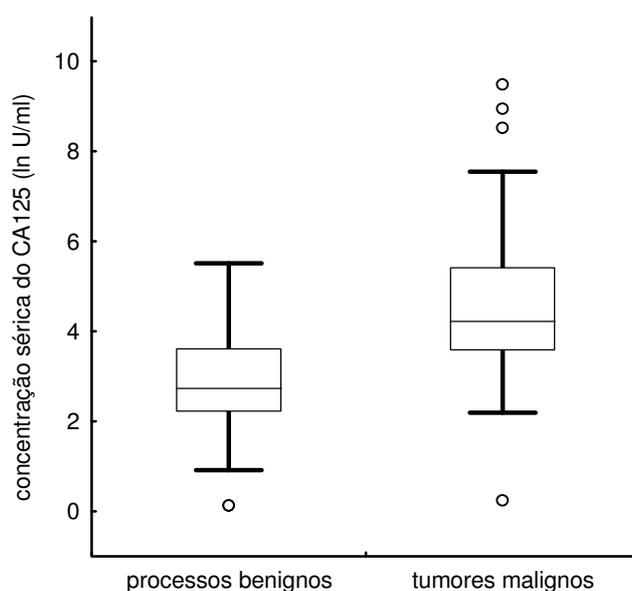


FIGURA 7 – *Box plots* para o logaritmo das concentrações séricas do CA125 (ln U/ml).

Sejam $\hat{\sigma}_1^2$ e $\hat{\sigma}_2^2$ estimadores não viciados para as variâncias das concentrações séricas de CA125, medidas em ambas amostras. O inter-

valo de confiança assintótico para a diferença entre as médias, com probabilidade de cobertura $1-\alpha$, é dado por

$$(\bar{x}_2 - \bar{x}_1 - t_{(1-\alpha/2)} \hat{ep}(\bar{x}_2 - \bar{x}_1); \bar{x}_2 - \bar{x}_1 + t_{(1-\alpha/2)} \hat{ep}(\bar{x}_2 - \bar{x}_1)), \quad (63)$$

onde $t_{(1-\alpha/2)}$ é o $(1-\alpha/2)$ -ésimo percentil de uma distribuição t de Student com n_1+n_2-2 graus de liberdade e $\hat{ep}(\bar{x}_2 - \bar{x}_1)$ é o erro padrão da diferença entre as médias, estimado por

$$\hat{ep}(\bar{x}_2 - \bar{x}_1) = \hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad (64)$$

(Altman & Gardner, 1989, p.21) sendo $\hat{\sigma}$ a raiz quadrada da média ponderada

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}}. \quad (65)$$

Tabela 3 – Médias amostrais da concentração sérica de CA125 (em ln U/ml), com seus respectivos erros padrão e intervalos de confiança 90%, para as amostras de 91 pacientes portadoras de processos benignos e 67 pacientes portadoras de tumores malignos.

	processos benignos	tumores malignos	diferença entre as médias
média amostral (ln U/ml)	2,8259	4,5477	1,7218
erro padrão	0,1028	0,2096	0,2160 (a) 0,2334 (b)
erro padrão <i>bootstrap</i>	0,1022	0,2094	0,2357
Intervalo assintótico (t de Student)	(2,655 a 2,997)	(4,198 a 4,897)	(1,364 a 2,079) (a) (1,334 a 2,109) (b)
Intervalo <i>bootstrap</i> padrão	(2,656 a 2,996)	(4,198 a 4,897)	(1,331 a 2,112)
Intervalo percentil	(2,661 a 2,996)	(4,212 a 4,890)	(1,343 a 2,104)
Intervalo BC _a	(2,662 a 2,997)	(4,219 a 4,897)	-

(a) considerando igualdade de variâncias

(b) considerando variâncias diferentes

O intervalo de confiança apresentado em (63) é similar àquele descrito em (35), onde $\hat{\theta}$ é o estimador da diferença entre as médias. Algumas propriedades deste intervalo, quando usadas pequenas amostras, são discutidas por Berengut (1981).

Estimados $\hat{\sigma}_1^2=0,9613$ e $\hat{\sigma}_2^2=2,9430$, encontrou-se o intervalo de confiança de 90% assintótico para a diferença entre as médias, conforme (63), mostrado na Tabela 3. Entretanto, este método tem como limitação o fato de considerar variâncias iguais, o que nem sempre é um pressuposto verdadeiro. Uma alternativa, assumindo variâncias diferentes, considera em (63) a quantidade $t_{1-\alpha/2}$ como o $(1-\alpha/2)$ -ésimo percentil de uma distribuição t de Student com ν graus de liberdade, onde ν é dado por

$$\nu = \frac{\left(\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\hat{\sigma}_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{\hat{\sigma}_2^2}{n_2}\right)^2}{n_2-1}}, \quad (66)$$

e

$$\hat{ep}(\bar{x}_2 - \bar{x}_1) = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}. \quad (67)$$

Verificou-se que os erros padrão das médias, estimados segundo a equação (7), são bastante próximos daqueles estimados pelo método *bootstrap*, utilizadas $B=10.000$ amostras. Conseqüentemente, os intervalos de confiança 90% estimados pelo método assintótico (utilizando percentis de uma distribuição t de Student) são similares àqueles estimados pelo método *bootstrap* padrão. Por construção, estes dois intervalos de confiança são simétricos.

O erro padrão *bootstrap* para a diferença entre as médias foi estimado segundo a equação (15), onde

$$s(\mathbf{x}^{*b}) = \bar{x}_2^{*b} - \bar{x}_1^{*b} \quad (68)$$

e

$$s(\cdot) = \frac{\sum_{b=1}^B (\bar{x}_2^{*b} - \bar{x}_1^{*b})}{B}, \quad (69)$$

sendo \bar{x}_2^{*b} e \bar{x}_1^{*b} , $b=1,2,\dots,B$, as médias das mensurações do CA125 (em ln U/ml) baseadas em $B=10.000$ amostras *bootstrap* independentes, estimadas respectivamente sobre as amostras de mulheres portadoras de tumores malignos e processos benignos.

As Figuras 8 e 9 mostram gráficos de probabilidade normal e histogramas para as médias geradas em cada uma das $B=10.000$ amostras *bootstrap* extraídas, respectivamente, das amostras de pacientes com processos benignos e tumores malignos. As curvas que se sobrepõem aos histogramas referem-se a curvas gaussianas com médias e variâncias iguais às calculadas sobre as distribuições amostrais das médias \bar{x}_1^{*b} e \bar{x}_2^{*b} , $b = 1,\dots,10.000$. Observa-se que ambas figuras sugerem distribuições satisfatoriamente simétricas para as médias, o que tem por consequência intervalos *bootstrap* estimados pelo método percentil, para as médias, relativamente simétricos (Tabela 4).

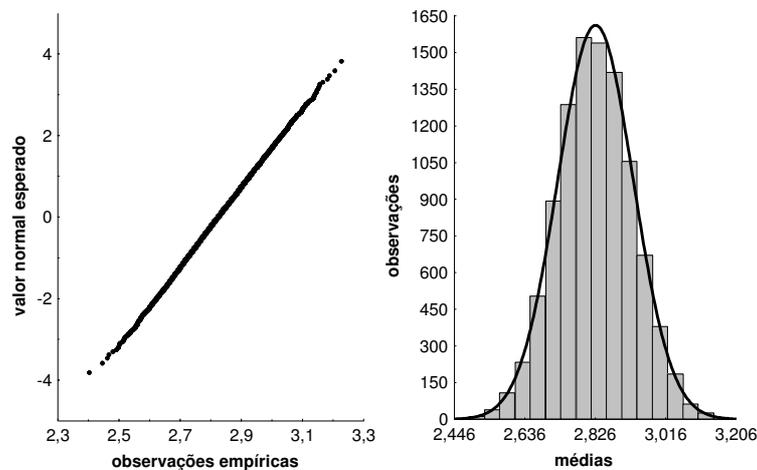


FIGURA 8 – Gráfico de probabilidade normal e histograma para as médias \bar{x}_1^{*b} das concentrações séricas do CA125 (em ln U/ml), para as pacientes portadoras de processos benignos, estimadas em $B=10.000$ amostras *bootstrap*

Na construção dos intervalos BC_a , estimou-se \hat{a} em 0,0015 e 0,0150 para, respectivamente, as amostras de pacientes portadoras de

processos benignos e tumores malignos. As constantes de correção para tendência, \hat{z}_0 , foram estimadas em, respectivamente, 0,0055 e -0,0023, para estas amostras, baseado nas $B=10.000$ amostras *bootstrap*. As quantidades α_1 e α_2 , de acordo com as equações apresentadas em (57), são dadas respectivamente por 0,0516 e 0,9515, para a amostra de pacientes portadoras de processos benignos e por 0,0537 e 0,9537, para a amostra de pacientes portadoras de tumores malignos. Dado que α_1 , em ambos casos, é maior que $\alpha/2 = 0,05$, os limites inferiores dos intervalos de confiança 90% BC_a para as respectivas médias refletem valores maiores que aqueles produzidos pelo método percentil. Analogamente, sendo α_2 maior que $1-\alpha/2 = 0,95$ também em ambos casos, os limites superiores produzidos pelo método BC_a também são maiores. Entretanto, as amplitudes $(L_S(x) - L_I(x))$ dos intervalos produzidos pelo método percentil são ligeiramente menores que aquelas produzidas pelos intervalos BC_a .

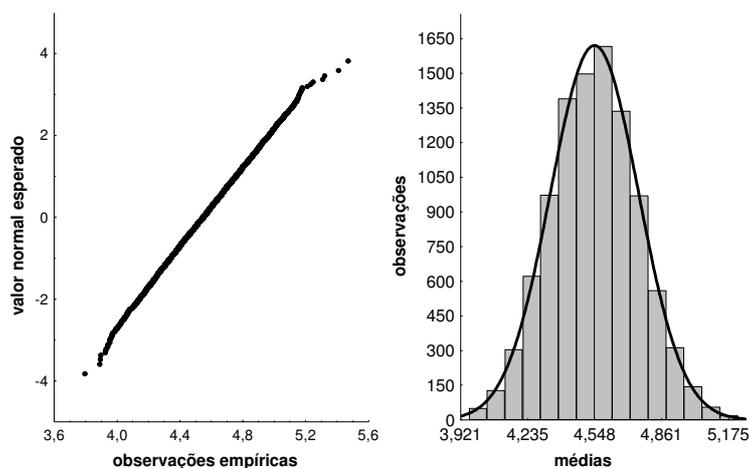


FIGURA 9 – Gráfico de probabilidade normal e histograma para as médias \bar{x}_2^{*b} das concentrações séricas do CA125 (em ln U/ml), para as pacientes portadoras de tumores malignos, estimadas em $B=10.000$ amostras *bootstrap*.

A construção de um intervalo de confiança para a diferença entre as médias, através do método BC_a , foi desconsiderada, devido à grande dificuldade de se obter uma estimativa de \hat{a} .

Na Tabela 4, observa-se que os intervalos BC_a não são tão simétricos quanto os intervalos estimados pelo método percentil, o que é consequência das correções para tendência e aceleração. Os intervalos de confiança assintóticos e estimados pelo método *bootstrap* padrão, por construção, são simétricos.

Tabela 4 – Índices de simetria, conforme (30), para os intervalos de confiança apresentados na Tabela 3.

	processos benignos	tumores malignos	diferença entre as médias
intervalos assintóticos	1,000	1,000	1,000
intervalo <i>bootstrap</i> padrão	1,000	1,000	1,000
intervalo percentil	0,969	0,981	0,991
intervalo BC_a	0,958	0,941	-

O gráfico de probabilidade normal apresentado na Figura 9 ilustra o efeito que os *outliers* verificados na Figura 7, para as concentrações séricas do CA125 (em ln U/ml) medidas em pacientes portadoras de tumores malignos, exercem sobre a distribuição das médias \bar{x}_2^{*b} , $b=1,2,\dots,B$. O mesmo fenômeno não ocorre para a distribuição de \bar{x}_1^{*b} , $b=1,2,\dots,B$ (Figura 8).

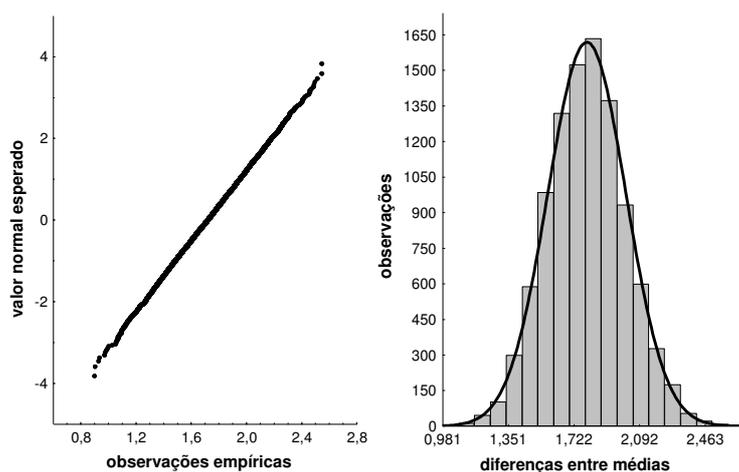


FIGURA 10 – Gráfico de probabilidade normal e histograma para a diferença entre as médias $\bar{x}_2^{*b} - \bar{x}_1^{*b}$, das concentrações séricas do CA125 (em ln U/ml), estimadas em $B=10.000$ amostras *bootstrap*.

8 Comentários finais

O método *bootstrap* constitui uma eficiente alternativa para a teoria usual de obtenção de estimativas do erro padrão de $\hat{\theta}$. Livre de complexidades algébricas, possibilita a obtenção de intervalos de confiança sem a necessidade de pressupostos sobre a distribuição do estimador. Vários tipos de intervalos de confiança *bootstrap* podem ser considerados, com variado custo computacional.

Os métodos aqui descritos para estimar erros padrão e gerar intervalos de confiança consideram parâmetros vindos de populações de tamanho infinito ou grandes o suficiente para que possam ser tratadas como de tamanho infinito. Métodos *bootstrap* adequados à populações finitas são apresentados por Booth *et al.* (1994).

O problema de determinar quantas replicações B são necessárias para a obtenção de boas estimativas dos limites inferior e superior de intervalos de confiança construídos pelo método *bootstrap* é discutido por Efron & Tibshirani (1986). Efron (1987) aborda este problema em relação à construção de intervalos BC_a . O trabalho de Hall (1986) explora a questão da determinação ótima de B com maior profundidade. Uma forma simples de determinar o número necessário de replicações B , consiste em conduzir um estudo empírico, replicando a aplicação do método *bootstrap* para diversos valores de B . A construção de gráficos como os apresentados nas Figuras 2 a 5 ilustra este procedimento. Observando o comportamento das estimativas *bootstrap* representadas nestes gráficos, regiões onde estas estimativas sugerem visualmente uma boa convergência fornecem uma boa aproximação do número ideal de replicações B . Entretanto, este método tem por limitação o fato de poder exigir um intenso esforço computacional.

Dentre os métodos Bayesianos, o *bootstrap* tem sua forma análoga discutida por Rubin (1981) e Davison & Hinkley (1997, p.512). Ao invés de simular a distribuição amostral da estatística utilizada para estimar θ , o *bootstrap* Bayesiano simula a distribuição *a posteriori* deste parâmetro.

Agradecimentos

Os autores agradecem a José Carlos Campos Torres, médico oncologista da Universidade Estadual de Campinas – UNICAMP, pela concessão dos dados utilizados no exemplo numérico, e a dois referêes pelos comentários e sugestões apresentadas. A pesquisa de Francisco Louzada-Neto é financiada parcialmente pelo CNPq.

MARTINEZ, E. Z., LOUZADA-NETO, F. Bootstrap confidence interval estimation. *Rev. Mat. Estat.* (São Paulo), v.19, p.217-251, 2001.

- **SUMMARY:** *Quantifying the error on parameter estimation is one of the major worries on statistical inference. Usually we focus on the estimation of the standard error of $\hat{\theta}$, the estimate of a parameter of interest θ . The problem is that analytical methods are not always feasible. In this context, bootstrap methods are an efficient alternative, presenting estimates of the standard error free of algebraic expressions. Also, the confidence intervals can be obtained without presuppositions about the distribution of the estimator. In this article we present the bootstrap method with its several different types of confidence intervals in a concise way. The methodology is illustrated by a real numerical example on a cancer study.*
- **KEYWORDS:** *Bootstrap, Confidence Intervals, Standard Deviation, Simulation, Biostatistics.*

Referências bibliográficas

ALTMAN, D. G. *Practical statistics for medical research*. London: Chapman & Hall, 1991. 611p.

ALTMAN, D. G., GARDNER, M. J. Calculating confidence intervals for means and their differences. In: GARDNER, M. J., ALTMAN, D. G. *Statistics with confidence: confidence intervals and statistical guidelines*. London: British Medical Journal, 1989. p.20-7.

BOOTH, J. G., BUTLER, R. W., HALL, P. Bootstrap methods for finite populations. *J. Am. Statist. Assoc.*, v.89, n.428, p.1282-9, 1994.

BICKEL, P. J., DOKSUM, K. A. *Mathematical statistics: basics ideas and selected topics*. San Francisco: Holden Day, 1977. 493p.

DAVISON, P. J., HINKLEY, D. V. *Bootstrap methods and their application*. Cambridge: Cambridge University Press, 1977. 582p.

- EFRON, B. Bootstrap methods: another look at the jackknife. *Ann. Stat.* v.7, n.1, p.1-26, 1979.
- EFRON, B. Better bootstrap confidence intervals. *J. Am. Stat. Assoc.*, v.82, n.397, p.171-200, 1987. Com comentários.
- EFRON, B., LEPAGE, R. Introduction to bootstrap. In: LEPAGE, R., BILLARD, L. *Exploring the limits of bootstrap*. New York: John Wiley, 1992. p.3-10.
- EFRON, B., TIBSHIRANI, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.*, v.1, n.1, p.54-77, 1986.
- EFRON, B., TIBSHIRANI, R. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993. 436p.
- FARANTE, G., CHACÓN, R. Utilidad de los marcadores tumorales en oncología. In: CHACÓN, R., FARANTE, G. *Oncología Médica*. Milano: European School of Oncology, 1990. p.49-68.
- FELLER, W. *An introduction to probability theory and its applications*. 2.ed. New York: John Wiley, 1971. v.2, 669p.
- HALL, P. On the number of bootstrap simulations required to construct a confidence interval. *Ann. Stat.*, v.14, p.1453-62, 1986.
- KARLSSON, S., LÖTHGREN, M. Computationally efficient double bootstrap variance estimation. *Comput. Stat. Data Anal.*, v.33, p.237-47, 2000.
- LEITE, J. G., SINGER, J. M. Métodos assintóticos em estatística: fundamentos e aplicações. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9., 1990, São Paulo. *Anais...* São Paulo: IME/USP, 1990. 130p.
- MILLER, R. G. The jackknife – a review. *Biometrika*, v.61, n.1, p.1-15, 1974.
- MOOD, A. M., GRAYBILL, F. A., BOES, D. C. *Introduction to the theory of statistics*. 3.ed. New York: McGraw-Hill, 1974. 564p.
- ROHATGI, V. K. *An introduction to probability theory and mathematical statistics*. New York: John Wiley, 1976. 684p.
- RUBIN, D. B. The Bayesian bootstrap. *Ann. Stat.* v.9, n.1, p.130-4, 1981.
- SILVA, D. N. *O método bootstrap e aplicações à regressão múltipla*. 1995. 158f. Dissertação (Mestrado em Estatística) – Instituto de Mate-

mática, Estatística e Ciência da Computação, Universidade Estadual de Campinas, Campinas, 1995.

SOUZA, G. S. Métodos Empíricos. In: SOUZA, G. S. Introdução aos modelos de regressão linear e não-linear. Brasília: Embrapa, 1998. p.445-83.

TORRES, J. C. C. *Índice de risco de malignidade associando CA125, ultra-sonografia e estado menstrual no diagnóstico pré-operatório de tumores pélvicos*. 1998. 113f. Dissertação (Mestrado em Medicina) – Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, 1998.

WARD, B. G., SHEPHERD, J. H., The role of monoclonal antibodies in the diagnosis and investigation of ovarian cancer. In: BLACKLEDGE, G., CHAN, K. K. *Management of ovarian cancer*. London: Butterworths, 1986. p.46-59.

Recebido em 16.8.2000.

Apêndice 1

Ordens de magnitude

O termo $O(\cdot)$ denota a ordem de magnitude de uma seqüência $\{a_n\}_{n \geq 1}$ de números reais, sendo que $a_n = O(b_n)$ indica existir um número real $K > 0$ e um número inteiro positivo $n_0 = n_0(K)$, tal que $|a_n / b_n| \leq K$, para todo $n \geq n_0$, e $\{b_n\}_{n \geq 1}$ é também uma seqüência de números reais (Leite & Singer, 1990, p.9).

A notação $O_p(\cdot)$ indica a ordem de magnitude de uma seqüência estocástica. Segundo Leite & Singer (1990, p.39), se $\{X_n\}_{n \geq 1}$ é uma seqüência de variáveis aleatórias e $\{b_n\}_{n \geq 1}$ é uma seqüência de números reais (ou variáveis aleatórias), então $X_n = O_p(b_n)$ se, para qualquer número real $\eta > 0$ existir um número real positivo $K = K(\eta)$ e um número inteiro positivo $n_0 = n_0(\eta)$, tais que

$$P\left(\left|\frac{X_n}{b_n}\right| \geq K\right) \leq \eta, \forall n \geq n_0. \quad (70)$$

Assim, X_n é igual a $O_p(b_n)$ se a seqüência $\{X_n/b_n\}_{n \geq 1}$ for limitada em probabilidade, para todo n suficientemente grande (Leite & Singer, 1990, p.39).

Apêndice 2

Expansões de Edgeworth

Segundo Efron & LePage (1992), um grande avanço na teoria do método bootstrap foi obtido em 1981, por Singh, que demonstrou que o método pivotal ou t-bootstrap, em muitos casos, acompanha a correção de Edgeworth. Estimando um ponto da distribuição amostral da estatística estudentizada, o método *bootstrap* automaticamente produz respostas tão boas quanto expansões de Edgeworth com um termo, sem a necessidade de cálculos teóricos especiais.

Um exemplo simples para a expansão de Edgeworth é dado por Bickel & Doksum (1977, p.33). A aproximação da média \bar{X} para a distribuição normal utiliza apenas os dois primeiros momentos desta estatística. Uma melhor aproximação, através da expansão de Edgeworth, é encontrada utilizando-se também o terceiro e quarto momentos. Seja F_n a distribuição de

$$T_n = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}, \quad (71)$$

e sejam $\gamma_1 n$ e $\gamma_2 n$, respectivamente, os coeficientes de assimetria e curtose de T_n . Sob certas condições,

$$F_n(x) = \Phi(x) - \phi(x) \left[\frac{1}{6} \gamma_1 n H_2(x) + \frac{1}{24} \gamma_2 n H_3(x) + \frac{1}{72} \gamma_1^2 n H_5(x) \right] + O(n^{-1}), \quad (72)$$

onde $\Phi(x)$ e $\phi(x)$ são, respectivamente, a função distribuição acumulada e a densidade de uma variável aleatória normal padrão, e H_2 , H_3 e H_5 são polinômios definidos por

$$H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x \quad \text{e} \quad H_5(x) = x^5 - 10x^3 + 15x. \quad (73)$$

Feller (1971, p.532) mostra que $H_k(x)$ é denominado polinômio de Hermite de grau k , definido por

$$\frac{d^k}{dx^k} \varphi_k(x) = (-1)^k H_k(x) \varphi(x). \quad (74)$$

Vamos agora buscar uma forma genérica para a expansão de Edgeworth. Seja X uma variável aleatória com distribuição de probabilidade F . Para um real ζ , a função característica de F (ou, de X) é uma função φ , conforme definido por Feller (1971, p.499). Supondo que os momentos μ_3, \dots, μ_r existem, e que $|\varphi|^v$ é integrável para algum $v \geq 1$, então f_n existe para $n \geq v$ e, quando $n \rightarrow \infty$,

$$f_n(x) - \varphi(x) - \varphi(x) \sum_{k=3}^r n^{-(k/2)+1} P_k(x) = O\left(n^{-(r/2)+1}\right), \quad (75)$$

uniformemente em x . Esta expansão é a chamada expansão de Edgeworth para f_n , sendo que P_k são polinômios que dependem apenas de μ_3, \dots, μ_r , mas não de n e r . Uma demonstração para (75) é encontrada em Feller (1971, p.535).

Sendo R_k polinômios obtidos pela relação

$$\varphi(x) P_k(x) = \frac{d}{dx} \varphi(x) R_k(x), \quad (76)$$

e aplicando integrais, a expressão (75) pode ser reescrita da forma

$$F_n(x) - \Phi(x) - \varphi(x) \sum_{k=3}^r n^{-(k/2)+1} R_k(x) = O\left(n^{-(r/2)+1}\right). \quad (77)$$

Nota-se que R_k também são polinômios que dependem apenas de μ_3, \dots, μ_r , mas não de n e r (Feller, 1971, p.541).