

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

REAMOSTRAGEM BOOTSTRAP EM AMOSTRAGEM POR
CONJUNTOS ORDENADOS E INTERVALOS DE CONFIANÇA
NÃO PARAMÉTRICOS PARA A MÉDIA

Cesar Augusto Taconeli

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Mestre em Estatística.

SÃO CARLOS
2005

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

REAMOSTRAGEM BOOTSTRAP EM AMOSTRAGEM POR
CONJUNTOS ORDENADOS E INTERVALOS DE CONFIANÇA
NÃO PARAMÉTRICOS PARA A MÉDIA

Cesar Augusto Taconeli

SÃO CARLOS

2005

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

T119rb

Taconeli, Cesar Augusto.

Reamostragem bootstrap em amostragem por conjuntos ordenados e intervalos de confiança não paramétricos para a média / Cesar Augusto Taconeli. -- São Carlos : UFSCar, 2005.

120 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2005.

1. Estatística matemática. 2. Bootstrap (estatística). 3. Amostragem por conjuntos ordenados. 4. Intervalos de confiança. I. Título.

CDD: 519.54 (20^a)

*Dedico este trabalho à memória de
minha saudosa avó Tereza.*

*“Concedei-nos Senhor, Serenidade
necessária para aceitar as coisas que
não podemos modificar, Coragem para
modificar aquelas que podemos e
Sabedoria para distinguirmos umas
das outras.”*

(Autor desconhecido)

Agradecimentos

A Deus, que me abençoa e dá forças diariamente para que eu possa seguir em frente;

à meus pais, Antonio e Alzira, que não pouparam esforços para que eu cumprisse mais esta etapa de minha formação;

à meus irmãos, Fábio e João Paulo, que me apoiaram e me motivaram desde o primeiro instante;

à minha orientadora, Profa. Dra. Maria Cecilia Mendes Barreto, pela confiança depositada em meu trabalho, pela amizade, pela dedicação e pelo modo brilhante com que conduziu nossos estudos;

aos professores, Dr. Francisco Antonio Rojas Rojas, Dra Lúcia Pereira Barroso e Dra. Teresa Cristina Martins Dias, cujas críticas e sugestões contribuíram para o aperfeiçoamento deste trabalho;

aos professores e funcionários do Departamento de Estatística da UFSCar, que sempre se puseram a disposição para me atender em minhas dificuldades;

aos meus colegas de mestrado, e a todos os meus amigos, que comigo compartilharam os bons e maus momentos vividos ao longo dos últimos anos;

à Coordenação de Aperfeiçoamento Pessoal de Ensino Superior – CAPES, pela assistência financeira;

a todos que estiveram comigo, de maneira direta ou indireta, durante o desenvolvimento deste trabalho.

Resumo

A amostragem por conjuntos ordenados é uma alternativa prática e eficiente no que concerne à obtenção de estimativas mais precisas frente à impossibilidade de extração de uma amostra numerosa, seja devido a dificuldades na mensuração da variável de interesse ou a um elevado custo inerente a obtenção de tais medidas. A aplicação deste delineamento amostral torna-se viável caso seja possível ordenar amostras extraídas aleatoriamente de maneira eficiente, de acordo com o valor da variável de interesse, sem de fato medi-las, mas baseado apenas em um critério pré-estabelecido, que pode ser alguma variável concomitante altamente correlacionada ou mesmo mediante algum julgamento pessoal.

Introduzida por McIntyre (1952), a amostragem por conjuntos ordenados propicia a estimação de diversos parâmetros com um relevante ganho em termos de precisão. Um estimador para a média populacional é a média da amostra por conjuntos ordenados ($\bar{\bar{X}}$), proposto por McIntyre com aplicações, inicialmente, na estimação da produção média de pastagens. Cesário e Barreto (2003) apresentam uma alternativa paramétrica na obtenção de intervalos de confiança bootstrap para a média de populações com distribuição normal via amostragem por conjuntos ordenados.

Dada a restrição quanto à seleção de grandes amostras, a suposição de alguma distribuição para a variável de interesse muitas vezes não é razoável, gerando estimativas pouco confiáveis. Neste contexto, o estudo e a proposição de estimadores intervalares não paramétricos para a média, elaborados a partir de um esquema de seleção de amostras capaz de gerar estimativas precisas sob circunstâncias adversas, como é a amostragem por conjuntos ordenados, mostra-se altamente relevante, sendo o objeto de estudo deste trabalho.

Os intervalos de confiança analisados são obtidos através de um esquema original de re-amostragem bootstrap, fundamentado em amostragem por conjuntos ordenados, seguindo algoritmos propostos neste trabalho. A análise das propriedades destes intervalos foi realizada a partir de um amplo estudo via simulação, que evidenciou uma significativa melhora das estimativas propostas, quando comparado àquelas convencionais, baseadas em amostragem aleatória simples, especialmente em relação à precisão de tais estimativas.

Abstract

Ranked set sampling is an efficient and practice way to obtain more precise estimative when the sample size is small because of the high cost or difficulties to measure the interest variable. Using rough and cheap qualitative or quantitative information, the sample units are ranked before their effective measurement.

In 1952, McIntyre introduced the ranked set sample design to estimate the average yields from plots of cropland, using the ranked set sample mean, \bar{X} . Cesario and Barreto (2003) have shown a parametric version of bootstrap confidence intervals for normal distribution mean.

Because of the restriction of small sample size, the distributional assumption may not be reasonable, producing no liable estimates. So the study and proposition of precise interval estimators of the population mean could be relevant and are the main interest of this work.

Using resampling methods, we propose in this work an extension of bootstrap resampling for ranked set sampling. A simulation study is conducted to the properties of single random sample bootstrap confidence intervals and the similar using our version for ranked set sampling. The analysis of the simulation study have shown the gain of precision for using the ranked set sampling bootstrap confidence intervals in the population mean.

Sumário

1 Introdução	1
2 Seleção da amostra por conjuntos ordenados e alguns estimadores para a média . .	4
2.1 Esquema de seleção de uma amostra por conjuntos ordenados	4
2.2 Critérios de ordenação e modelos	6
2.3 Algumas propriedades de estatísticas de ordem	7
2.4 Propriedades de variáveis com distribuição da família locação e escala	9
2.5 Estimadores da média via RSS	10
2.6 Modelo de ordenação por julgamento pessoal	13
2.7 Modelo de ordenação por variável concomitante	15
3 Estimadores para outros parâmetros e intervalos de confiança sob o delineamento de Amostragem por Conjuntos Ordenados	17
3.1 Estimadores para a variância	17
3.2 Estimadores para os quantis	19
3.3 Estimador da função distribuição acumulada (F^*)	22
3.3.1 Caracterização de uma amostra por conjuntos ordenados	24
3.3.2 Eficiência de F^*	25
3.4 Estimação intervalar via amostragem por conjuntos ordenados	27
4 Intervalos de confiança Bootstrap sob Amostragem Aleatória Simples	29
4.1 O bootstrap	29
4.2 Bootstrap paramétrico	30
4.3 Bootstrap não paramétrico	31
4.4 Intervalos de confiança bootstrap	32

4.4.1- Acurácia dos intervalos de confiança	32
4.4.2 Intervalo bootstrap normal	33
4.4.3 Intervalo bootstrap básico	34
4.4.4 Intervalo bootstrap studentizado	36
4.4.5 Intervalo bootstrap percentil	37
4.4.6 O método BCa	38
5 Propostas originais de reamostragem bootstrap em Amostragem por Conjuntos Ordenados	41
5.1 Proposta original de reamostragem bootstrap em conjuntos ordenados perfeitamente e extensões da estimação intervalar para a média populacional	42
5.2 Proposta original de reamostragem bootstrap em conjuntos ordenados com ordenação imperfeita	45
5.3 Aplicação dos intervalos propostos em dados reais	47
5.4 Procedimento para análise da adequabilidade dos intervalos propostos e planejamento do estudo via simulação	49
5.5 Determinação do número de reamostragens e intervalos de confiança simulados via análise de convergência	51
6 Estudo via simulação das propriedades dos intervalos de confiança originários dos esquemas de reamostragem propostos	55
6.1 Estimação da média das medições médias diárias de MP ₁₀ na estação Ibirapuera, nos anos de 2000 e 2001	56
6.2 Estimação da média de uma população com distribuição normal	63
6.3 Estimação da média de uma população com distribuição exponencial	70
6.4 Estimação da média de uma população com distribuição uniforme	77
6.5 Estimação da média de uma população com distribuição lognormal	84
6.6 Estimação da média de uma população com distribuição Gumbell	91
7 Conclusão e propostas futuras	98

Apêndice	100
Apêndice A: Programa para gerar intervalos de confiança bootstrap não paramétricos dos tipos normal, básico e percentil, via RSS com ordenação perfeita, para a média de MP10 na estação Ibirapuera em 2000/2001	100
Apêndice B: Programa para gerar intervalos de confiança bootstrap não paramétricos do tipo BCa, via RSS com ordenação imperfeita, para a média de uma variável com distribuição normal (0,1)	106
Apêndice C: Programa para gerar intervalos de confiança bootstrap não paramétricos do tipo studentizado, via RSS com ordenação perfeita, para a média de uma variável com distribuição lognormal (0,1)	112
 Referências	 117

Lista de figuras

Figura 1 -Representação da seleção de uma amostra por conjuntos ordenados de tamanho $n = 3$, com $m = 2$ réplicas	6
Figura 2 - Valores de $Var_B(\bar{\bar{X}})$ calculados via bootstrap a partir de B reamostras, originárias de cinco amostras distintas extraídas das medições médias diárias de MP ₁₀	52
Figura 3 -Valores de $\bar{\bar{X}}_{(0,05)}$ e $\bar{\bar{X}}_{(0,95)}$ para diferentes valores de B , calculados via bootstrap, a partir de cinco amostras distintas extraídas das medições médias diárias de MP ₁₀	53
Figura 4 -Probabilidade de cobertura calculada para diferentes números de intervalos de confiança bootstrap do tipo percentil de 90, 95 e 99% de confiança para a média das medições médias diárias de MP ₁₀	54
Figura 5 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média das verificações médias diárias de MP10 na estação Ibirapuera (2000/2001), via SRS e RSS com ordenação perfeita e imperfeita	57
Figura 6 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média das verificações médias diárias de MP10 na estação Ibirapuera (2000/2001), via SRS e RSS com ordenação perfeita e imperfeita	58
Figura 7 - Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição normal, via SRS e RSS com ordenação perfeita e imperfeita.	64
Figura 8 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição normal, via SRS e RSS com ordenação perfeita e imperfeita	65

Figura 9 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição exponencial, via SRS e RSS com ordenação perfeita e imperfeita	71
Figura 10 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição exponencial, via SRS e RSS com ordenação perfeita e imperfeita	72
Figura 11 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição uniforme, via SRS e RSS com ordenação perfeita e imperfeita	78
Figura 12 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição uniforme, via SRS e RSS com ordenação perfeita e imperfeita	79
Figura 13 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição lognormal, via SRS e RSS com ordenação perfeita e imperfeita	85
Figura 14 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição lognormal, via SRS e RSS com ordenação perfeita e imperfeita	86
Figura 15 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição Gumbell, via SRS e RSS com ordenação perfeita e imperfeita	92
Figura 16 -Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição Gumbell, via SRS e RSS com ordenação perfeita e imperfeita	93

Capítulo 1

Introdução

A dificuldade em se obter uma amostra numerosa torna necessária a utilização de um procedimento mais refinado de extração de amostras, capaz de compensar tal limitação produzindo estimadores mais precisos para os parâmetros de interesse. A Amostragem por Conjuntos Ordenados (*Ranked Set Sampling – RSS*) é uma boa alternativa para os casos em que se tem uma variável de interesse cuja mensuração seja cara ou de difícil obtenção, havendo no entanto a possibilidade de ordenação das unidades amostrais utilizando algum critério subjetivo, muitas vezes visual, ou alguma variável concomitante fortemente correlacionada, de maneira simples, barata e eficiente.

Introduzida por McIntyre (1952) com vistas à estimação da produção média de pastagens, a amostragem por conjuntos ordenados vem sendo estudada e difundida desde então. Sua aplicação torna-se conveniente à medida que possibilita a obtenção de estimativas mais precisas para uma grande variedade de funções e parâmetros populacionais mediante extração de uma amostra de tamanho reduzido.

Diversas publicações atestam a conveniência, em termos de ganho de precisão, da utilização da *RSS* com relação à Amostragem Aleatória Simples (*Simple Random Sample – SRS*). Takahasi e Wakimoto (1968) comprovam a menor variância do estimador média amostral via *RSS* em relação ao mesmo estimador obtido via *SRS*, sob ordenação perfeita (quando o critério utilizado para ordenação é preciso), enquanto Dell e Clutter (1972) a validade desta propriedade ainda que na presença de erros de ordenação, embora os resultados sejam menos satisfatórios para critérios de ordenação pouco precisos.

Stokes (1980) apresenta um estimador da variância populacional usando amostragem por conjuntos ordenados, que é assintoticamente não viciado e tem menor erro quadrático médio que o seu análogo para *SRS*, embora a melhora não seja tão significativa quanto no caso da estimação da média. Stokes e Sager (1988) apresentam um estudo da estimação da função distribuição através da distribuição empírica obtida a partir da amostra por conjuntos ordenados e Chen (1999) apresenta um estudo da estimação da função densidade, que mais uma vez evidencia a eficiência de *RSS* sobre *SRS*. O trabalho de Chen (2000) introduz um estimador não paramétrico para o p -quantil, estuda algumas de suas propriedades e apresenta um intervalo de confiança assintótico.

Recentes estudos têm abordado a estimação paramétrica via *RSS*. Stokes (1995) introduz a análise paramétrica por conjuntos ordenados na estimação dos parâmetros de locação e de escala das distribuições pertencentes à família locação e escala. Barnett (1999) apresenta resultados alcançados através de simulação que ratificam um relevante aumento na precisão de diferentes estimadores de μ_X , a média populacional, via *RSS*, em relação ao estimador média amostral via *SRS* (\bar{X}), para amostras da distribuição lognormal e Gumbell. Yu et al (1999) propõem um estimador paramétrico para a variância, como uma extensão não viciada do estimador proposto por Stokes (1980).

Cesário e Barreto (2003), Raminelli (2003) e Garcia (2004) apresentam, respectivamente, algumas alternativas paramétricas na obtenção de intervalos de confiança para a média, quantis e variância de populações com distribuição normal via *RSS*. Utilizando o método de reamostragem bootstrap (Davison e Hinkley, 1997), alcançam resultados bastante expressivos, viabilizando a aplicação de tais estimativas. Taconeli e Barreto (2003) se atêm ao estudo de propriedades assintóticas e proposição de intervalos de confiança para a média populacional.

Frente à inexistência de outros estudos acerca da estimação intervalar não paramétrica sob *RSS*, este trabalho tem como objetivo principal propor e avaliar, através de um detalhado estudo via simulação, um esquema original de reamostragem bootstrap baseado em amostragem por conjuntos ordenados quando a ordenação é realizada mediante julgamento visual, com aplicação direta na obtenção de intervalos de confiança para a média populacional. É abordado um procedimento de reamostragem baseado exclusivamente nas amostras selecionadas (bootstrap não paramétrico), buscando re-adequar as diferentes metodologias aplicáveis na determinação de

intervalos de confiança bootstrap, baseadas originalmente em amostragem aleatória simples, para o caso em que as amostras são selecionadas por conjuntos ordenados.

As seções iniciais do Capítulo 2 apresentam alguns conceitos referentes a estatísticas de ordem e variáveis com distribuição da família locação e escala, fundamentais para o estudo do vício e da precisão de estimadores presentes nos Capítulos 2 e 3. Na seqüência, são apresentados o esquema de seleção de uma amostra por conjuntos ordenados, além de alguns estimadores para a média populacional sob tal delineamento. Uma explanação acerca dos dois modelos com erros de ordenação encerra o capítulo.

A aplicação da amostragem por conjuntos ordenados na estimação de quantis, da variância e da função distribuição de probabilidades é enfocada no Capítulo 3. Estudos e resultados relativos à estimação intervalar paramétrica da média, variância e quantis complementam o capítulo.

O Capítulo 4, por sua vez, trata exclusivamente do procedimento de reamostragem bootstrap, descrevendo de maneira detalhada o método em suas duas versões (paramétrica e não paramétrica) e apontando sua aplicabilidade na obtenção de diferentes estimativas intervalares sob o delineamento de *SRS*.

O Capítulo 5 apresenta as propostas originais de reamostragem bootstrap por conjuntos ordenados, bem como a extensão na determinação de estimadores intervalares para a média. Traz ainda o planejamento do estudo via simulação, utilizado na avaliação da acurácia e da precisão das estimativas sugeridas.

Os resultados alcançados através deste estudo para amostras geradas a partir de cinco distribuições teóricas (normal, exponencial, uniforme, lognormal e Gumbell), ou extraídas de uma população real, com dados fornecidos pela CETESB (Companhia de Tecnologia de Saneamento Ambiental), podem ser verificados no Capítulo 6.

Finalizando, o Capítulo 7 apresenta as principais conclusões extraídas deste estudo, acompanhadas por algumas propostas para estudos futuros.

Capítulo 2

Seleção da amostra por conjuntos ordenados e alguns estimadores para a média

Neste capítulo, são apresentados e discutidos diversos aspectos relativos a extração de uma amostra por conjuntos ordenados e aos estimadores existentes para a média populacional. A Seção 2.1 descreve o procedimento utilizado na obtenção da amostra. Os critérios utilizados para a ordenação das amostras são descritos e classificados na Seção 2.2. As Seções 2.3 e 2.4 introduzem alguns conceitos referentes a estatísticas de ordem e à distribuição de variáveis da família locação escala, para a compreensão de algumas propriedades dos estimadores da média populacional presentes na Seção 2.5 e de outros parâmetros presentes no Capítulo 3. As Seções 2.6 e 2.7 apresentam os diferentes modelos com erros de ordenação, conseqüentes da utilização dos dois diferentes tipos de critérios na ordenação das amostras: julgamento pessoal ou variável concomitante.

2.1 Esquema de seleção de uma amostra por conjuntos ordenados

O esquema de seleção de uma amostra por conjuntos ordenados é apresentado na seqüência.

1. Seleção de n amostras aleatórias simples de tamanho n da população de interesse, com reposição;

2. Ordenação de cada uma das n amostras, em ordem crescente da variável de interesse, baseado em algum critério subjetivo (julgamento pessoal) ou através de uma variável concomitante fortemente correlacionada, sem de fato medir as unidades amostrais;
3. Mensuração na primeira amostra daquela observação que se julgou apresentar o menor valor da variável em estudo, na segunda amostra da unidade amostral que foi julgada como tendo a segunda menor medida, e assim por diante, até a n -ésima amostra, em que se deve medir a n -ésima menor unidade amostral (u.a.), ou seja, aquela que se julgou apresentar a maior medida;
4. Repetição dos passos 1 a 3, m vezes.

Ao final deste procedimento, são medidas $N = mn$ observações, que compõem a amostra por conjuntos ordenados, representada por

$$(x_{[r]i}; r = 1, 2, \dots, n; i = 1, 2, \dots, m). \quad (2.1)$$

Neste caso, $x_{[r]i}$ refere-se à observação que foi julgada como estatística de ordem r , no i -ésimo ciclo realizado. Subentende-se que o índice r denota, além da ordem que foi julgada a unidade amostral, o “número” da amostra de origem, ou seja, $x_{[r]i}$ foi julgada como de ordem r na r -ésima amostra do i -ésimo ciclo ou repetição. Vale lembrar que caso a ordenação seja perfeita, o índice r de fato identifica a qual classe das estatísticas de ordem corresponde a observação. Ressalte-se o fato de as observações serem independentes, já que provém de amostras selecionadas independentemente, mas não identicamente distribuídas, devido à classificação em diferentes ordens.

A utilização de réplicas possibilita a obtenção de uma amostra maior sem a necessidade de ordenação de uma grande quantidade de unidades amostrais. Com o objetivo de selecionar uma amostra final de tamanho $N = 15$, amostras de tamanho 15 deverão ser ordenadas, caso não sejam utilizadas réplicas. A comparação e posterior ordenação de quinze unidades amostrais pode não ser uma tarefa fácil. Neste caso, a utilização de $m = 3$ réplicas permitiria uma redução no número de elementos a serem ordenados em cada amostra para $n = 5$, mantendo o tamanho final

da amostra em $N = 15$, mas implicando em uma maior facilidade e eficiência no processo de ordenação.

A Figura 1 ilustra o esquema de seleção de uma amostra de conjuntos ordenados de tamanho $n = 3$, com duas replicações ($m = 2$). Supõe-se que a variável de interesse é a altura das árvores e o critério utilizado na sua ordenação é o tamanho aproximado das mesmas, acessado através de uma simples inspeção visual. Cada uma das amostras inicialmente selecionadas encontra-se disposta na horizontal, estando assinaladas as unidades amostrais que irão compor a amostra final, sendo as únicas de fato mensuradas quanto à variável de interesse.

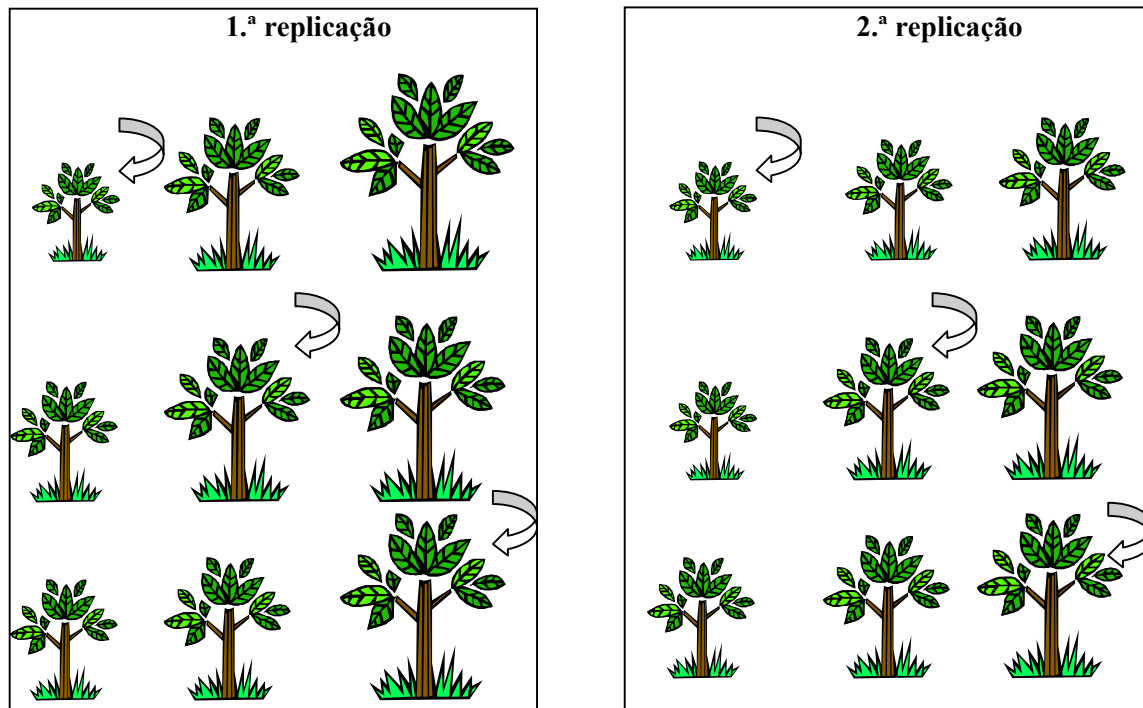


Figura 1- Representação da seleção de uma amostra por conjuntos ordenados de tamanho $n = 3$, com $m = 2$ réplicas.

2.2 Critérios de ordenação e modelos

Pode-se classificar o procedimento de amostragem em conjuntos ordenados de duas maneiras, de acordo com o desempenho do procedimento adotado para ordenação das unidades amostrais: ordenação perfeita, quando o método utilizado não apresenta erros, e ordenação

imperfeita, nas situações em que a ordenação pode apresentar falhas. Como dito anteriormente, o critério de ordenação das amostras pode se basear em uma opinião subjetiva, fornecida pelo próprio pesquisador, ou em alguma variável concomitante, que seja de fácil mensuração e apresente uma forte correlação com a variável de interesse. Cada uma destas possibilidades de critérios para ordenação especifica um dos dois diferentes modelos aplicáveis em amostragem por conjuntos ordenados.

Caso se queira inferir sobre a altura de árvores, por exemplo, as unidades amostrais podem ser ordenadas de maneira eficiente de acordo com julgamento visual, como exemplificado na Seção 2.1. Da mesma forma, amostras de volumes de água de um rio acondicionadas em recipientes transparentes podem ser ordenadas de acordo com julgamento visual, caso se tenha interesse em estimar o índice de uma determinada impureza. O modelo de julgamento pessoal se faz útil em situações semelhantes a estas.

Considerando ainda a amostra de árvores, pode-se ter ciência de uma forte correlação entre a altura e o perímetro do tronco a uma determinada distância do solo, permitindo a ordenação das unidades amostrais sem maiores custos ou dificuldades. Caso o interesse seja ordenar famílias de acordo com suas condições econômicas, pode-se utilizar algum critério que atribua pontos de acordo com as posses apresentadas. Desta maneira, utilizar-se-ia as pontuações obtidas pelas famílias como base para o procedimento de ordenação. Nestas situações, aplica-se o modelo de ordenação via variável concomitante.

2.3 Algumas propriedades de estatísticas de ordem

Para uma amostra aleatória simples de tamanho n (x_1, x_2, \dots, x_n) , considera-se as unidades amostrais como independentes e identicamente distribuídas, com função distribuição acumulada F_X e função densidade de probabilidade f_X . Tal propriedade não se estende para uma amostra por conjuntos ordenados, uma vez que, ordenada a i -ésima amostra, verifica-se a seguinte situação

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} \quad (2.2)$$

sendo $X_{(r)}$ ($r=1,2,\dots,n$) a estatística de ordem r . Neste caso, as v.a.s são independentes, mas identicamente distribuídas somente para algum valor fixado de r . Ou seja, $X_{(r)}$ tem função distribuição acumulada F_r , que segundo David (1981) pode ser determinada da seguinte forma

$$\begin{aligned} F_r(x) &= P\{X_{(r)} \leq x\} = \\ &= P\{\text{pelo menos } r \text{ dos } X_i \text{ são menores ou iguais a } x\} = \\ &= \sum_{i=r}^n \binom{n}{i} [F(x)]^i [1-F(x)]^{n-i}. \end{aligned} \quad (2.3)$$

Ainda segundo David (1981), a função $F_r(x)$ pode ser expressa da seguinte forma

$$F_r(x) = \int_0^{F(x)} \frac{1}{B(r, n-r+1)} t^{r-1} (1-t)^{n-r} dt \quad (2.4)$$

sendo $B(r, n-r+1) = \int_0^1 x^{r-1} (1-x)^{n-r} dx$.

Assumindo que X seja uma v.a. contínua, David (1981) determina a fdp da r -ésima estatística de ordem simplesmente derivando $F_r(x)$, obtendo a seguinte expressão

$$\begin{aligned} f_r(x) &= \frac{1}{B(r, n-r+1)} \frac{d}{dx} \int_0^{F(x)} t^{r-1} (1-t)^{n-r} dt = \\ &= \frac{1}{B(r, n-r+1)} [F(x)]^{r-1} [1-F(x)]^{n-r}. \end{aligned} \quad (2.5)$$

Conhecendo f_r , tem-se condições de calcular a média e a variância de $X_{(r)}$, como segue

$$E[X_{(r)}] = \mu_{(r)} = \int_{-\infty}^{\infty} x f_r(x) dx = \frac{1}{B(r, n-r+1)} \int_{-\infty}^{\infty} x [F(x)]^{r-1} [1-F(x)]^{n-r} dx \quad (2.6)$$

$$Var[X_{(r)}] = \sigma_{(r)}^2 = E[X_{(r)}^2] - [E[X_{(r)}]]^2 = \frac{1}{B(r, n-r+1)} \left[\int_{-\infty}^{\infty} x^2 [F(x)]^{r-1} [1-F(x)]^{n-r} dx - \mu_{(r)}^2 \right] \quad (2.7)$$

2.4 Propriedades de variáveis com distribuição da família locação e escala

Algumas propriedades de estimadores obtidos via amostragem por conjuntos ordenados envolvem conhecimentos de variáveis aleatórias da família locação e escala. Considere uma variável aleatória X com distribuição $F_{\mu,\sigma}(X)$, sendo μ e σ , respectivamente, os parâmetros de locação e escala. Segundo Pigato e Barreto (2004), a função de distribuição desta variável pode ser expressa na forma

$$F_0\left(\frac{X - \mu}{\sigma}\right). \quad (2.8)$$

A média, μ_x , e a variância, σ_x^2 , são funções dos parâmetros de locação e escala (μ e σ), e da média, μ_0 , e variância, σ_0^2 , da distribuição básica F_0 . Tais relações podem ser expressas da seguinte forma: $\mu_x = \mu + \sigma\mu_0$, $\sigma_x^2 = \sigma^2\sigma_0^2$. Quando F_0 define uma distribuição simétrica, $\mu_0 = 0$, de modo que $\mu_x = \mu$.

Considere $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ estatísticas de ordem de uma amostra de tamanho n e seja

$$U_{(r)} = \frac{X_{(r)} - \mu}{\sigma}, \quad r = 1, \dots, n \quad (2.9)$$

as variáveis ordenadas reduzidas. Para $r = 1, \dots, n$, $E(U_{(r)}) = \alpha_r$ e $Var(U_{(r)}) = v_r$ dependem apenas do tamanho da amostra, da ordem e da função distribuição da variável reduzida $(X - \mu)/\sigma$ (David, 1981). Tais quantidades podem ser facilmente determinadas mediante a utilização de algum software estatístico ou consultando resultados em bibliografias adequadas. Pearson e Hartley (1976) apresentam os momentos das estatísticas de ordem para a distribuição normal e tamanhos de amostras até 20.

2.5 Estimadores da média via RSS

O estimador não viciado para $E(X)$ via *RSS* proposto por McIntyre (1952) é dado pela média da amostra por conjuntos ordenados

$$\bar{\bar{X}} = \frac{1}{mn} \sum_{r=1}^n \sum_{i=1}^m X_{[r]i} \quad (2.10)$$

com variância dada por

$$Var(\bar{\bar{X}}) = \frac{1}{m^2 n^2} Var\left(\sum_{i=1}^m \sum_{r=1}^n x_{[r]i}\right) = \frac{1}{m^2 n^2} m Var\left(\sum_{r=1}^n (\mu + \sigma U_{[r]})\right) = \frac{1}{mn^2} \sigma^2 \sum_{r=1}^n v_r. \quad (2.11)$$

A expressão (2.11) pode ser reescrita da seguinte maneira

$$Var(\bar{\bar{X}}) = \frac{\sigma_X^2}{mn} - \frac{1}{mn^2} \sigma^2 \sum_{r=1}^n (\alpha_r - \bar{\alpha})^2 \quad (2.12)$$

que indica o ganho em termos de precisão decorrente da utilização do estimador $\bar{\bar{X}}$ em detrimento a \bar{X} , o estimador média amostral via *SRS*.

A precisão relativa deste estimador com respeito ao estimador \bar{X} , para uma amostra de tamanho mn , é

$$RP = \frac{var(\bar{X})}{var(\bar{\bar{X}})} = \frac{\frac{1}{mn} \sigma^2}{\frac{1}{n^2 m} \sum_{r=1}^n v_r} = \left(1 - \frac{1}{n} \sum_{r=1}^n \frac{(\alpha_r - \bar{\alpha})^2 \sigma^2}{\sigma_X^2}\right)^{-1}. \quad (2.13)$$

Diversas outras publicações atestam a conveniência, em termos de ganho de precisão, da utilização de $\bar{\bar{X}}$ em comparação com \bar{X} . Takahasi e Wakimoto (1968) comprovam a menor variância do estimador da média via *RSS* em relação a \bar{X} sob ordenação perfeita. Dell e Clutter (1972) apresentam bons resultados mesmo com erros de ordenação, mas os resultados tornam-se menos satisfatórios conforme aumenta o erro na ordenação das unidades amostrais.

Existem outros estimadores para a média populacional baseados na amostra por conjuntos ordenados. O BLUE (*best linear unbiased estimator*), por exemplo, consiste na ponderação da amostra, atribuindo diferentes pesos às unidades amostrais. Barnett e Moore (1997) comprovam o relevante ganho alcançado quando da utilização deste estimador considerando variáveis aleatórias com distribuição simétrica e assimétrica.

Os estimadores do parâmetro de locação e do parâmetro de escala, considerando uma única replicação do processo de amostragem, extraídos de Barnett e Moore (1997), são dados por

$$\mu^* = \sum_{r=1}^n \gamma_r x_r \quad \text{e} \quad \sigma^* = \sum_{r=1}^n \eta_r x_r \quad (2.14)$$

respectivamente, sendo

$$\gamma_r = \frac{(1/v_r) \left[\sum_{j=1}^n (\alpha_j^2 / v_j) - \alpha_r \sum_{j=1}^n (\alpha_j / v_j) \right]}{\Delta} \quad ; \quad \eta_r = \frac{(1/v_r) \left[\alpha_r \sum_{j=1}^n (1/v_j) - \sum_{j=1}^n (\alpha_j / v_j) \right]}{\Delta} \quad (2.15)$$

com

$$\Delta = \left[\sum_{r=1}^n \left(\frac{\alpha_r^2}{v_r} \right) \right] \left[\sum_{r=1}^n \left(\frac{1}{v_r} \right) \right] - \left[\sum_{r=1}^n \frac{\alpha_r}{v_r} \right]^2. \quad (2.16)$$

Quando a variável de interesse é assimétrica, o BLUE constitui uma combinação linear de μ^* e σ^* , dada por

$$\mu_x^* = \mu^* + \bar{\alpha} \sigma^* \quad (2.17)$$

onde μ_x^* é o BLUE da média populacional por RSS e $E(U) = \bar{\alpha}$.

Barnett e Moore (1997) obtiveram também a variância deste estimador

$$Var(\mu_x^*) = \frac{\sigma^2}{\Delta} \left\{ \sum_{r=1}^n \left(\frac{\alpha_r^2}{v_r} \right) - 2\bar{\alpha} \sum_{r=1}^n \left(\frac{\alpha_r}{v_r} \right) + \bar{\alpha}^2 \sum_{r=1}^n \left(\frac{1}{v_r} \right) \right\} = \frac{\sigma^2}{\Delta} \sum_{r=1}^n \left(\frac{(\alpha_r - \bar{\alpha})^2}{v_r} \right). \quad (2.18)$$

A precisão relativa de μ_x^* para o estimador $\bar{\bar{X}}$ é, então, dada por

$$RP(\mu_x^*, \bar{\bar{X}}) = \frac{Var(\bar{\bar{X}})}{Var(\mu_x^*)} = \frac{\sum_{r=1}^n v_r}{\left\{ n^2 \sum_{r=1}^n \left[\frac{(\alpha_r - \bar{\alpha})^2}{v_r} \right] \right\}}. \quad (2.19)$$

É possível verificar que μ_x^* não pode ser menos eficiente que $\bar{\bar{X}}$, e, portanto o valor acima será necessariamente maior ou igual a um.

Kaur et al (1997) propõem uma alternativa na estimação da média de variáveis com distribuição assimétrica à direita. Este estimador parte de uma alteração do delineamento amostral. Aqui, seleciona-se $n + k - 1$ amostras de tamanho n , mensurando-se na primeira aquela observação que se julga apresentar a menor medida quanto à variável de interesse, na segunda amostra a segunda menor observação, e assim por diante, até que nas k amostras finais, mensuram-se as u.a.s que se julga terem os maiores valores. Desta forma, utiliza-se a média destes k valores, aliada às demais observações tomadas, buscando melhor representar o efeito da assimetria presente na distribuição dos dados, produzindo o seguinte estimador

$$\bar{\bar{X}}_r = \sum_{i=1}^m \left[\sum_{r=1}^{n-1} X_{[r]i} + \sum_{t=1}^k X_{[n]t} / k \right] / mn \quad (2.20)$$

cuja variância é dada por

$$Var(\bar{\bar{X}}_r) = \left[\sum_{r=1}^{n-1} v_r + v_n / k \right] / mn^2. \quad (2.21)$$

Para se comparar o estimador $\bar{\bar{X}}_r$ com $\bar{\bar{X}}$, deve-se utilizar tamanhos de amostras idênticos. Deste modo, para o estimador usual ($\bar{\bar{X}}$) considera-se uma amostra de tamanho $m(n-1+r)$, que origina a seguinte expressão para a precisão relativa:

$$RP(\bar{X}_r, \bar{X}) = \frac{n^2 \sum_{r=1}^{h-1+r} v_r}{(n-1+k)^2 \left(\sum_{r=1}^{h-1} v_r + v_n / k \right)} . \quad (2.22)$$

Barnett (1999) apresenta aplicações destes dois estimadores em dados com distribuição lognormal e Gumbell, e verifica expressivas melhoras quanto à precisão da estimação da média populacional, em relação ao estimador \bar{X} .

Baseado na média das observações julgadas como sendo as medianas das amostras selecionadas, o *Median Ranked Set Sampling* (MRSS) fornece uma estimativa não viciada para a média populacional, apenas nos casos em que a distribuição de X é simétrica. O estimador, proposto em Muttlak (1997) e citado em Muttlak (1998), considera duas variantes, conseqüentes da utilização de amostras de tamanho par ou ímpar. Para simplificar as notações, considere $X_{(r:n/2)j}$ a mediana da r -ésima amostra do j -ésimo ciclo. Pode-se escrever o estimador média amostral via MRSS da seguinte forma

$$\bar{X}_{MRSS} = \frac{1}{mn} \sum_{r=1}^n \sum_{j=1}^m X_{(r:n/2)j} \quad (2.23)$$

cuja variância é dada por

$$Var(\bar{X}_{MRSS}) = \frac{1}{m^2 n} \sum_{r=1}^n \sigma_{[r:n/2]}^2 \quad (2.24)$$

onde $\sigma_{[r:n/2]}^2 = E(X_{[r:n/2]} - E(X_{[r:n/2]}))^2$

Muttlak (1998) apresenta uma comparação deste estimador com a média amostral via conjuntos ordenados (\bar{X}), verificando maior precisão do estimador \bar{X}_{MRSS} em todas as situações analisadas. Considerando estimadores do tipo regressão, comprova-se maior eficiência quando as unidades amostrais são ordenadas através de uma variável concomitante, a menos que as variáveis apresentem uma correlação excessivamente acentuada.

2.6 Modelo de ordenação por julgamento pessoal

O modelo com erros de ordenação por julgamento pessoal pode ser escrito da seguinte maneira

$$Y_i = X_i + e_i \quad , \quad e_i \sim N(0, \sigma_e^2) \quad (2.25)$$

A variável Y representa o valor da variável X (que é de interesse), acrescido de um erro aleatório, conseqüente das falhas de ordenação. Este erro é normalmente distribuído com média zero e variância σ_e^2 .

Na presença de erros de ordenação, Dell e Clutter (1972) verificam, através de simulação, um decréscimo da precisão relativa apresentada em (2.13) decorrente do aumento da variabilidade causado por incorreções na ordenação da amostra, para dados com distribuições simétricas (uniforme e normal) e assimétrica (exponencial). Este decréscimo torna-se menos relevante à medida que o julgamento utilizado como critério de ordenação torna-se mais preciso.

Nahhas et al (2003) mostraram que, usando o modelo de erros de ordenação de Dell e Clutter (1972) e supondo que a variável X é normalmente distribuída com variância σ^2 e Y tem variância $\sigma^2 + \sigma_e^2$, a precisão relativa pode ser escrita como

$$RP = \left\{ 1 - \rho_{xy}^2 \sum_{r=1}^n \alpha_{[r]}^2 / n \right\}^{-1} \quad (2.26)$$

onde $\alpha_{[r]}$ é a média da r -ésima estatística de ordem de uma amostra de tamanho n com distribuição normal padronizada e $\rho_{xy} = \text{corr}(X, Y) = \sigma / (\sigma_e^2 + \sigma^2)^{1/2}$.

Nahhas e Wolfe (manuscrito não publicado, citado em Nahhas et al (2003)) mostraram que, se a variável de interesse tiver distribuição normal, a probabilidade esperada de ordenar incorretamente um par (X_i, Y_i) é

$$p = \left[\{1/2\} - \left\{ \tan^{-1}(\sigma / \sigma_e) \right\} / \pi \right]. \quad (2.27)$$

Tem-se então que

$$(\sigma / \sigma_e) = \left[\tan^2 \{ \pi(1-2p) / 2 \} \right]^{-1} \quad (2.28)$$

e

$$\rho_{xy}^2 = f(p) = \left[1 + \left[\tan^2 \left\{ \pi(1-2p)/2 \right\} \right]^{-1} \right]. \quad (2.29)$$

Segue que

$$RP = \left\{ 1 - \frac{f(p)}{n} \sum_{r=1}^n \alpha_{[r]}^2 \right\}^{-1}. \quad (2.30)$$

Ou seja, a variância do estimador média amostral, na presença de erros de ordenação, quando a distribuição da variável de interesse é normal, é dada como a seguir

$$Var(\bar{X}) = \left\{ 1 - \frac{f(p)}{n} \sum_{r=1}^n \mu_{[r]}^2 \right\}^{-1} n / \sigma^2. \quad (2.31)$$

Tais resultados possibilitam a estimação da variância de maneira adequada, servindo como base para a obtenção de intervalos de confiança, sob a condição de que a variável de interesse tenha distribuição normal.

2.7 Modelo de ordenação por variável concomitante

Em alguns casos, a ordenação pode ser realizada mediante a mensuração de uma variável concomitante, fortemente correlacionada com a variável de interesse. A utilização de uma variável Y (por exemplo, alguma medida externa de pacientes ou cobaias) pode servir de base para a ordenação de uma variável X (por exemplo, alguma característica interna dos mesmos).

O modelo de erro de ordenação com variável concomitante foi primeiramente abordado por Stokes (1977), que verificou que o aumento da precisão do estimador depende de uma maior correlação entre as variáveis.

O processo de seleção da amostra se dá com a obtenção de n amostras de tamanho n , nas quais mede-se o valor da variável Y , que será responsável pela ordenação das amostras. Como para o modelo com julgamento pessoal, farão parte da amostra por conjuntos ordenados e serão mensuradas as unidades amostrais julgadas como tendo a menor medida na primeira amostra, a de segunda menor medida na segunda e assim por diante até a n -ésima amostra, para a qual deve-se medir a observação que apresentar maior valor para a variável de interesse, de acordo com o critério utilizado para ordenação.

Este procedimento, aqui também, pode ser replicado m vezes com o intuito de se obter estimativas mais precisas.

Obviamente, o benefício proveniente da utilização deste método de amostragem depende de quanta informação a variável Y produz sobre X . Se X e Y forem independentes, o estimador $\bar{\bar{X}}$ deve ser equivalente em precisão a \bar{X} . Por outro lado, quanto mais correlacionadas forem estas variáveis, mais a precisão relativa se aproximará daquela obtida com ordenação perfeita. Stokes (1977) assume que há uma relação linear entre X e Y , com coeficiente de correlação ρ_{xy} , produzindo o seguinte resultado:

$$RP = \left[1 - \frac{\rho_{xy}^2}{n} \sum_{r=1}^n \left\{ \frac{E(Y_{(r)}) - \mu_y}{\sigma_y} \right\}^2 \right]^{-1}. \quad (2.32)$$

O capítulo seguinte apresenta outros estimadores, pontuais e intervalares, para a média e diversos outros parâmetros e funções populacionais, obtidos mediante amostragem por conjuntos ordenados.

Capítulo 3

Estimadores para outros parâmetros e intervalos de confiança sob o delineamento de Amostragem por Conjuntos Ordenados

A amostragem por conjuntos ordenados propicia a estimação eficiente de diversos parâmetros e funções populacionais além da média. Neste capítulo, são apresentados e discutidos alguns estimadores paramétricos e não paramétricos para a variância (Seção 3.1), quantis (Seção 3.2) e função distribuição acumulada (Seção 3.3) via *RSS*. A Seção 3.4 descreve resultados de estudos anteriores relativos à determinação de intervalos de confiança para a média, os quantis, a variância e os parâmetros do modelo de regressão linear simples, para dados com distribuição normal. Procura-se ressaltar os bons resultados obtidos mediante utilização do bootstrap, em sua versão paramétrica, na determinação de tais estimativas.

3.1 Estimadores para a variância

Stokes (1980) propõe um estimador assintoticamente não viciado ($\hat{\sigma}^2$) para a variância populacional e comprova a maior eficiência deste estimador em relação ao estimador originário de uma amostra aleatória simples, mesmo na presença de erros de ordenação. Este ganho em precisão, no entanto, não é tão drástico quanto o observado na estimação da média.

$$\hat{\sigma}^2 = \frac{1}{mn-1} \sum_{i=1}^m \sum_{r=1}^n \left(X_{[r]i} - \bar{\bar{X}} \right)^2 \quad (3.1)$$

tem média e variância dadas respectivamente por

$$E(\hat{\sigma}^2) = \sigma_X^2 + \frac{\sum_{r=1}^n \tau_{(r)}^2}{n(mn-1)}, \quad (3.2)$$

e

$$\begin{aligned} Var(\hat{\sigma}^2) = & \frac{m}{(mn-1)^2} \left\{ \left(\frac{mn-1}{mn} \right)^2 \sum_{r=1}^n E(X_{(r)} - \mu_{(r)})^4 \right. \\ & + 4 \sum_{r=1}^n (\mu_{(r)} - \mu_X)^2 \sigma_{(r)}^2 + 4 \left(\frac{mn-1}{mn} \right) \sum_{r=1}^n (\mu_{(r)} - \mu_X) E(X_{(r)} - \mu_{(r)})^2 \\ & \left. + 4 \frac{m}{(mn)^2} \sum_{r=1}^n \sum_{r < s} \sigma_{(r)}^2 \sigma_{(s)}^2 + \frac{2(m-1) - (mn-1)^2}{(mn)^2} \sum_{r=1}^n [\sigma_{(r)}^2] \right\}, \quad (3.3) \end{aligned}$$

sendo $\tau_{(r)} = \mu_{(r)} - \mu_X$ e $\mu_{(r)}$ a média da r -ésima estatística de ordem definida em (2.6). e $\sigma_{(r)}^2$ a variância da r -ésima estatística de ordem, definida em (2.7).

Uma alternativa paramétrica aplicável na estimação da variância populacional é apresentada em Yu et al (1999). Este estimador, versão não viciada de $\hat{\sigma}^2$, é dado por

$$\hat{\sigma}_{MS}^2 = \frac{1}{\left[mn - 1 + \sum_{r=1}^n \frac{\alpha_r^2}{n} \right]} \sum_{i=1}^m \sum_{r=1}^n (X_{(r)i} - \bar{X})^2 \quad (3.4)$$

onde α_r representa a média da r -ésima estatística de ordem reduzida de uma variável com distribuição normal. A variância do estimador $\hat{\sigma}_{MS}^2$ é apresentada na seqüência

$$\begin{aligned} Var(\hat{\sigma}_{MS}^2) = & \frac{m\sigma^4}{\left(mn - 1 + \sum_{r=1}^n \frac{\alpha_r^2}{n} \right)} \left\{ \left(\frac{mn-1}{mn} \right)^2 \sum_{r=1}^n E(U_{(r)} - \alpha_r)^4 \right. \\ & + 4 \sum_{r=1}^n \alpha_r^2 E(U_{(r)} - \alpha_r)^2 + 4 \left(\frac{mn-1}{mn} \right) \sum_{r=1}^n \alpha_r E(U_{(r)} - \alpha_r)^2 \\ & + 4 \frac{m}{(mn)^2} \sum_{r=1}^n \sum_{r < s} E(U_{(r)} - \alpha_r)^2 E(U_{(s)} - \alpha_s)^2 \\ & \left. + \frac{2(m-1) - (mn-1)^2}{(mn)^2} \sum_{r=1}^n [E(U_{(r)} - \alpha_r)^2]^2 \right\}. \quad (3.5) \end{aligned}$$

3.2 Estimadores para os quantis

Baseado no estudo de propriedades assintóticas, Chen (1999) propõe um intervalo de confiança não paramétrico para os quantis populacionais, a partir do quantil estimado de uma amostra por conjuntos ordenados.

Para uma v.a. X com distribuição F , o quantil populacional de ordem p é definido da seguinte maneira

$$\xi_p = F^{-1}(p) = \inf\{x : F(x) \geq p\}. \quad (3.6)$$

Pode-se estimar ξ_p através do quantil amostral calculado a partir de uma amostra selecionada aleatoriamente ou mediante utilização de esquemas amostrais mais refinados. Considere uma amostra por conjuntos ordenados de tamanho $N = mn$, extraída de uma população com distribuição F e função densidade de probabilidades f . A função distribuição empírica via *RSS* é dada por

$$F^*(x) = \frac{1}{N} \sum_{i=1}^m \sum_{r=1}^n I_{(-\infty, x)}(X_{(r)i}) \quad (3.7)$$

sendo $N = mn$ o tamanho da amostra selecionada e $I_0(\cdot)$ a função indicadora. Para $0 < p < 1$, seja ξ_p o quantil populacional de ordem p e $\hat{\xi}_{N,p}$ o quantil estimado a partir de F^* da seguinte forma

$$\hat{\xi}_{N,p} = \inf\{x : F^*(x) \geq p\}. \quad (3.8)$$

Os resultados assintóticos apresentados na seqüência foram extraídos de Chen (1999). Primeiramente, pode-se verificar que o p -ésimo quantil amostral em conjuntos ordenados converge quase certamente para o p -ésimo quantil populacional, pois com probabilidade um

$$\left| \hat{\xi}_{N,p} - \xi_p \right| \leq \frac{2(\log N)^2}{f(\xi_p)N^{1/2}} \quad (3.9)$$

para valores de N suficientemente grandes. Outro resultado diz respeito à convergência em distribuição de $\hat{\xi}_{N,p}$. Segundo Chen (1999), se a função densidade f for positiva em uma vizinhança de ξ_p e contínua em ξ_p , considerando ordenação perfeita tem-se

$$\sqrt{N}(\hat{\xi}_{N,p} - \xi_p) \xrightarrow{D} N\left(0, \frac{\sigma_{n,p}^2}{f^2(\xi_p)}\right) \quad (3.10)$$

sendo $\sigma_{n,p}^2 = \frac{1}{n} \sum_{r=1}^n B(r, n+r-1, p)[1 - B(r, n+r-1, p)]$ e $B(r, s, x)$ a função distribuição acumulada da distribuição beta com parâmetros r e s . Está implícito neste resultado o fato de que $B(r, n+r-1) = F_{(r)}(\xi_p)$ não depende de F , provendo condições de se determinar intervalos de confiança para o quantil populacional.

Desta forma, baseado no resultado referente à convergência em distribuição de $\hat{\xi}_{N,p}$, um intervalo de confiança para o p -ésimo quantil é

$$\left[\hat{\xi}_{N,p} - \frac{z_\alpha \sigma_{n,p}}{\sqrt{N} f(\xi_p)}; \hat{\xi}_{N,p} + \frac{z_\alpha \sigma_{n,p}}{\sqrt{N} f(\xi_p)} \right] \quad (3.11)$$

sendo z_α o α -ésimo quantil de ϕ , a distribuição normal padrão. As expressões (3.10) e (3.11) continuam válidas quando f é estimada.

Para dados com distribuição normal, Barnett e Brown (2002) obtiveram o estimador linear não viciado ótimo por conjuntos ordenados (*ranked set sample best linear unbiased quantile estimator – RSBLUQE*) do p -ésimo quantil populacional, que é apresentado a seguir

$$\hat{Q}_p = \hat{\mu} + \zeta_{1-p} \hat{\sigma} = \frac{\sum_{r=1}^n (X_{(r)} / \nu_r)}{\sum_{r=1}^n (1 / \nu_r)} + \zeta_{1-p} \frac{\sum_{r=1}^n (\alpha_r X_{(r)} / \nu_r)}{\sum_{r=1}^n (\alpha_r^2 / \nu_r)} \quad (3.12)$$

sendo α_r e ν_r a média e a variância da r -ésima estatística de ordem padronizada, enquanto $\hat{\mu}$ e $\hat{\sigma}$ são versões simplificadas dos BLUEs μ^* e σ^* , citados em (2.14), podendo ser re-escritos, na condição de normalidade, por

$$\hat{\mu} = \frac{\sum_{r=1}^n (X_{(r)} / v_r)}{\sum_{r=1}^n (1 / v_r)} \quad (3.13)$$

$$\hat{\sigma} = \frac{\sum_{r=1}^n (\alpha_r X_{(r)} / v_r)}{\sum_{r=1}^n (\alpha_r^2 / v_r)}. \quad (3.14)$$

A variância do RSBLUQE é dada por

$$Var(\hat{Q}_p) = Var(\hat{\mu} + \zeta_{1-p} \hat{\sigma}) = \sigma^2 \left(\frac{1}{\sum_{r=1}^n (1 / v_r)} + \zeta_{1-p}^2 \frac{1}{\sum_{r=1}^n (\alpha_r^2 / v_r)} \right). \quad (3.15)$$

Uma extensão do RSBLUQE na presença de réplicas é dada por

$$\hat{Q}_{rep} = \frac{\sum_{r=1}^n (\overline{X_{(r)\cdot}} / v_r)}{\sum_{r=1}^n (1 / v_r)} + \zeta_{1-p} \frac{\sum_{r=1}^n (\alpha_r X_{(r)\cdot} / v_r)}{\sum_{r=1}^n (\alpha_r^2 / v_r)} \quad (3.16)$$

sendo $\overline{X_{(r)\cdot}} = \frac{\sum_{k=1}^m X_{(r)k}}{m}$ e m o número de replicações realizadas. A variância de \hat{Q}_{rep} é dada por

$$Var(\hat{Q}_{rep}) = \frac{\sigma^2}{r} \left[\frac{1}{\sum_{r=1}^n (1 / v_r)} + \zeta_{1-p}^2 \frac{1}{\sum_{r=1}^n (\alpha_r^2 / v_r)} \right]. \quad (3.17)$$

Barnett e Brown (2002) verificaram um relevante ganho em eficiência devido a utilização do RSBLUQE, ao invés do BLUQE (*best linear unbiased quantile estimator*) baseado em SRS e proposto em Hassanein et al (1986).

3.3 Estimador da função distribuição acumulada

Considere X_1, X_2, \dots, X_N uma amostra aleatória simples extraída de uma população com função distribuição F . Esta função na maior parte das vezes é desconhecida, devendo ser de alguma forma estimada.

A maneira mais usual de se estimar F é através da função distribuição empírica (\hat{F}), obtida a partir das unidades amostrais selecionadas. Esta função caracteriza-se por saltos de tamanho $1/N$, que ocorrem em cada um dos pontos amostrais. Pode-se representar a função distribuição empírica (f.d.e.) da seguinte maneira

$$\hat{F}(t) = \frac{1}{N} \sum_{j=1}^N I_{(-\infty, x_j)}(t), \quad (3.18)$$

onde $I_0(\cdot)$ é a função indicadora.

Um primeiro resultado importante relativo à f.d.e. diz respeito ao fato desta ser um estimador não viciado de F , o que pode ser verificado como a seguir

$$E(\hat{F}(t)) = E\left(\frac{1}{N} \sum_{j=1}^N I_{(-\infty, x_j)}(t)\right) = \frac{1}{N} E\left(\sum_{j=1}^N I_{(-\infty, x_j)}(t)\right). \quad (3.19)$$

Mas as variáveis aleatórias $I_{(0, x_1)}(t), I_{(0, x_2)}(t), \dots, I_{(0, x_N)}(t)$ têm distribuição de Bernoulli, com parâmetro $F(t)$. Logo,

$$\frac{1}{N} E\left(\sum_{j=1}^N I_{(-\infty, x_j)}(t)\right) = \frac{1}{N} \sum_{j=1}^N F_{X_j}(t) = \frac{1}{N} (NF(t)) = F(t). \quad (3.20)$$

Como X_1, X_2, \dots, X_N são independentes, $I_{(0, x_1)}(t), I_{(0, x_2)}(t), \dots, I_{(0, x_N)}(t)$ também o são. Logo, a variância de \hat{F} é dada por

$$Var(\hat{F}(t)) = Var\left(\frac{1}{n} \sum_{j=1}^n I_{(-\infty, x_j)}(t)\right) = \frac{1}{n^2} Var\left(\sum_{j=1}^n I_{(-\infty, x_j)}(t)\right) =$$

$$= \frac{1}{n^2} nF(t)[1-F(t)] = \frac{F(t)[1-F(t)]}{n}. \quad (3.21)$$

Considere agora uma amostra por conjuntos ordenados de tamanho $N = mn$, dada por $(X_{[1]1}, X_{[2]1}, \dots, X_{[n]1}, X_{[1]2}, X_{[2]2}, \dots, X_{[n]2}, X_{[1]m}, X_{[2]m}, \dots, X_{[n]m})$, originária de uma população com distribuição F . A função distribuição empírica, calculada a partir da amostra em conjuntos ordenados, é dada por

$$F^*(t) = \frac{1}{mn} \sum_{r=1}^n \sum_{i=1}^m I_{(-\infty, X_{[r]i}]}(t). \quad (3.22)$$

As observações $\{X_{[r]i}\}$ são independentes, mas identicamente distribuídas com distribuição $F_{[r]}$ apenas para algum r fixado. Logo, $I_{(0, X_{[r]1}]}(t), I_{(0, X_{[r]2}]}(t), \dots, I_{(0, X_{[r]m}]}(t)$ têm distribuição Bernoulli com parâmetro $F_{[r]}(t)$. Deste modo,

$$\begin{aligned} E(F^*(t)) &= E\left(\frac{1}{mn} \sum_{r=1}^n \sum_{i=1}^m I_{(-\infty, X_{[r]i}]}(t)\right) = \\ &= \frac{1}{mn} E\left(\sum_{r=1}^n \sum_{i=1}^m I_{(-\infty, X_{[r]i}]}(t)\right) = \frac{1}{n} \sum_{r=1}^n F_{[r]}(t). \end{aligned} \quad (3.23)$$

Dell e Clutter (1972) apresentam o seguinte resultado

$$\frac{1}{n} \sum_{r=1}^n f_{[r]} = f. \quad (3.24)$$

Utilizando o resultado 3.20, tem-se

$$E(F^*(t)) = \frac{1}{n} \sum_{r=1}^n F_{[r]}(t) = F(t). \quad (3.25)$$

A variância de F^* é calculada da seguinte maneira

$$\begin{aligned} \text{Var}(F^*(t)) &= \frac{1}{(mn)^2} \sum_{r=1}^n \sum_{i=1}^m \text{var}(I_{(-\infty, X_{[r]i}]}(t)) = \\ &= \frac{1}{m^2 n^2} n \sum_{r=1}^n F_{[r]}(t)[1-F_{[r]}(t)] = \frac{1}{mn^2} \sum_{r=1}^n F_{[r]}(t)[1-F_{[r]}(t)]. \end{aligned} \quad (3.26)$$

3.3.1 Caracterização de uma amostra por conjuntos ordenados

Seja $T = (T_1, T_2, \dots, T_n)$ um vetor aleatório multinomial com mn realizações e vetor de probabilidades $p = (1/n, 1/n, \dots, 1/n)$. Considere mn variáveis aleatórias, representadas por y_1, y_2, \dots, y_{mn} , obtidas observando T e selecionando T_r observações da população com f.d.p. $f_{[r]}$ $r = 1, 2, \dots, n$. Stokes e Sager (1988) mostram que

- I. Y_1, Y_2, \dots, Y_{mn} tem a mesma estrutura probabilística de X_1, X_2, \dots, X_{mn} , uma amostra aleatória simples de uma população com f.d.p f_X ;
- II. $\{Y_1, Y_2, \dots, Y_{mn} / T = (m, m, \dots, m)\}$ tem a mesma estrutura probabilística de $\{X_{[r]i}\}$, $r = 1, 2, \dots, n; i = 1, 2, \dots, m$, uma amostra por conjuntos ordenados da mesma população.

Dado que os estimadores \hat{F} e F^* são não viciados, pode-se compará-los simplesmente através de suas variâncias.

$$\begin{aligned} \text{Var}(\hat{F}(t)) &= E(\text{Var}(\hat{F}(t) / T)) + \text{Var}(E(\hat{F}(t) / T)) = \\ &= E\left(\frac{1}{(mn)^2} \sum_{r=1}^n T_r F_{[r]}(t) [1 - F_{[r]}(t)]\right) + \text{Var}\left(\frac{1}{mn} \sum_{r=1}^n T_r F_{[r]}(t)\right) = \\ &= \text{Var}[F^*(t)] + \text{Var}\left[\frac{1}{mn} \sum_{r=1}^n T_r F_{[r]}(t)\right]. \end{aligned} \quad (3.27)$$

O segundo termo corresponde ao ganho na estimação de F utilizando F^* ao invés de \hat{F} . Caso o critério utilizado para ordenação das amostras seja muito ruim (aleatório), $F_{[r]}(t) = F(t)$, produzindo

$$\frac{1}{mn} \sum_{r=1}^n T_r F_{[r]}(t) = F(t) \quad (3.28)$$

e, conseqüentemente

$$Var(\hat{F}(t)) = Var(F^*(t)) \quad (3.29)$$

3.3.2 Eficiência de F^*

A função distribuição empírica pode ser usada tanto na estimação pontual de $F(t)$, para algum valor de t , quanto para toda a distribuição populacional.

Um resultado importante apresentado por Stokes e Sager (1988) corresponde a uma extensão do Teorema Central do Limite para o estimador F^*

$$[F^* - E(F^*)]/[Var(F^*)]^{1/2} \xrightarrow{D} z \sim N(0,1). \quad (3.30)$$

Utilizando este resultado, pode-se construir intervalos de $100(1-\alpha)\%$ de confiança para $F(t)$, com limites

$$F^* \pm z_{\alpha/2} \sqrt{Var(F^*(t))}. \quad (3.31)$$

Um estimador não viciado de $Var(F^*(t))$ é dado em 3.32

$$\hat{Var}(F^*(t)) = \frac{1}{(m-1)n^2} \sum_{r=1}^n \hat{F}_{[r]}(t) [1 - \hat{F}_{[r]}], \quad (3.32)$$

onde $\hat{F}_{[r]}(t) = (1/m) \sum_{i=1}^m I[X_{[r]i} \leq t]$ é a função distribuição empírica baseada nas m estatísticas de ordem julgadas na r -ésima posição.

A precisão relativa (RP) de \hat{F} em relação a F^* é apresentada a seguir

$$\begin{aligned} RP(t) &= [Var(\hat{F}(t))]/[Var(F^*(t))] = \\ &= F(t)[1 - F(t)] / \left\{ F(t) - \sum_{r=1}^n [F_{[r]}(t)]^2 / n \right\}. \end{aligned} \quad (3.33)$$

O valor de RP pode ser estimado utilizando estimadores para as variâncias de \hat{F} e F^* , ou seja

$$RP^* = \hat{V}ar[\hat{F}(t)] / \hat{V}ar[F^*(t)] \quad (3.34)$$

sendo $\hat{V}ar(F^*(t))$ como definida em (3.32) e $\hat{V}ar(\hat{F}(t))$ dada por

$$\hat{V}ar[\hat{F}(t)] = \left\{ F^*(t) - F^{*2}(t) + \frac{1}{(m-1)n^2} \sum_{r=1}^n [F_{[r]}(t) - F_{[r]}^2(t)] \right\} / mn \quad (3.35)$$

que, de acordo com Stokes e Sager (1988), é um estimador não viciado de $Var[\hat{F}(t)]$.

Duas situações extremas podem ser consideradas para melhor interpretar a expressão (3.34), correspondente à precisão relativa do estimador F^* em relação ao estimador \hat{F} . Primeiramente, se o critério de ordenação for aleatório, pode-se concluir da expressão (3.33) que o valor de RP será igual a 1, uma vez que $F_{[r]}(t) = F(t)$, para $r = 1, 2, \dots, n$. Por outro lado, caso o critério de ordenação seja indefectível (ordenação perfeita), $F_{[r]}(t) = F_{(r)}(t)$, a função distribuição acumulada da r -ésima estatística de ordem, tornando possível representar RP da seguinte maneira:

$$RP(t) = \frac{F(t)[1-F(t)]}{F(t) - \sum_{r=1}^n [I_{F(t)}(r, n-r+1)]^2 / n} \quad (3.36)$$

sendo $[I_{F(t)}(r, n-r+1)]$ a função beta incompleta, igual à distribuição da r -ésima estatística de ordem.

Chen (1999) propõe um método de estimação da função densidade de probabilidades a partir da amostra por conjuntos ordenados, utilizando o método kernel, e conclui, a partir da comparação dos erros médios quadráticos integrados, que este estimador é mais eficiente que o correspondente estimador obtido via *SRS*.

3.4 Estimação intervalar via amostragem por conjuntos ordenados

A maior parte dos estudos citados nesta seção baseia-se na proposição de métodos paramétricos de reamostragem, no caso o bootstrap, visando determinar intervalos de confiança para diversos parâmetros. O Capítulo 4 apresenta de maneira detalhada o funcionamento do bootstrap, bem como sua aplicabilidade na obtenção de algumas estimativas intervalares.

O bootstrap é um procedimento de reamostragem que possibilita, dentre outras coisas, a estimação do erro padrão de diversos estimadores, permitindo a obtenção de intervalos de confiança e testes de hipóteses. Dada uma amostra original, o bootstrap, em sua versão paramétrica, consiste na geração de reamostras a partir de uma distribuição F , cuja forma é conhecida, mas com parâmetros estimados via amostra. As estimativas produzidas por cada reamostra fornecem condições de se construir diferentes tipos de intervalos de confiança.

A estimação paramétrica via RSS com aplicação na proposição de métodos de reamostragem bootstrap é abordada em estudos recentes. Cesário e Barreto (2003), Raminelli (2003) e Garcia (2004) utilizam amplamente esta técnica de reamostragem na proposição de estimadores intervalares para os parâmetros de interesse.

A determinação e avaliação de intervalos de confiança bootstrap para a média e para os parâmetros de regressão linear simples, considerando populações com distribuição normal, foi o objeto de estudo de Cesário e Barreto (2003). Comprovou-se, neste trabalho, a adequabilidade dos estimadores propostos, mediante análise de resultados obtidos via simulação. Na estimação da média populacional, baseado nas probabilidades de cobertura obtidas, obteve-se bons resultados, principalmente na condição de conhecimento do valor do parâmetro de escala (σ). Para os casos em que não se conhece este valor, os intervalos do tipo studentizado apresentaram melhor desempenho. Bons resultados foram alcançados também na estimação dos parâmetros de regressão, mas somente para números elevados de replicações da variável dependente (m superior a 7).

Raminelli (2003) apresenta algumas alternativas na obtenção de intervalos de confiança para quantis populacionais via RSS , para dados com distribuição normal. Considera, na estimação dos quantis, o intervalo assintótico proposto por Chen (1999) e apresentado em (3.11), além de diferentes intervalos de confiança bootstrap, baseados no estimador linear não viciado ótimo na presença de réplicas (3.16). Através de simulação, foi verificada boa adequabilidade dos métodos

bootstrap para diferentes quantis e tamanhos de amostra em relação aos intervalos assintóticos, sobretudo para amostras de tamanho reduzido. Ratificou, além disso, a conclusão obtida por Chen (1999), referente a um ganho mais expressivo, em termos de eficiência, na estimação de quantis centrais.

Estudo análogo foi realizado por Garcia (2004), visando analisar a adequabilidade da utilização do bootstrap paramétrico na construção de estimativas intervalares para a variância populacional. Foram considerados, dentre outros, o estimador não paramétrico proposto por Stokes (1980) e dado em (3.1), além de sua versão modificada, $\hat{\sigma}_{MS}^2$, apresentado em (3.4). Foi verificado, via simulação, um melhor desempenho das estimativas baseadas em $\hat{\sigma}_{MS}^2$, devido a um menor vício e conseqüentemente menor erro quadrático médio, se comparado aos demais estimadores. Constatou, além disso, uma boa performance dos intervalos propostos, verificando probabilidades de cobertura estimadas bem próximas aos níveis de confiança pré-estabelecidos, principalmente para os intervalos do tipo studentizado, como será detalhado no próximo capítulo.

Taconeli e Barreto (2003), por sua vez, verificaram um bom desempenho de intervalos de confiança assintóticos para variáveis com distribuição normal, quando a ordenação das amostras é perfeita ou apresenta pequenas falhas. Na presença de erros de ordenação com maiores magnitudes, detectaram uma subestimação da variância do estimador média amostral. Para variáveis com distribuição exponencial verificaram que a distribuição da média amostral é gama, e atestam a inviabilidade dos intervalos propostos nesta situação. Estudo semelhante foi realizado por Brandão (2003), considerando variáveis aleatórias com distribuição uniforme, lognormal e Gumbell, em que se diagnosticou a simetria da distribuição do estimador \bar{X} para diversos tamanhos de amostras, além da adequabilidade dos intervalos assintóticos normais na estimação da média de uma distribuição uniforme via RSS.

O capítulo seguinte apresenta uma revisão bibliográfica do método de reamostragem bootstrap, detalhando o procedimento em suas duas versões (paramétrica e não paramétrica), além de expor diversos intervalos de confiança originados por esta metodologia, caracterizando-os de acordo com suas propriedades assintóticas.

Capítulo 4

Intervalos de confiança Bootstrap sob Amostragem Aleatória Simples

Desenvolvido por Efron (1979), o bootstrap é um método amplamente utilizado na obtenção de estimativas pontuais e intervalares, bem como na avaliação da acurácia de estimativas e testes. Consiste, basicamente, na replicação do processo de estimação via reamostragem, através da amostra ou da sua distribuição, com parâmetros estimados via amostra. Efron e Tibshirani (1993) e Davison e Hinkley (1997) apresentam de maneira detalhada várias aplicações dos métodos bootstrap em diversas metodologias estatísticas.

A Seção 4.1 descreve o funcionamento do bootstrap e a maneira como este procedimento pode ser utilizado na estimação de um parâmetro populacional qualquer. Já as Seções 4.2 e 4.3 apresentam suas versões paramétrica e não paramétrica, respectivamente. Alguns intervalos de confiança obtidos via bootstrap são encontrados na Seção 4.4, acompanhados de um breve estudo relativo a acurácia de estimativas intervalares, necessário para um bom entendimento das propriedades assintóticas dos intervalos apresentados.

4.1 O bootstrap

Seja X_1, X_2, \dots, X_n uma amostra aleatória simples de tamanho n com função distribuição de probabilidade F . Suponha que se esteja interessado em inferir sobre uma quantidade populacional $\theta = t(F)$, e seja $\hat{\theta}$ um estimador desta quantidade. Segundo Efron e Tibshirani (1993), com o procedimento de reamostragem bootstrap tem-se condições de fazer inferência

sobre a distribuição de $\hat{\theta} - \theta$. Isto é possível a partir da determinação de uma estimativa da distribuição F , por exemplo \tilde{F} , de modo que

$$t(\tilde{F}) = \hat{\theta}. \quad (4.1)$$

De acordo com Canty (2000), usando reamostragem bootstrap, pode-se obter réplicas de $\hat{\theta}$, indicadas por $\hat{\theta}^*$, tais que

$$(\hat{\theta} - \theta) \stackrel{D}{\approx} (\hat{\theta}^* - \hat{\theta}), \quad (4.2)$$

possibilitando a estimação de propriedades de $\hat{\theta}$, como seu vício e variância, da seguinte maneira

$$b(\hat{\theta}) = E(\hat{\theta} - \theta / X \sim F) \approx E(\hat{\theta}^* - \hat{\theta} / X^* \sim \tilde{F}) \quad (4.3)$$

$$\begin{aligned} Var(\hat{\theta}) &= Var(\hat{\theta} - \theta / X \sim F) \\ &\approx Var(\hat{\theta}^* - \hat{\theta} / X^* \sim \tilde{F}). \end{aligned} \quad (4.4)$$

Ressaltando o fato de que a distribuição \tilde{F} é conhecida, não há maiores dificuldades na determinação da distribuição de $(\hat{\theta}^* - \hat{\theta} / X^* \sim \tilde{F})$, ratificando a conveniência da aproximação apresentada em (4.2). Esta aproximação pode se dar de maneira analítica, o que nem sempre é possível, ou através de métodos computacionalmente intensivos, como Monte Carlo. O processo de simulação de amostras bootstrap pode ser paramétrico, caso se conheça a distribuição de X , ou não paramétrico, nas situações em que não se tenha tal conhecimento.

4.2 Bootstrap paramétrico

O processo de reamostragem bootstrap paramétrico baseia-se em modelos probabilísticos com parâmetros estimados via amostra original. Suponha F uma distribuição indexada por um parâmetro ψ , com estimador $\hat{\psi}$, baseado na amostra original X_1, X_2, \dots, X_n , e considere $\hat{\theta} = t(F_{\hat{\psi}})$. Deste modo, pode-se considerar $F_{\hat{\psi}}$ como estimador de F_{ψ} . Sejam $X_1^*, X_2^*, \dots, X_B^*$ amostras da distribuição $F_{\hat{\psi}}$ e $\hat{\psi}_1^*, \hat{\psi}_2^*, \dots, \hat{\psi}_B^*$ estimativas de $\hat{\psi}$ para estas amostras. Assim,

obtem-se B estimativas para o parâmetro de interesse $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*)$. O bootstrap paramétrico consiste, finalmente, na aproximação da distribuição $t(F_{\psi^*}) - t(F_{\psi})$ por $t(F_{\hat{\psi}^*}) - t(F_{\hat{\psi}})$.

4.3 Bootstrap não paramétrico

O bootstrap não paramétrico é usado quando não se tem conhecimento da distribuição F . Neste caso, deve-se gerar B amostras com reposição e de mesmo tamanho da amostra original, a partir da distribuição \hat{F} , que corresponde à distribuição empírica dos dados:

$$\hat{F}(x) = \frac{\#(x_j \leq x)}{n}, \quad (4.5)$$

onde $\#(\cdot)$ representa o número de vezes em que a condição é verdadeira. $\hat{F}(x)$ assume os valores $(1/n, 2/n, \dots, 1)$ e também é conhecida como frequência relativa acumulada.

Como no caso paramétrico, o procedimento consiste na obtenção de amostras $x_j^* = (x_{j1}^*, x_{j2}^*, \dots, x_{jn}^*)$, $j = 1, 2, \dots, B$, sendo x_j^* obtida via reamostragem dos dados originais, responsáveis pela produção das B estimativas bootstrap $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*)$. Ao possibilitar a extração de conclusões baseadas exclusivamente na amostra, o bootstrap não paramétrico é responsável pela produção de resultados confiáveis em situações em que suposições relativas à distribuição populacional são questionáveis.

Davison e Hinkley (1997) apresentam alguns resultados para o caso em que o estimador de interesse é a média de uma *SRS*. Aqui, o asterisco presente em E^* e Var^* indica que tais quantidades são calculadas de acordo com a distribuição ajustada, \hat{F}

$$E^*(\bar{X}^*) = E^*\left(\frac{1}{n} \sum_{j=1}^n x_j^*\right) = \bar{X} \quad (4.6)$$

e similarmente

$$\begin{aligned}\text{var}^*(\bar{X}^*) &= \frac{1}{n} \text{var}^*(X^*) = \frac{1}{n} E^*\{X^* - E^*(X^*)\}^2 = \\ &= \frac{(n-1)}{n} \frac{1}{n(n-1)} \sum_{j=1}^n (x_j - \bar{x})^2\end{aligned}\quad (4.7)$$

que pode ser escrito na forma

$$\text{var}^*(\bar{X}^*) = \frac{(n-1)s^2}{n^2}.\quad (4.8)$$

4.4 Intervalos de confiança bootstrap

A simples utilização de uma estimativa pontual $\hat{\theta}$ no processo de inferência de um determinado parâmetro θ mostra-se, na maioria das vezes, imprecisa, uma vez que não leva em consideração informações referentes à precisão do estimador e ao erro decorrente da estimação. A construção de intervalos de confiança leva em consideração tais valores, propiciando a obtenção de estimativas mais confiáveis.

Baseado na aproximação (4.2), torna-se possível a obtenção de estimativas intervalares do parâmetro que se deseja inferir. Isto pode ocorrer sob diversas abordagens, produzindo diferentes intervalos, cujos desempenhos são verificados no presente trabalho.

Duas características primordiais na análise do desempenho de intervalos de confiança são a precisão (referente à variabilidade das estimativas) e a acurácia, ou seja, a proximidade da probabilidade de cobertura de tais intervalos com o nível de confiança almejado. Efron e Tibshirani (1993) apresentam algumas classificações para intervalos de confiança, baseadas na acurácia dos mesmos, que são apresentadas na Seção 4.4.1.

4.4.1- Acurácia dos intervalos de confiança

Considere um intervalo de confiança unilateral, com limite superior igual a $\hat{\theta}_\alpha$, tal que

$$P(\theta \leq \hat{\theta}_\alpha) \approx \alpha. \quad (4.9)$$

Segundo Efron e Tibshirani (1993), $\hat{\theta}_\alpha$ é denominado acurado de primeira ordem se

$$P(\theta \leq \hat{\theta}_\alpha) = \alpha + O(n^{-1/2}) \quad (4.10)$$

sendo n o tamanho da amostra. Ou seja, $P(\theta \leq \hat{\theta}_\alpha)$ irá diferir de α por um termo cuja magnitude é, no máximo, de $n^{-1/2}$, para valores de n suficientemente grandes. Por outro lado, $\hat{\theta}_\alpha$ é dito acurado de segunda ordem se

$$P(\theta \leq \hat{\theta}_\alpha) = \alpha + O(n^{-1}) \quad (4.11)$$

com interpretação análoga à anterior, ou seja, $P(\theta \leq \hat{\theta}_\alpha)$ irá diferir de α por um termo cuja magnitude é, no máximo, de n^{-1} , para valores de n suficientemente grandes.

Davison e Hinkley (1997) apresentam com detalhes a determinação da acurácia dos limites de confiança dos intervalos de confiança bootstrap. Alguns destes resultados são apresentados ao longo das seções seguintes, acompanhados dos respectivos intervalos.

4.4.2 Intervalo bootstrap normal

Considere $\hat{\theta}$ um estimador de um parâmetro θ . Uma primeira alternativa consiste na aproximação da distribuição de $(\hat{\theta} - \theta)$ por uma normal, de tal forma que

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{Var(\hat{\theta})}} \sim N(0,1) \quad (4.12)$$

produzindo o seguinte intervalo bilateral com $(1-2\alpha)\%$ de confiança

$$\left(\hat{\theta} - z_{1-\alpha} \sqrt{\text{Var}(\hat{\theta})}, \hat{\theta} - z_{\alpha} \sqrt{\text{Var}(\hat{\theta})} \right), \quad (4.13)$$

sendo z_{α} e $z_{1-\alpha}$ quantis da distribuição normal padrão. O bootstrap pode ser aplicado, neste caso, no cálculo da variância do estimador ($\text{Var}_B(\hat{\theta})$), bem como na obtenção de seu vício estimado ($b_B(\hat{\theta})$), produzindo um fator de correção para a estimativa pontual. Tais quantidades são obtidas mediante a geração de B amostras bootstrap, e o cálculo das quantidades requeridas se dá da seguinte forma

$$\text{Var}_B(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2 \quad (4.14)$$

$$b_B(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}), \quad (4.15)$$

onde $\hat{\theta}_b^*$ refere-se à estimativa obtida através da b -ésima reamostra e $\bar{\hat{\theta}}^*$ à média das B estimativas

$$\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*. \quad (4.16)$$

Finalmente, um intervalo de confiança usando a aproximação normal, com vício e variância calculados via bootstrap, é dado da seguinte maneira

$$\left(\hat{\theta} - b_B(\hat{\theta}) - z_{1-\alpha} \sqrt{\text{Var}_B(\hat{\theta})}, \hat{\theta} - b_B(\hat{\theta}) - z_{\alpha} \sqrt{\text{Var}_B(\hat{\theta})} \right). \quad (4.17)$$

Martinez e Louzada-Neto (2000) apresentam este intervalo como intervalo bootstrap padrão.

4.4.3 Intervalo bootstrap básico

Suponha que a distribuição de $\hat{\theta} - \theta$ seja conhecida. Neste caso, é possível determinar um valor α que satisfaça

$$\begin{aligned}
P(\hat{\theta} - \theta < a_\alpha) &= \alpha \\
P(\theta < \hat{\theta} - a_\alpha) &= 1 - \alpha.
\end{aligned}
\tag{4.18}$$

Através de (4.18), torna-se possível a obtenção do seguinte limite de confiança $(1-\alpha)$

$$\hat{\theta}_{1-\alpha} = \hat{\theta} - a_\alpha \tag{4.19}$$

e, analogamente, um intervalo de confiança bilateral $(1-2\alpha)$ para θ tem a seguinte forma

$$(\hat{\theta} - a_{1-\alpha}, \hat{\theta} - a_\alpha). \tag{4.20}$$

Os quantis da distribuição $\hat{\theta} - \theta$ na maior parte das vezes não são conhecidos. Esta dificuldade pode ser relevada ao se utilizar a aproximação proposta em (4.2) e o bootstrap na obtenção da distribuição $(\hat{\theta}^* - \hat{\theta})$. Seria possível, finalmente, determinar as quantidades a_α e $a_{1-\alpha}$, através deste procedimento.

Considere G a função distribuição de $\hat{\theta} - \theta$, e G^* a fdp de $\hat{\theta}^* - \hat{\theta}$. Temos, então

$$G(a_\alpha) = \alpha \Rightarrow a_\alpha = G^{-1}(\alpha). \tag{4.21}$$

Usando bootstrap, através da distribuição aproximada, podemos determinar os quantis

$$a_{B;\alpha} = (G^*)^{-1}(\alpha) \tag{4.22}$$

após a obtenção de B estimativas $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*)$ provenientes de B reamostras geradas via bootstrap, $(x_1^*, x_2^*, \dots, x_B^*)$. A fda utilizada em (4.21) é calculada empiricamente a partir das estimativas ordenadas $(\hat{\theta}_{(1)}^* - \hat{\theta}, \hat{\theta}_{(2)}^* - \hat{\theta}, \dots, \hat{\theta}_{(B)}^* - \hat{\theta})$, ou seja, G^* é aproximada por

$$\hat{G}^*(u) = \frac{1}{B} \sum_{b=1}^B I\{\hat{\theta}_b^* - \hat{\theta} \leq u\}. \tag{4.23}$$

A partir da função distribuição acumulada G^* , o quantil a_α da distribuição de $\hat{\theta} - \theta$ é estimado pelo quantil $a_{B;\alpha}$, dado por

$$\hat{a}_{B;\alpha} = \hat{\theta}_{((B+1)\alpha)}^* - \hat{\theta}. \tag{4.24}$$

Davison e Hinkley (1997) apresentam o seguinte intervalo de confiança para θ a partir dos quantis estimados em (4.24) e baseado no intervalo proposto em (4.20):

$$\begin{aligned}
& \left(\hat{\theta} - a_{B;1-\alpha}, \hat{\theta} - a_{B;\alpha} \right) = \\
& = \left(\hat{\theta} - \left(\hat{\theta}_{(B+1)(1-\alpha)}^* - \hat{\theta} \right), \hat{\theta} - \left(\hat{\theta}_{(B+1)\alpha}^* - \hat{\theta} \right) \right) = \\
& = \left(2\hat{\theta} - \hat{\theta}_{(B+1)(1-\alpha)}^*, 2\hat{\theta} - \hat{\theta}_{(B+1)\alpha}^* \right). \tag{4.25}
\end{aligned}$$

O intervalo de confiança bootstrap básico é acurado de primeira ordem (Davison e Hinkley, 1997).

O número de replicações a serem realizadas deve ser elevado, a fim de que os resultados sejam precisos. Além disso, deve-se determinar um valor de B tal que os valores $(B+1)\alpha$ e $(B+1)(1-\alpha)$ sejam inteiros. Caso contrário, será necessário o uso de interpolação, considerando a escala de probabilidade normal, uma vez que a interpolação linear não se ajusta bem neste caso (maiores detalhes, ver Davison e Hinkley (1997)).

4.4.4 Intervalo bootstrap studentizado

A aproximação apresentada em (4.12) tem bons resultados para valores grandes de n . Quando dispomos de amostras de tamanhos reduzidos, uma melhor aproximação é dada por

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\widehat{Var}(\hat{\theta})}} \sim t_{n-1}. \tag{4.26}$$

Deste modo, seria possível a obtenção de um intervalo de confiança para θ , baseado na distribuição t-Student como

$$\left(\hat{\theta} - z_{(1-\alpha)} \sqrt{\widehat{Var}(\hat{\theta})}, \hat{\theta} - z_{(\alpha)} \sqrt{\widehat{Var}(\hat{\theta})} \right) \tag{4.27}$$

sendo $z_{(1-\alpha)}$ e $z_{(\alpha)}$ quantis de uma distribuição t-Student com $n-1$ graus de liberdade.

Uma alternativa consiste em obter a distribuição da variável studentizada Z diretamente dos dados. Para isso, utiliza-se o método t-bootstrap, calculando a estimativa $\hat{\theta}_b^*$, proveniente da reamostra x_b^* e a quantidade

$$Z_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{\sqrt{Var(\hat{\theta}_b^*)}} \quad (4.28)$$

para cada uma das B reamostras.

O calculo da estatística Z_b^* requer o conhecimento de $Var(\hat{\theta}_b^*)$. Quando não se dispõe de tal quantidade, uma maneira alternativa para se calcular o valor de $Var(\hat{\theta}_b^*)$ é através de duplo bootstrap, ou seja, tomando um número M de reamostras da b -ésima reamostra bootstrap e calculando a variância das M estimativas obtidas, da maneira como descrito em 4.15, apenas substituindo B por M . O elevado número requerido de reamostragens neste processo representa um custo computacional bastante grande.

Tendo em mãos os valores $z_1^*, z_2^*, \dots, z_B^*$, o α -ésimo quantil da distribuição de Z^* deve ser estimado por $\hat{z}_{B;\alpha} = z_{((B+1)\alpha)}^*$ (Davison e Hinkley, 1997). Finalmente, o intervalo de confiança bootstrap studentizado é

$$\left(\hat{\theta} - z_{(B+1)(1-\alpha)}^* \sqrt{Var_B(\hat{\theta})}, \hat{\theta} - z_{(B+1)\alpha}^* \sqrt{Var_B(\hat{\theta})} \right). \quad (4.29)$$

O intervalo de confiança studentizado ou t-bootstrap é acurado de segunda ordem, ao contrário dos métodos apresentados nas seções anteriores (Davison e Hinkley, 1997).

4.4.5 Intervalo bootstrap percentil

Tendo gerado um número B de reamostras bootstrap $(x_1^*, x_2^*, \dots, x_B^*)$, que originaram B estimativas para o parâmetro de interesse $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*)$, um intervalo de confiança percentil $(1-2\alpha)$ será determinado pelos quantis α e $1-\alpha$ de \hat{G}^* , a fda de $\hat{\theta}^*$ (Efron e Tibshirani, 1993). Pode-se escrever este intervalo da seguinte forma:

$$\left(\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)\right). \quad (4.30)$$

Sabendo que $\hat{G}^{-1}(\alpha)$ refere-se ao α -ésimo quantil da distribuição bootstrap, o intervalo apresentado em (4.30) pode ser re-escrito como

$$\left(\hat{\theta}_{\alpha}^*, \hat{\theta}_{1-\alpha}^*\right). \quad (4.31)$$

As expressões 4.30 e 4.31, segundo Efron e Tibshirani (1993) referem-se à situação ideal, em que o número de replicações bootstrap é infinita. Na prática, utiliza-se um número finito B de reamostras bootstrap $(x_1^*, x_2^*, \dots, x_B^*)$ e, conseqüentemente, de B estimativas bootstrap $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*)$. Os percentis da distribuição de $\hat{\theta}^*$ possibilitam, finalmente, a construção de intervalos de confiança do tipo bootstrap percentil, como o apresentado em 4.32:

$$\left(\hat{\theta}_{(\alpha)}^{*B}, \hat{\theta}_{(1-\alpha)}^{*B}\right). \quad (4.32)$$

O intervalo bootstrap percentil é destacado devido à simplicidade do método e propriedade de invariância a transformações monótonas (Martinez e Louzada Neto, 2000). Este método, assim como os métodos bootstrap normal e básico, é acurado de primeira ordem.

4.4.6 O método BCa

O método BCa (forma abreviada para “*bias corrected and accelerated*”) constitui uma versão aperfeiçoada do método percentil, apresentado em 4.4.5. Os percentis utilizados na delimitação dos intervalos de confiança dependem de duas quantidades \hat{a} e \hat{z}_0 denominados, respectivamente, como correção para tendência e aceleração.

De acordo com Efron e Tibshirani (1993), um intervalo de confiança BCa $(1-2\alpha)$ é dado por

$$\left(\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}\right) \quad (4.33)$$

onde

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right) \quad (4.34)$$

sendo que $\Phi(\cdot)$ refere-se à função distribuição acumulada de uma normal padrão e $z^{(\alpha)}$ ao percentil 100α desta distribuição. O intervalo de confiança percentil pode ser considerado um caso particular do intervalo BCa, em que \hat{a} e \hat{z}_0 são ambos iguais a zero.

O valor \hat{z}_0 mede a discrepância entre a mediana de $\hat{\theta}^*$ e $\hat{\theta}$, em unidades normais. Seu valor é computado como a seguir

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\hat{\theta}_b^* < \hat{\theta})}{B}\right) \quad (4.35)$$

sendo $\Phi^{-1}(\cdot)$ a inversa da distribuição acumulada normal padrão.

A interpretação do valor \hat{a} , segundo Canty (2000), não é simples. Uma estimativa não paramétrica para esta quantidade é apresentada na seqüência

$$\hat{a} = \frac{\sum l_j^3}{6\left(\sum l_j^2\right)^{3/2}} \quad (4.36)$$

sendo l uma função conveniente denominada função influência. Martinez e Louzada-Neto (2000) apresentam um método de obtenção do valor \hat{a} via jackknife.

Elaborado por volta de 1950 por Maurice Quenouille e aprimorado posteriormente por Tukey, ambos citados em Martinez e Louzada Neto (2000), o jackknife é uma alternativa não paramétrica para a obtenção de estimativas de erros padrão. Consiste, basicamente, na extração de todas as possíveis amostras de tamanho $n-1$ de uma amostra de tamanho n , gerando $n-1$ reamostras, sendo a i -ésima reamostra composta pelas observações originais, com exceção da i -ésima ($i = 1, 2, \dots, n$). As reamostras geradas podem ser denotadas da seguinte maneira:

$$\begin{aligned}
x_{(1)} &= (x_2, x_3, \dots, x_n) \\
x_{(2)} &= (x_1, x_3, \dots, x_n) \\
&\vdots \\
x_{(n)} &= (x_1, x_2, \dots, x_{n-1}).
\end{aligned}
\tag{4.37}$$

Para cada uma das amostras geradas, calcula-se a estimativa de interesse $\hat{\theta}_{(i)}, i = 1, 2, \dots, n$.

Considere $\hat{\theta}$ a estimativa de θ , obtida a partir dos dados originais. Sendo

$$\hat{\theta}_{(\cdot)} = \frac{\sum_{i=1}^n \hat{\theta}_{(i)}}{n},
\tag{4.38}$$

baseado no procedimento descrito, pode-se obter o valor de \hat{a} como apresentado

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6 \left[\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2 \right]^{3/2}}.
\tag{4.39}$$

O método BCa, assim como o método t-bootstrap, é acurado de segunda ordem (Davison e Hinkley, 1997).

Outro método de obtenção de intervalos de confiança bootstrap, conhecido como ABC (*approximate bootstrap confidence interval*) corresponde a uma aproximação analítica por expansão de Taylor da segunda reamostragem no método BC_α (Davison e Hinkley, 1997). O Capítulo 5 apresenta os esquemas originais de reamostragem bootstrap via RSS, sob ordenação perfeita e imperfeita, e os conseqüentes intervalos de confiança, além do planejamento do estudo realizado via simulação.

Capítulo 5

Propostas originais de reamostragem bootstrap em Amostragem por Conjuntos Ordenados

No Capítulo 3 foram apresentados alguns aspectos positivos referentes à utilização da função distribuição empírica de uma amostra por conjuntos ordenados (F^*) em detrimento a \hat{F} . Verificou-se o fato de F^* ser um estimador não viciado, com variância inferior à do estimador \hat{F} , possibilitando que a função distribuição F seja estimada de maneira mais eficiente.

Tendo em vista os benefícios decorrentes da adoção da amostragem em conjuntos ordenados no processo de estimação de diversos parâmetros, propõe-se neste trabalho estender de maneira original a metodologia bootstrap selecionando amostras e gerando reamostras de acordo com o processo de conjuntos ordenados.

A Seção 5.1 propõe o algoritmo para obtenção de reamostras bootstrap quando o delineamento é por conjuntos ordenados perfeitamente e sua utilização na obtenção de intervalos de confiança. A Seção 5.2 contém um segundo algoritmo para gerar as reamostras quando a ordenação apresenta falhas. Na Seção 5.3 aborda-se a origem e algumas características dos dados fornecidos pela CETESB (Companhia de Tecnologia de Saneamento Ambiental) e utilizados com o objetivo de verificar algumas propriedades das estimativas propostas considerando uma população real. Uma descrição detalhada do estudo computacional realizado está presente na Seção 5.4, enquanto a Seção 5.5 se atém aos critérios utilizados na determinação do número de reamostragens e intervalos de confiança simulados para cada configuração. Tais critérios são essencialmente fundamentados em análises de convergência.

5.1 Proposta original de reamostragem bootstrap em conjuntos ordenados perfeitamente e extensões da estimação intervalar para a média populacional

A presente proposta de reamostragem bootstrap para conjuntos ordenados tem por princípio retirar, com reposição, amostras por conjuntos ordenados com reposição da amostra por conjuntos ordenados original.

Algoritmo 1

Seja $\{X_{[r]i} : r = 1, \dots, n; i = 1, \dots, m\}$ uma amostra por conjuntos ordenados e \bar{X} o estimador da média populacional produzido por esta amostra. O algoritmo aqui proposto de obtenção de reamostras bootstrap em conjuntos ordenados perfeitamente é descrito pelos seguintes passos:

1. Extração de B amostras $\{X_{[r]i}^* : r = 1, \dots, n; i = 1, \dots, m; j = 1, \dots, B\}$ com reposição, também via amostragem por conjuntos ordenados, da amostra original;
2. Cálculo de $\bar{X}_j^*, j = 1, \dots, B$ as respectivas médias amostrais.

Baseado neste procedimento de reamostragem bootstrap original, são propostos os seguintes intervalos de confiança bootstrap para μ_X :

a) Intervalo bootstrap normal

$$\left(\bar{X} - b_B(\bar{X}) - z_{1-\alpha} \sqrt{Var_B(\bar{X})}, \bar{X} - b_B(\bar{X}) - z_{\alpha} \sqrt{Var_B(\bar{X})} \right) \quad (5.1)$$

sendo $b_B(\bar{X}) = \frac{1}{B} \sum_{b=1}^B (\bar{X}_b^* - \bar{X})$ o vício do estimador, enquanto $Var_B(\bar{X})$ é a variância do estimador \bar{X} , estimada através das amostras replicadas da seguinte maneira

$$Var_B(\bar{X}) = \frac{1}{B-1} \sum_{b=1}^B (\bar{X}_b^* - \bar{\bar{X}}^*)^2 \quad (5.2)$$

sendo $\bar{\bar{X}}^* = \frac{1}{B} \sum_{k=1}^B \bar{X}_k^*$. O fator de correção para o vício não se faz necessário, uma vez que, como visto anteriormente, o estimador \bar{X} é não viciado para a média populacional. Sua utilização, portanto, poderia ser descartada sem maiores prejuízos.

b) Intervalo bootstrap básico

$$\left(2\bar{X} - \left(\bar{X}_{(B+1)(1-\alpha)}^* \right), 2\bar{X} - \left(\bar{X}_{(B+1)\alpha}^* \right) \right) \quad (5.3)$$

sendo $\bar{X}_{((B+1)(1-\alpha))}^*$ e $\bar{X}_{((B+1)\alpha)}^*$ obtidos empiricamente a partir das B estimativas \bar{x}_j^* calculadas em cada reamostra.

c) Intervalo bootstrap studentizado

$$\left(\bar{X} - z_{((B+1)(1-\alpha))}^* \sqrt{Var_B(\bar{X})}, \bar{X} - z_{((B+1)\alpha)}^* \sqrt{Var_B(\bar{X})} \right) \quad (5.4)$$

sendo $Var_B(\bar{X})$ calculado como em (5.2) e $z_{((B+1)(1-\alpha))}^*$ e $z_{((B+1)\alpha)}^*$ os quantis $(1-\alpha)$ e α da distribuição empírica obtida a partir dos valores $z_1^*, z_2^*, \dots, z_B^*$, calculados em cada uma das reamostras como

$$Z_b^* = \frac{\bar{X}_b^* - \bar{X}}{\sqrt{Var(\bar{X}_b^*)}}, b = 1, \dots, B \quad (5.5)$$

e $Var(\bar{X}_b^*)$ estimada através de um duplo bootstrap por

$$Var_B(\bar{X}_b^*) = \frac{1}{R-1} \sum_{k=1}^R (\bar{X}_{b,k}^{**} - \bar{\bar{X}}_b^{**}) \quad (5.6)$$

com $\bar{\bar{X}}_b^{**} = \frac{1}{R} \sum_{k=1}^R \bar{X}_{b,k}^{**}$.

d) Intervalo bootstrap percentil

$$\left(\bar{\bar{X}}_{(B+1)\alpha}^* , \bar{\bar{X}}_{(B+1)(1-\alpha)}^* \right) \quad (5.7)$$

onde $\bar{\bar{X}}_{(B+1)(1-\alpha)}^*$ e $\bar{\bar{X}}_{(B+1)\alpha}^*$ são os quantis $(1-\alpha)$ e α da distribuição empírica de $\bar{\bar{X}}^*$, calculada por meio das B estimativas obtidas.

e) Intervalo bootstrap BCa

$$\left(\bar{\bar{X}}^{*(\alpha_1)} , \bar{\bar{X}}^{*(\alpha_2)} \right) \quad (5.8)$$

onde

$$\begin{aligned} \alpha_1 &= \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})} \right) \\ \alpha_2 &= \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})} \right) \end{aligned} \quad (5.9)$$

sendo

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\#(\bar{X}_b^* < \bar{\bar{X}})}{B} \right) \quad (5.10)$$

e

$$\hat{a} = \frac{\sum l_j^3}{6(\sum l_j^2)^{3/2}} \quad (5.11)$$

obtido via jackknife, analogamente como descrito no Capítulo 4, da seguinte forma

$$\hat{a} = \frac{\sum_{i=1}^n \left(\bar{\bar{X}}_{(\cdot)} - \bar{\bar{X}}_{(i)} \right)^3}{6 \left[\sum_{i=1}^n \left(\bar{\bar{X}}_{(\cdot)} - \bar{\bar{X}}_{(i)} \right)^2 \right]^{3/2}}. \quad (5.12)$$

5.2 Proposta original de reamostragem bootstrap em conjuntos ordenados com ordenação imperfeita

O critério adotado para ordenação das unidades amostrais, como dito anteriormente, pode apresentar falhas, resultando em uma menor precisão dos estimadores dos parâmetros de interesse. Logicamente, esta perda de precisão tornar-se-á mais acentuada à medida que a ordenação seja realizada de maneira mais imprecisa.

Visando analisar o impacto da presença de erros de ordenação na estimação da média populacional, bem como verificar a adequabilidade dos intervalos de confiança propostos, segue um esquema para a incorporação dos erros no processo de simulação com amostras por conjuntos ordenados, mediante ordenação por julgamento pessoal.

Algoritmo 2

1. Geração de uma amostra aleatória simples da população de interesse: X_1, X_2, \dots, X_n ;
2. Geração de erros aleatórios e_1, e_2, \dots, e_n segundo uma $N(0, \sigma_\epsilon^2)$;
3. Criação da variável aleatória Y , originária da soma das variáveis criadas nos passos anteriores: $Y_1 = X_1 + e_1, Y_2 = X_2 + e_2, \dots, Y_n = X_n + e_n$;
4. Cálculo do *rank* de Y : r_1, r_2, \dots, r_n , onde r_j são números entre 1 e n que não se repetem;
5. Definição de $X_{[1]}$ como o valor gerado no primeiro passo cuja ordem no passo 4 é 1;
6. Os passos 1 a 5 são repetidos mais $n-1$ vezes, só que, no passo 5 define-se $X_{[r]}$ como o valor gerado no primeiro passo cujo rank obtido no passo 4 é r , $r = 2, \dots, n$;

7. Os passos 1 a 6 são repetidos m vezes. A amostra de conjuntos ordenados com m replicações de cada posição i é representada por $\{X_{[r]i} : r = 1, 2, \dots, n; i = 1, 2, \dots, m\}$, cujo termo geral, $X_{[r]i}$ é uma forma abreviada de $X_{[r]ri}$.

Conforme detalhado anteriormente, o bootstrap consiste na extração de reamostras por meio de uma função *plug-in* que se assemelhe à função distribuição da amostra original. Neste caso, deve-se procurar uma alternativa que torne possível a propagação dos erros de ordenação na geração das reamostras como da seleção da amostra, de forma que os estimadores produzidos sejam adequados. Neste contexto, propõe-se aqui o seguinte procedimento de re-amostragem para esta situação.

Algoritmo 3

1. Seleção de uma amostra por conjuntos ordenados $\{X_{[r]i}, r = 1, 2, \dots, n; i = 1, 2, \dots, m\}$ conforme o Algoritmo 2;
2. Ordenação da amostra obtida no passo 1 através dos valores da variável Y : Y_1, Y_2, \dots, Y_{mn} ;
3. Atribuição de *ranks* de 1 a $m*n$ às unidades amostrais selecionadas no passo 1, de acordo com a ordenação estabelecida no passo 2: r_1, r_2, \dots, r_{mn} ;
4. Geração de B reamostras $X_1^*, X_2^*, \dots, X_B^*$, segundo procedimento de amostragem por conjuntos ordenados, utilizando como critério para ordenação das u.a.s os *ranks* determinados no passo 3.

Repare que ao ordenar as amostras bootstrap de acordo com a variável aleatória Y , tem-se a garantia de que os erros de ordenação influenciam diretamente na ordem das reamostras, evitando, desta maneira, problemas referentes à acurácia dos estimadores considerados..

Na prática, esta ordenação poderia se realizar após a seleção da amostra utilizando o mesmo critério de ordenação utilizado durante o procedimento de amostragem. Deste modo, o pesquisador ordenaria a amostra de tamanho mn não de acordo com os valores mensurados da variável de interesse, que a esta altura já estão acessíveis, mas conforme seu próprio julgamento, de modo a conservar os erros inerentes do processo de ordenação.

Vale ressaltar que a utilização de grandes amostras poderia acarretar maiores dificuldades quando da ordenação da amostra final, uma vez que seria necessária a ordenação de mn unidades amostrais. As propostas apresentadas neste trabalho, no entanto, tem por objetivo proporcionar estimativas eficientes frente à impossibilidade de mensuração de um elevado número de elementos, não levando em conta, neste momento, possíveis dificuldades e custos relativos à ordenação das amostras.

Os programas computacionais que implementam os algoritmos relacionados são apresentados nos Apêndices A, B e C. Toda a programação foi implementada utilizando o Software R (Ihaka e Gentleman, 1996), versão 1.8.

5.3 Aplicação dos intervalos propostos em dados reais

Criada em 24 de julho de 1968, a CETESB - Companhia de Tecnologia de Saneamento Ambiental é a agência do Governo do Estado de São Paulo responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição, visando preservar e recuperar a qualidade das águas, do ar e do solo.

Um dos serviços realizados pela CETESB é a coleta e divulgação de informações referentes à qualidade do ar. Para isso, conta desde 1996 com 29 estações, responsáveis pela medição dos índices de diferentes tipos de impurezas atmosféricas, distribuídas em diferentes pontos das cidades de São Paulo, Cubatão, Campinas, São José dos Campos, Sorocaba e Paulínia. Dispõe ainda de estações móveis, que possibilitam o monitoramento da qualidade do ar em um número maior de localidades, de maneira menos sistemática. Os índices de interesse são coletados a cada hora. Os resultados são disponibilizados no site www.cetesb.sp.gov.br.

Os índices considerados pela CETESB seguem o padrão estabelecido pelos EUA e Canadá, determinados a partir de um longo período de experiência. As substâncias contempladas pela estrutura do índice são as seguintes:

- ✓ Dióxido de enxofre;
- ✓ Partículas totais em suspensão;
- ✓ Material particulado inalável (MP₁₀);
- ✓ Fumaça;
- ✓ Monóxido de carbono;
- ✓ Ozônio;
- ✓ Dióxido de nitrogênio.

A média diária, para cada uma das quantidades verificadas, é calculada a partir das 24 medidas tomadas entre as 16 horas do dia anterior até as 15 horas do dia de referência.

Uma das características da qualidade do ar se refere à quantidade de material particulado maior que 10 micra (MP₁₀). Constituem este material partículas sólidas ou líquidas que ficam suspensas no ar, na forma de poeira, neblina, aerossol, fumaça ou fuligem, com tamanho inferior a 10 micra, originados, principalmente, a partir de processos de combustão (industrial, automobilística) e aerossol secundário (formação na atmosfera). Altos níveis de MP₁₀ são responsáveis por um aumento no número de atendimentos hospitalares (causados, sobretudo, por problemas respiratórios), além de maior incidência de morte prematura (Revista Meio Ambiente Industrial, 2000).

O padrão de qualidade adotado pela CETESB, referente à concentração atmosférica de MP₁₀, é apresentado na Tabela 1.

Tabela 1- Padrão de qualidade do ar quanto à concentração atmosférica de MP₁₀

Faixa de Concentração	Qualidade	Índice	
0 - 50 µg /m ³	Boa	0-50	Atende
51 - 150 µg /m ³	Regular	51 - 100	ao padrão
151 - 250 µg /m ³	Inadequada	101 - 198	Não
251 - 420 µg /m ³	Má	200 - 299	atende
421 - 500 µg /m ³	Péssima	300 - 398	ao
> 500 µg /m ³	Critica	> 400	padrão

De acordo com resolução da CETESB, não mais do que 35 médias diárias ao ano podem ultrapassar o limite de $50 \mu\text{g}/\text{m}^3$.

O objetivo desta aplicação reside na construção e análise do desempenho de intervalos de confiança bootstrap para a média de MP_{10} registrada na estação Ibirapuera no biênio 2000/2001, via amostragem por conjuntos ordenados, como os propostos no Capítulo 4. Sabe-se de antemão, de acordo com os dados fornecidos pela CETESB, que a quantidade média de MP_{10} resultante de um total de 722 medições realizadas no período e local considerados foi $42.26 \mu\text{g}/\text{m}^3$.

5.4 Procedimento para análise da adequabilidade dos intervalos propostos e planejamento do estudo via simulação

Os intervalos de confiança propostos na Seção 5.1 foram avaliados de acordo com a acurácia e a precisão apresentadas. Para isso, foram simulados K intervalos de 90, 95 e 99% de confiança para cada uma das possíveis configurações, resultantes das combinações de diferentes tamanhos de amostras e números de replicações, distribuição da população de origem da amostra e grau de imperfeição na ordenação das unidades amostrais, para os quais considerou-se:

- ✓ Tamanhos de amostra e números de replicação: $m = 2$ e $n = 4$, $m = 3$ e $n = 3$, $m = 3$ e $n = 5$, $m = 5$ e $n = 3$. Estes valores foram escolhidos de modo a se ter a possível influência do número de réplicas com um tamanho final fixado e entre 3 e 20 (Barnett, 1999);
- ✓ Distribuições: dados gerados a partir de distribuições da família locação-escala (normal, exponencial, uniforme, lognormal e Gumbell), com parâmetro de locação igual a zero e parâmetro de escala igual a um, além das médias diárias de MP_{10} (Seção 5.3);
- ✓ Graus de imperfeição: $\sigma_\varepsilon^2 = 0$ (ordenação perfeita), $\sigma_\varepsilon^2 = 0,05$ (pequena incorreção) e $\sigma_\varepsilon^2 = 0,5$ (grau de imperfeição mais elevado), valores análogos aos utilizados em Dell e Cuttler(1972), no estudo da precisão do estimador $\bar{\bar{X}}$.

Os graus de imperfeição utilizados para os dados reais são diferentes, devido à discrepância entre as grandezas destas observações em relação aos valores gerados a partir das

distribuições teóricas. Buscou-se, entretanto, manter uma correspondência com a influência dos valores de σ_ϵ^2 adotados para as demais distribuições. O valor de σ_ϵ^2 responsável por uma maior imperfeição, neste caso, foi considerado $\sigma_\epsilon^2 = \sigma_X^2 / 2 = 238,59$, enquanto $\sigma_\epsilon^2 = \sigma_X^2 / 20 = 23,85$ foi escolhido para a situação em que os erros de ordenação têm menor interferência.

A partir dos K intervalos de confiança bootstrap simulados de cada uma das configurações consideradas, tomou-se como medida de acurácia a probabilidade de cobertura constatada, ou seja, a proporção dos de intervalos de confiança simulados que abrangem o valor do parâmetro de interesse (a média populacional). Sejam $\hat{\theta}_{Ik}$ e $\hat{\theta}_{Sk}$ os limites (inferior e superior) do k -ésimo intervalo de confiança gerado, $k = 1, 2, \dots, K$. Então, a probabilidade de cobertura corresponde a

$$PC = \frac{\sum_{k=1}^K I_{(\hat{\theta}_{Ik}, \hat{\theta}_{Sk})}(\mu)}{K}. \quad (5.13)$$

Logo, um intervalo bem acurado apresenta probabilidade de cobertura próxima ao nível de confiança fixado.

Como medida de precisão, utilizou-se o comprimento médio dos intervalos, além do desvio padrão dos comprimentos. Obviamente, intervalos precisos são compactos, com pequeno comprimento médio e baixa variabilidade quanto a sua amplitude. Seja $d_k = \hat{\theta}_{Sk} - \hat{\theta}_{Ik}$. Então,

$$CM = \frac{1}{K} \sum_{k=1}^K d_k \quad (5.14)$$

$$DPC = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (d_k - \bar{d})^2}. \quad (5.15)$$

O número de reamostragens a serem realizadas (B , ou R para o duplo bootstrap), bem como o total de intervalos de confiança simulados (K), foram determinados com base nos resultados obtidos em uma detalhada análise de convergência. Alguns destes resultados são apresentados na seqüência, acompanhados dos valores escolhidos para K , B e R .

5.5 Determinação do número de reamostragens e intervalos de confiança simulados via análise de convergência

Alguns dos intervalos propostos, como o studentizado e o normal, requerem a estimação de $Var_B(\bar{X})$, como descrito na Seção 5.1. Outros baseiam-se apenas na distribuição empírica do estimador, como os intervalos percentil, básico e BCa, delimitados por seus percentis ou funções dos mesmos. A qualidade de todas estas estimativas está diretamente relacionada ao número B de reamostras utilizado na construção de cada intervalo. Valores pequenos de B podem proporcionar estimativas ruins da variância do estimador, bem como dos seus percentis, originando, conseqüentemente, intervalos de confiança pouco precisos. Por outro lado, o uso de um número elevado de reamostras torna-se inviável, à medida que demanda um grande esforço computacional.

Com o objetivo de determinar os números adequados de reamostras, que possibilite a obtenção de estimativas precisas com o mínimo esforço computacional, realizou-se preliminarmente um estudo da convergência dos estimadores envolvidos na obtenção dos limites dos intervalos de confiança. O procedimento adotado consiste no cálculo das quantidades de interesse a cada nova reamostragem realizada. Ou seja, considerando um total de B reamostras, estimativas para os percentis e para a variância do estimador devem ser tomadas a cada reamostra gerada, ou seja, para $B = 2, 3, \dots, 999, 1000$. Com isto, têm-se condições de determinar a partir de qual valor de B tais estimativas se estabilizem, fazendo deste valor o número ideal de reamostras a serem geradas.

Esta proposta de análise de convergência foi realizada para os intervalos construídos a partir de dados reais, fornecidos pela CETESB, referentes aos resultados das medições diárias de MP_{10} na estação Ibirapuera em São Paulo, SP (Seção 5.3), considerando $m = 2$ e $n = 4$. Esta escolha se deve ao fato da distribuição dos dados ser assimétrica, aliado ao menor tamanho de amostra considerado, constituindo a situação mais “crítica” analisada. Resultados considerados adequados para esta configuração, relativos ao número necessário de reamostras são estendidos para todas as demais configurações.

Com o propósito de desfazer qualquer distorção causada por alguma determinada amostra, a análise da convergência foi verificada, quando possível, a partir de cinco amostras distintas. A

Figura 2 apresenta a análise da convergência de $(Var_B(\bar{\bar{X}}))$. Pode-se observar uma boa convergência mesmo para valores de B próximos a 200. A fim de garantir resultados precisos, não se utilizou neste trabalho menos do que 200 reamostras com o intuito de se estimar a variância do estimador média amostral. Foram simuladas 1000 reamostras para a estimação de $Var_B(\bar{\bar{X}})$ quando na obtenção dos intervalos bootstrap normais e 200 reamostras para os dois ciclos bootstrap necessários da obtenção de intervalos bootstrap do tipo studentizado.

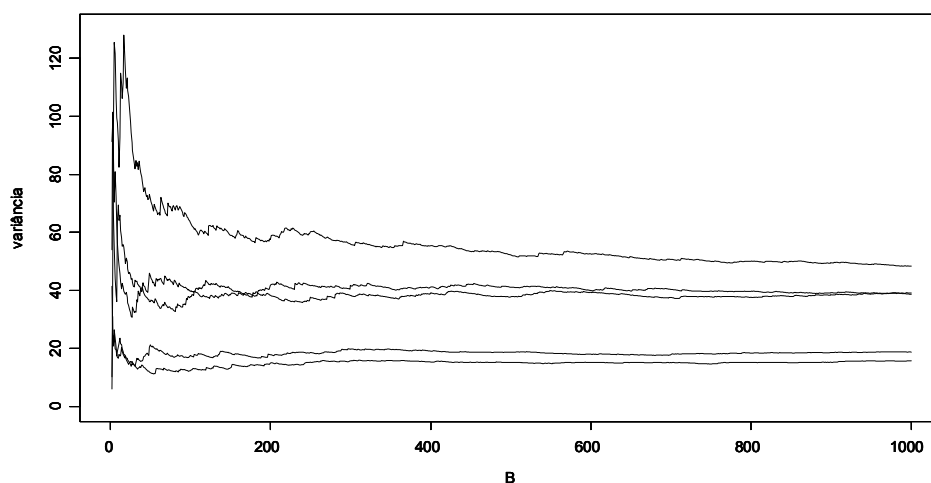


Figura 2- Valores de $Var_B(\bar{\bar{X}})$ calculados via bootstrap a partir de B reamostras, originárias de cinco amostras distintas extraídas das medições médias diárias de MP_{10} .

Estudo análogo foi realizado para os intervalos baseados em percentis. Mais uma vez, analisou-se a convergência, agora dos quantis ou de funções deles responsáveis pelos limites dos respectivos intervalos, a partir de cinco amostras distintas. A Figura 3 apresenta a análise da convergência para os limites inferiores e superiores de intervalos bootstrap do tipo percentil com 90% de confiança. Pode-se observar maior instabilidade nas estimativas para valores de B inferiores a 200, e uma boa convergência para valores de B superiores a 400. Estudo similar foi realizado para intervalos com 95 e 99% de confiança, assim como para intervalos de confiança bootstrap dos tipos básico e BCa. Em todos os casos verificou-se que um total de 400 reamostras seria adequado, do ponto de vista da convergência dos resultados alcançados. Assim, devido à grande semelhança dos resultados alcançados, foram suprimidas as demais figuras, buscando

compactar e dinamizar este trabalho No estudo via simulação, utilizou-se 1000 reamostras na determinação de intervalos de confiança bootstrap percentil, básico e BCa.

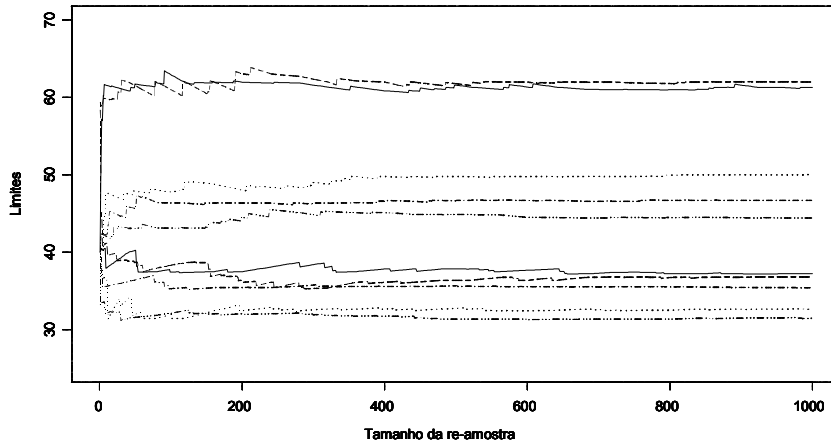


Figura 3- Valores de $\bar{\bar{X}}_{(0,05)}$ e $\bar{\bar{X}}_{(0,95)}$ para diferentes valores de B , calculados via bootstrap, a partir de cinco amostras distintas extraídas das medições médias diárias de MP_{10} . (Diferentes tracejados indicam diferentes amostras originais).

Deve-se verificar também um número adequado de intervalos de confiança a serem gerados, de modo a estimar de maneira precisa a probabilidade de cobertura dos intervalos propostos. Para isso, utilizando a mesma configuração (população e tamanho de amostra) utilizada na análise da convergência anterior, procedeu-se com a análise da convergência da probabilidade de cobertura. Foi utilizado como limitante um total de 1000 intervalos, visando não alongar por demais o tempo gasto nas simulações. Para todos os intervalos, fixou-se $B=1000$, exceto para o intervalo studentizado, para o qual utilizou-se 200 reamostras em cada um dos dois bootstraps requeridos na obtenção dos limites de confiança.

Mais uma vez, os resultados verificados foram bastante similares para todos os tipos de intervalos de confiança bootstrap considerados. Por praticidade, é apresentada apenas a curva de convergência referente aos 1000 intervalos de confiança bootstrap percentil (Figura 4). Verificou-se, em todos os casos, que a convergência é alcançada para 400 intervalos de confiança, número utilizado para intervalos bootstrap do tipo studentizado. Para os demais, devido ao menor tempo de simulação, adotou-se um total de 1000 intervalos de confiança.

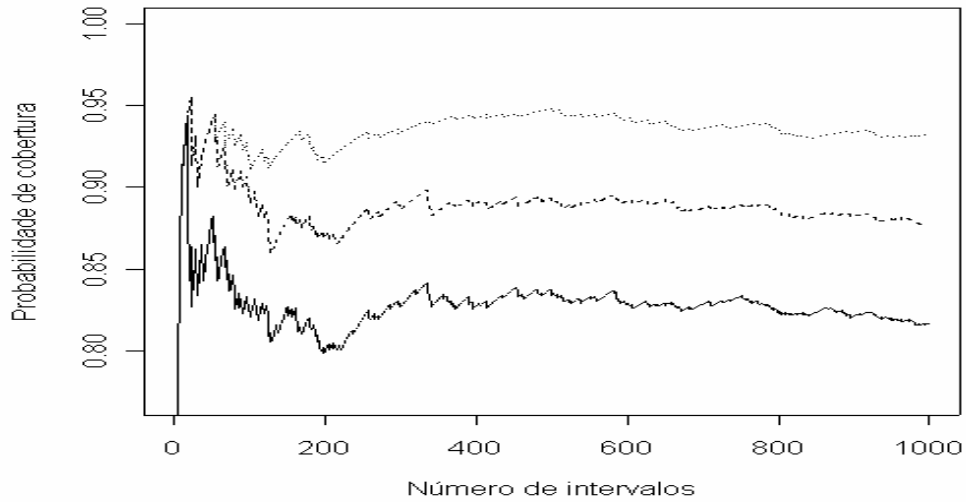


Figura 4- Probabilidade de cobertura calculada para diferentes números de intervalos de confiança bootstrap do tipo percentil de 90 (linha contínua), 95 (tracejado mais forte) e 99% de confiança (tracejado mais fraco) para a média das medições médias diárias de MP_{10} .

Tendo definido o número de reamostragens e o total de intervalos de confiança a serem simulados, pôde-se, finalmente, analisar o desempenho dos intervalos de confiança bootstrap não paramétricos via *RSS* através de um amplo estudo via simulação. Os principais resultados alcançados são apresentados no Capítulo 6, através de figuras e tabelas que trazem uma série de informações relativas ao estudo realizado.

Capítulo 6

Estudo via simulação das propriedades dos intervalos de confiança originários dos esquemas de reamostragem propostos

Neste capítulo são apresentados os resultados do estudo realizado via simulação, através de representações gráficas da probabilidade de cobertura e do comprimento médio verificados para os intervalos gerados sob cada uma das configurações consideradas. As Seções 6.1 a 6.6 contêm as análises originárias das seis distribuições estudadas. Cada figura e tabela retrata o desempenho dos intervalos propostos na estimação da média da população considerada naquela seção. Foram simulados também intervalos de confiança baseados em amostragem aleatória simples, permitindo uma comparação direta com a performance dos intervalos propostos sob ordenação perfeita e imperfeita.

Algumas observações são necessárias antes da exposição dos resultados. Com o objetivo de não tornar maçante a descrição dos resultados, construiu-se gráficos apenas para os resultados provenientes dos intervalos bootstrap percentil e studentizado, uma vez que o desempenho dos demais intervalos apresenta uma elevada similaridade quanto ao observado para os intervalos do tipo percentil. Os resultados, no entanto, podem ser observados na íntegra nas tabelas apresentadas em cada seção. Estas tabelas contêm também o desvio padrão dos comprimentos dos intervalos simulados, outra medida relativa à precisão das estimativas geradas.

Nas tabelas de probabilidades de cobertura, valores em negrito indicam que os intervalos gerados a partir de *RSS* apresentam melhor acurácia em relação aos intervalos bootstrap obtidos via *SRS*, para a configuração escolhida de tamanho de amostra, número de réplicas e grau de imperfeição dos erros. Valores acompanhados de um asterisco indicam que, para essa configuração, não se pode rejeitar a hipótese da probabilidade de cobertura ser a inicialmente

desejada (90, 95 ou 99%) a um nível de 95% de confiança. Tal hipótese é verificada através de um teste para proporções, baseado na distribuição binomial com parâmetros n e p , sendo n o número de intervalos de confiança simulados e p o nível de confiança considerado.

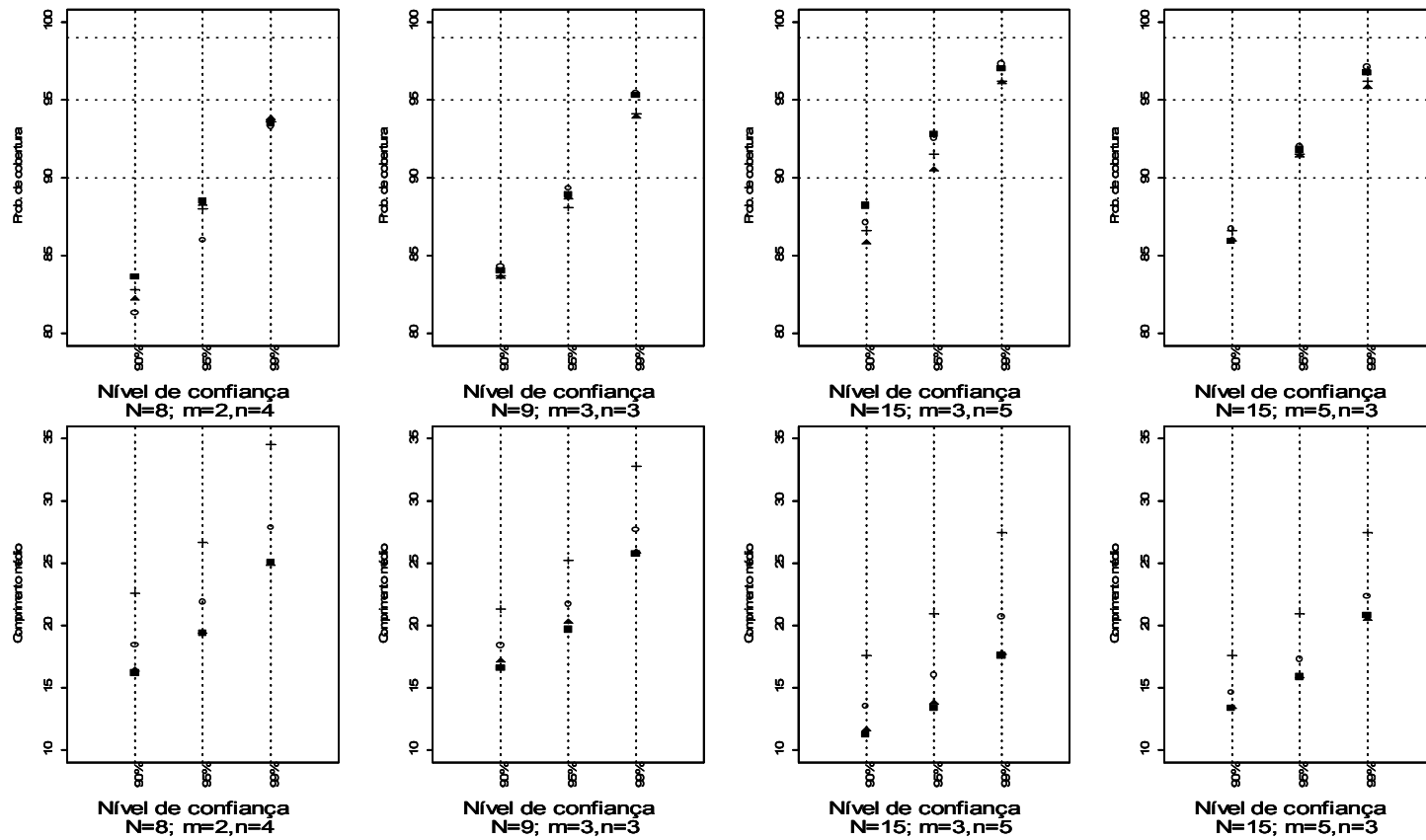
As figuras contrapõem resultados obtidos via *SRS* com amostras de tamanho N e via *RSS* com amostras de tamanho n e m réplicas, de tal forma que $N=mn$.

6.1 Estimação da média das medições médias diárias de MP10 na estação Ibirapuera, nos anos de 2000 e 2001

As Figuras 5 e 6 e as Tabelas 2 e 3 contêm algumas características, obtidas através de simulação, referentes ao desempenho dos intervalos de confiança bootstrap na estimação intervalar da média das verificações médias diárias de MP_{10} , no local e período anteriormente especificados. A performance dos intervalos de confiança gerados através da metodologia bootstrap original e a proposta neste trabalho foram comparadas de acordo com a acurácia e a precisão das estimativas intervalares produzidas pelos dois métodos.

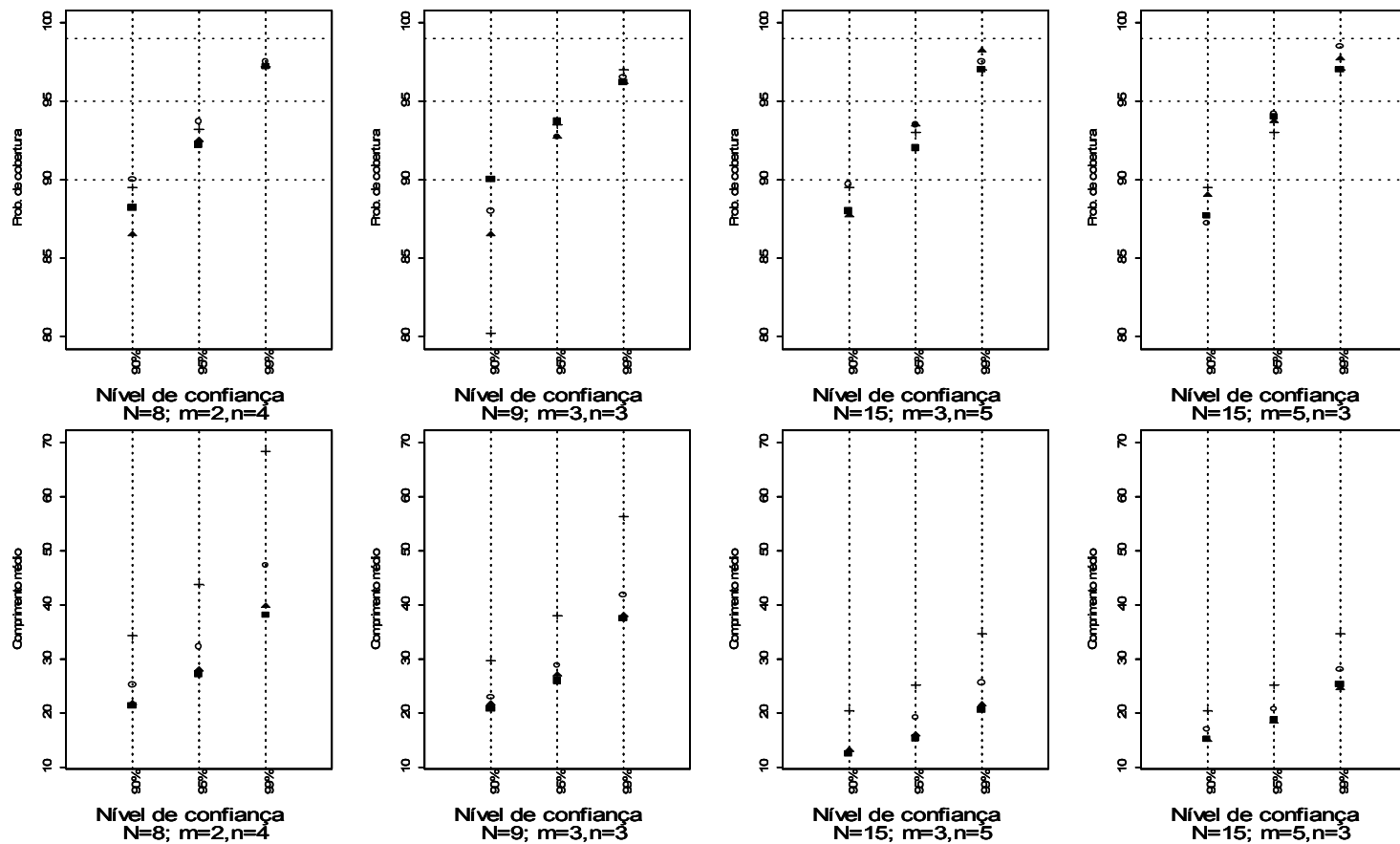
Pode-se reparar, através da Tabela 2, que na maior parte das configurações consideradas houve uma discreta melhora quanto a acurácia dos intervalos gerados via *RSS* em relação àqueles originados via *SRS*. Os intervalos do tipo studentizado confirmaram ser os mais acurados, sendo capazes de produzir probabilidades de cobertura mais próximas aos níveis de confiança pré-estabelecidos. Os demais intervalos ficaram aquém ao desejado, em relação às probabilidades de cobertura obtidas, sob ambos delineamentos amostrais.

Verifica-se também, através da Tabela 3, uma substancial melhora quanto à precisão das estimativas obtidas através do esquema de reamostragem proposto em relação às estimativas convencionais, baseadas em amostragem aleatória simples. Observe que os comprimentos médios dos intervalos de confiança via *RSS* com $mn=8$ pouco diferem daqueles observados via *SRS* com amostras de tamanho $N=15$, principalmente quando o critério utilizado para ordenação das amostras é eficiente. Este fato indica que o procedimento de reamostragem proposto neste trabalho viabiliza o uso de uma amostra bem mais reduzida, sem grande prejuízo na precisão das estimativas geradas.



(+) SRS (■) RSS (Ordenação perfeita) (▲) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.05$)
 (○) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 5- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média das verificações médias diárias de MP10 na estação Ibirapuera (2000/2001), via SRS e RSS com ordenação perfeita e imperfeita.



(+) SRS (■) RSS (Ordenação perfeita) (▲) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.05$)

(○) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 6- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média das verificações médias diárias de MP10 na estação Ibirapuera (2000/2001), via SRS e RSS com ordenação perfeita e imperfeita.

Tabela 2- Probabilidade de cobertura de intervalos de confiança bootstrap para a média das medições médias diárias de MP₁₀, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita.

		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS
Normal	90%	82,3	82,6	81,3	80,7	84,2	84,0	84,0	80,8
	95%	87,6	88,3	86,2	87,0	89,2	88,7	89,3	86,9
	99%	93,8	94,4	93,4	93,7	94,8	94,2	95,4	94,2
Básico	90%	83,0	82,8	81,7	81,4	83,7	83,9	84,0	83,4
	95%	88,0	87,9	86,1	86,1	89,1	88,2	88,9	88,1
	99%	94,7	93,3	92,1	92,7	95,0	93,5	94,3	93,7
Percentil	90%	83,6	82,2	81,3	82,8	84,0	83,6	84,3	83,7
	95%	88,5	88,3	86,0	88,0	88,9	88,7	89,3	88,1
	99%	93,5	93,8	93,3	93,6	95,3	93,9	95,4	94,1
BCa	90%	83,5	83,1	83,0	82,2	84,7	84,8	81,8	81,9
	95%	88,8	88,5	88,2	87,9	89,9	89,8	88,8	87,2
	99%	94,3	94,7	93,7	93,2	93,8	95,4	94,7	92,3
Studentizado	90%	88,2*	86,5	90,0*	89,5*	90,0*	86,5	88,0*	90,2*
	95%	92,2	92,5	93,7*	93,2*	93,7*	92,7	92,7*	93,5*
	99%	97,2	97,2	97,5	97,2	96,2	96,2	96,5	97,0

Tabela 2- (cont) Probabilidade de cobertura de intervalos de confiança bootstrap para a média das medições médias diárias de MP₁₀, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita.

	Confiança	N=15; m=3,n=5				N=15; m=5,n=3			
		RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS
Normal	90%	85,6	85,3	86,7	84,6	85,8	85,4	86,5	84,6
	95%	90,3	90,9	92,2	89,5	91,5	91,3	91,9	89,5
	99%	96,5	96,5	97,6	95,3	96,2	95,6	97,2	95,3
Básico	90%	85,4	84,9	86,4	85,1	87,2	85,6	86,6	85,1
	95%	91,9	90,3	91,3	89,4	91,4	90,7	91,4	89,4
	99%	96,1	95,5	96,9	94,3	96,0	95,1	96,6	94,3
Percentil	90%	88,2	85,8	87,1	86,6	85,9	86,0	86,7	86,6
	95%	92,8	90,5	92,5	91,5	91,8	91,4	92,0	91,5
	99%	97,0	96,1	97,3	96,2	96,7	95,8	97,1	96,2
BCa	90%	86,3	88,1	87,1	86,0	85,0	85,3	85,7	86,0
	95%	91,3	92,9	91,3	91,3	89,7	91,1	91,4	91,3
	99%	96,8	97,0	97,3	96,8	94,4	96,6	96,7	96,8
Studentizado	90%	88,0*	87,7*	89,7*	89,5*	87,7*	89,0*	87,2*	89,5*
	95%	92,0	93,5*	93,5*	93,0*	94,0*	93,7*	94,2*	93,0*
	99%	97,0	98,2*	97,5	97,0	97,0	97,7	98,5*	97,0

Tabela 3- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média das medições médias diárias de MP₁₀, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita.

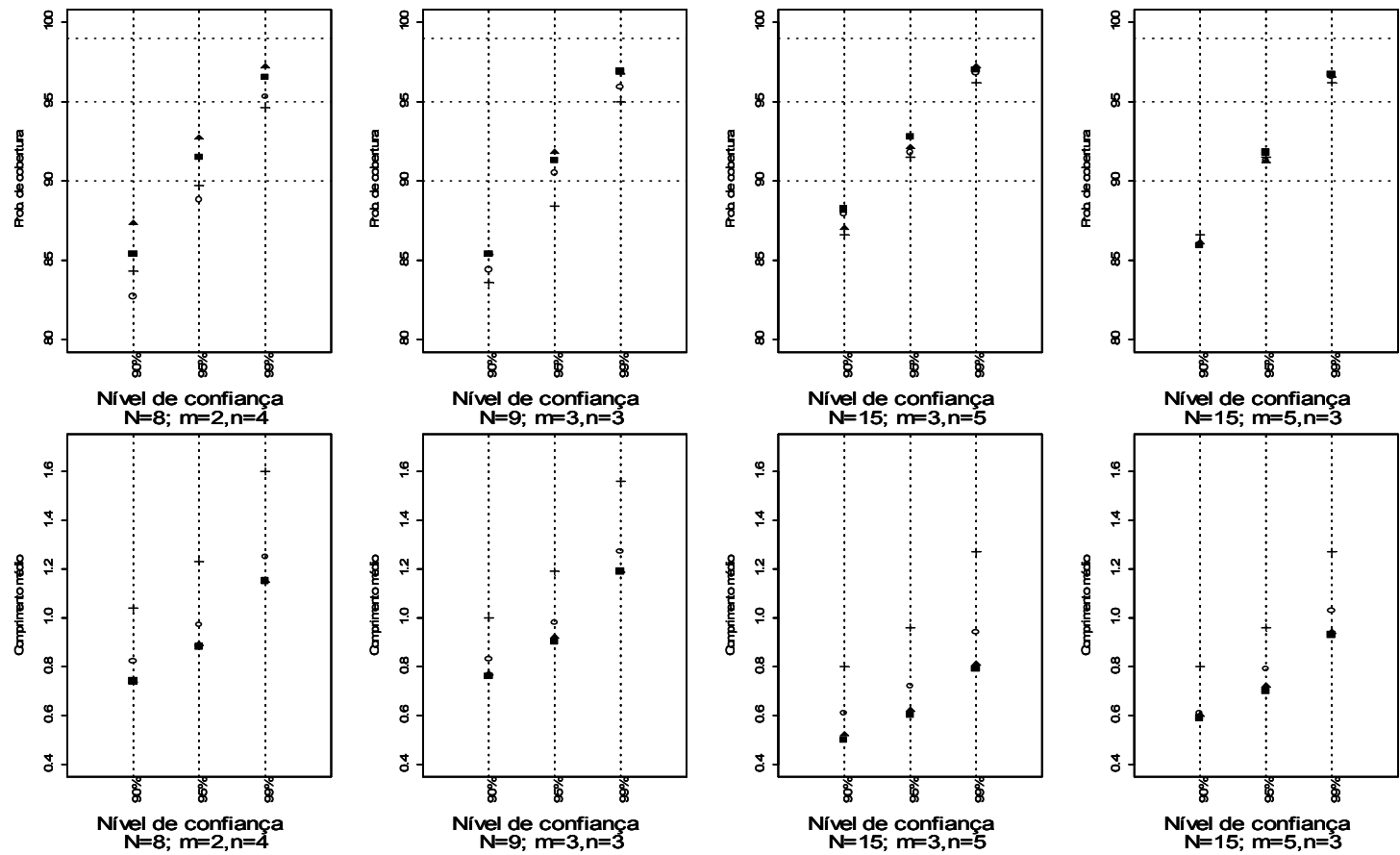
		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS
Normal	90%	15,73 (5,19)	16,44 (5,14)	18,49 (5,92)	22,42 (7,22)	16,58 (5,00)	17,13 (5,39)	18,40 (5,22)	21,67 (6,76)
	95%	18,80 (6,20)	19,65 (6,14)	22,09 (7,07)	26,80 (8,63)	19,81 (5,98)	20,48 (6,45)	21,99 (6,25)	25,42 (8,08)
	99%	24,70 (8,15)	25,82 (8,07)	29,03 (9,29)	35,20 (11,35)	26,03 (7,85)	26,91 (8,47)	28,89 (8,21)	33,40 (10,62)
Básico	90%	16,21 (5,30)	16,40 (5,08)	18,44 (5,85)	22,49 (7,71)	16,70 (5,05)	17,15 (5,40)	18,41 (5,23)	21,54 (6,70)
	95%	19,31 (6,39)	19,45 (6,08)	21,87 (7,00)	26,53 (9,03)	19,77 (5,96)	20,25 (6,32)	21,72 (6,16)	25,45 (7,85)
	99%	24,81 (7,90)	24,90 (7,57)	27,85 (8,71)	33,96 (11,49)	25,50 (7,57)	25,88 (7,90)	27,71 (7,71)	32,71 (10,00)
Percentil	90%	16,19 (5,13)	16,40 (5,08)	18,44 (5,85)	22,60 (7,50)	16,59 (5,16)	17,15 (5,40)	18,41 (5,23)	21,32 (6,60)
	95%	19,37 (6,29)	19,45 (6,08)	21,87 (7,00)	26,67 (8,81)	19,68 (6,09)	20,25 (6,32)	21,72 (6,16)	25,23 (7,74)
	99%	25,08 (7,78)	24,90 (7,57)	27,85 (8,71)	34,54 (11,33)	25,71 (7,95)	25,88 (7,90)	27,71 (7,71)	32,78 (9,85)
BCa	90%	16,21 (5,68)	16,87 (5,54)	18,68 (5,75)	22,84 (7,96)	16,88 (5,43)	16,91 (5,30)	18,44 (5,86)	22,10 (7,44)
	95%	19,28 (6,70)	20,06 (6,50)	22,20 (6,75)	27,19 (9,58)	20,10 (6,45)	20,15 (6,34)	21,95 (6,54)	26,37 (8,93)
	99%	24,79 (8,47)	25,69 (8,18)	28,39 (8,73)	34,51 (11,87)	25,87 (8,26)	25,86 (7,98)	28,06 (8,18)	33,71 (11,44)
Studentizado	90%	21,30 (12,99)	21,75 (12,94)	25,23 (13,47)	34,30 (17,88)	20,76 (11,20)	21,74 (10,86)	22,90 (9,15)	29,71 (13,78)
	95%	27,20 (17,62)	27,98 (17,33)	32,28 (18,52)	43,77 (23,41)	26,04 (13,87)	27,07 (13,75)	28,83 (12,28)	38,04 (19,05)
	99%	38,08 (29,46)	39,77 (29,01)	47,35 (28,73)	68,38 (48,82)	37,46 (23,93)	38,06 (21,76)	41,81 (22,52)	56,33 (33,57)

Tabela 3 (cont)- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média das medições médias diárias de MP₁₀, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita.

		<i>N=15; m=3,n=5</i>				<i>N=15; m=5,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 23,85$)	RSS- O.I. ($\sigma_e^2 = 238,59$)	SRS
Normal	90%	11,33 (2,82)	11,62 (2,63)	13,51 (2,67)	17,34 (4,24)	13,28 (3,21)	13,39 (3,22)	14,60 (3,20)	17,34 (4,24)
	95%	13,54 (3,37)	13,88 (3,14)	16,15 (3,19)	20,73 (5,07)	15,64 (3,83)	16,00 (3,85)	17,45 (3,82)	20,73 (5,07)
	99%	17,79 (4,43)	18,24 (4,13)	21,21 (4,19)	27,24 (6,67)	20,54 (5,04)	21,03 (5,07)	22,93 (5,02)	27,24 (6,67)
Básico	90%	11,45 (2,90)	11,62 (2,63)	13,51 (2,67)	17,59 (4,47)	13,27 (3,20)	13,39 (3,22)	14,60 (3,20)	17,59 (4,47)
	95%	13,61 (3,44)	13,76 (3,10)	16,03 (3,18)	20,87 (5,31)	15,75 (3,79)	15,88 (3,81)	17,29 (3,79)	20,87 (5,31)
	99%	17,72 (4,43)	17,75 (3,96)	20,69 (4,08)	27,10 (6,82)	20,47 (4,83)	20,50 (4,90)	22,35 (4,86)	27,10 (6,82)
Percentil	90%	11,23 (2,66)	11,62 (2,63)	13,51 (2,67)	17,60 (4,17)	13,34 (3,23)	13,39 (3,22)	14,60 (3,20)	17,60 (4,17)
	95%	13,37 (3,16)	13,76 (3,10)	16,03 (3,18)	20,93 (4,93)	15,84 (3,84)	15,88 (3,81)	17,29 (3,79)	20,93 (4,93)
	99%	17,56 (4,11)	17,75 (3,96)	20,69 (4,08)	27,45 (6,49)	20,82 (5,04)	20,50 (4,90)	22,35 (4,86)	27,45 (6,49)
BCa	90%	11,37 (2,85)	11,79 (2,75)	13,68 (2,91)	17,80 (4,57)	13,36 (3,68)	13,60 (3,40)	14,89 (3,47)	17,80 (4,57)
	95%	13,56 (3,40)	14,06 (3,28)	16,31 (3,50)	21,28 (4,59)	15,97 (4,42)	16,20 (4,07)	17,76 (4,16)	21,28 (4,59)
	99%	17,54 (4,37)	18,19 (4,21)	21,10 (4,59)	27,45 (7,04)	20,61 (5,70)	20,99 (5,26)	22,96 (5,40)	27,45 (7,04)
Studentizado	90%	12,55 (4,31)	13,17 (4,50)	15,63 (4,04)	20,42 (6,23)	15,31 (5,15)	14,98 (4,36)	17,01 (5,22)	20,42 (6,23)
	95%	15,39 (5,65)	16,00 (5,62)	19,16 (5,18)	25,18 (8,01)	18,69 (6,38)	18,32 (5,65)	20,72 (6,62)	25,18 (8,01)
	99%	20,65 (7,84)	21,55 (8,89)	25,56 (7,33)	34,72 (12,74)	25,28 (9,91)	24,52 (8,13)	28,09 (10,33)	34,72 (12,74)

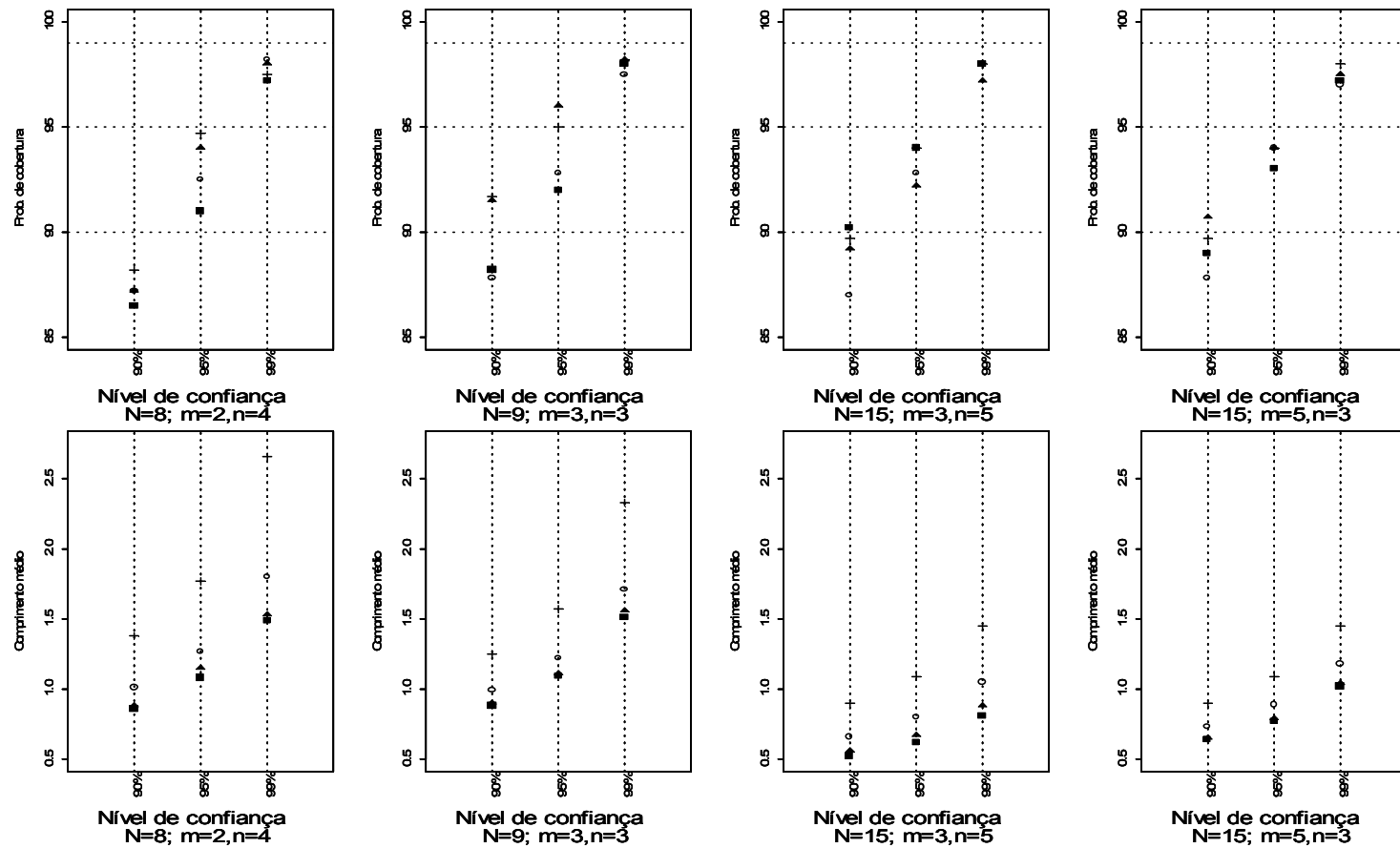
6.2 Estimação da média de uma população com distribuição normal

As Figuras 7 e 8 e as Tabelas 4 e 5 apresentam alguns resultados relativos à performance dos intervalos de confiança propostos na estimação da média de uma população com distribuição normal. Mais uma vez observa-se um ganho significativo de precisão dos intervalos propostos, sem prejuízo quanto à acurácia das estimativas, se comparado aos intervalos concebidos via *SRS*. Repare, por exemplo, que para o menor tamanho de amostra considerado ($N = 8 ; m = 2, n = 4$), o comprimento médio dos intervalos bootstrap do tipo percentil, originários de uma amostra por conjuntos ordenados, é igual a 0.74, considerando ordenação perfeita e imperfeita com $\sigma_\varepsilon^2 = 0.05$, enquanto via *SRS*, o comprimento médio é 1.04. Já para $N = 15 ; m = 3, n = 5$, o comprimento médio varia de 0.50, via *RSS*, para 0.80, considerando *SRS*. Verifica-se também um menor afastamento das probabilidades de cobertura em relação aos níveis de confiança fixados para todos os intervalos de confiança bootstrap considerados, em relação aos resultados referentes à estimação da média de MP_{10} (Tabela 2). No geral, o intervalo de confiança bootstrap studentizado forneceu as estimativas mais acuradas.



(+) *SRS* (■) *RSS* (Ordenação perfeita) (▲) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.05$)
 (○) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 7- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição normal, via *SRS* e *RSS* com ordenação perfeita e imperfeita.



(+) *SRS* (■) *RSS* (Ordenação perfeita) (▲) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.05$)

(○) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 8- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição normal, via *SRS* e *RSS* com ordenação perfeita e imperfeita.

Tabela 4- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição normal.

		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	85,9	84,2	82,9	82,7	86,3	85,6	84,2	83,9
	95%	90,2	89,3	89,5	89,2	91,7	92,1	90,9	90,6
	99%	97,1	97,0	95,8	95,6	96,6	96,8	96,8	96,2
Básico	90%	87,7	84,3	83,1	83,7	87,7	86,1	83,8	85,3
	95%	92,3	89,4	89,3	89,1	93,0	91,9	90,0	91,3
	99%	96,6	96,4	95,0	95,9	97,5	96,4	96,0	96,1
Percentil	90%	85,4	84,3	82,7	84,3	85,4	85,4	84,4	83,6
	95%	91,5	89,0	88,7	89,7	91,3	91,8	90,5	88,4
	99%	96,5	95,7	94,9	94,6	96,9	96,8	95,9	95,0
BCa	90%	85,6	87,3	82,7	82,8	84,1	85,1	83,6	86,0
	95%	91,5	92,7	88,8	88,2	89,8	91,4	89,5	91,3
	99%	96,3	97,2	95,3	95,1	95,1	96,7	95,2	95,5
Studentizado	90%	86,5	87,2*	87,2*	88,2*	88,2*	91,5*	87,8*	91,7*
	95%	91,0	94,0*	92,5	94,7*	92,0	96,0	92,8*	95,0*
	99%	97,2	98,0*	98,2*	97,5	98,0*	98,2	97,5	98,2*

Tabela 4 (cont)- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição normal.

		<i>N=15; m=3, n=5</i>				<i>N=15; m=5, n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	89,3*	87,8	86,0	86,1	88,6*	87,6	87,5	86,1
	95%	94,4*	93,5	92,2	91,5	93,5	92,6	92,7	91,5
	99%	98,7*	98,3	97,8	96,7	97,6	97,9	98,2	96,7
Básico	90%	90,0*	87,8	86,1	86,5	89,1*	87,7	87,9	86,5
	95%	94,3*	93,9*	91,7	91,5	93,9*	92,3	92,8	91,5
	99%	98,3*	98,4*	97,6	97,4	98,2	97,7	98,0	97,4
Percentil	90%	87,7	87,3	86,4	87,4	88,6*	88,0	87,7	87,4
	95%	93,3	93,0	92,5	93,9*	94,1*	92,7	92,5	93,9*
	99%	98,4*	97,9	97,4	97,6	97,9	97,7	97,6	97,6
BCa	90%	88,8*	87,9	87,8	84,5	89,4*	88,0	87,8	84,5
	95%	94,4*	93,5	93,1	90,4	93,6*	93,3	93,4	90,4
	99%	98,2	98,1	97,7	96,3	97,3	96,9	96,7	96,3
Studentizado	90%	90,2*	89,2*	87,0	89,7*	89,0*	90,7*	87,8	89,7*
	95%	94,0*	92,2	92,8*	94,0*	93,0*	94,0*	94,0*	94,0*
	99%	98,0*	97,2	98,0*	98,0*	97,2	97,5	97,0	98,0*

Tabela 5- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição normal.

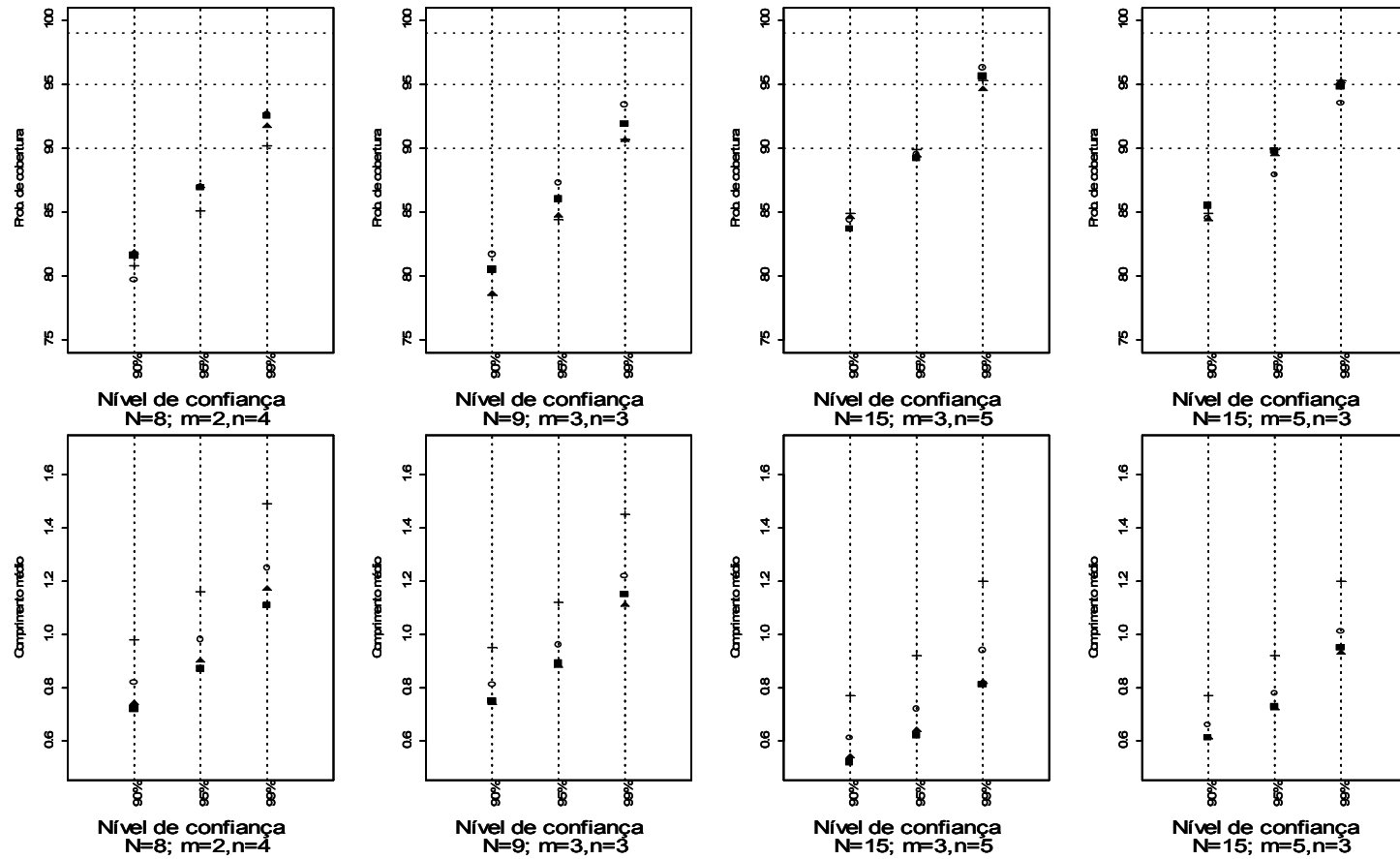
		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	0,74 (0,17)	0,74 (0,17)	0,82 (0,19)	1,05 (0,29)	0,76 (0,18)	0,77 (0,17)	0,83 (0,18)	1,01 (0,26)
	95%	0,88 (0,21)	0,89 (0,21)	0,98 (0,23)	1,25 (0,34)	0,91 (0,22)	0,92 (0,20)	0,99 (0,21)	1,20 (0,31)
	99%	1,16 (0,28)	1,17 (0,27)	1,29 (0,30)	1,65 (0,45)	1,20 (0,29)	1,21 (0,27)	1,30 (0,28)	1,58 (0,41)
Básico	90%	0,74 (0,18)	0,74 (0,17)	0,82 (0,19)	1,05 (0,28)	0,77 (0,18)	0,77 (0,17)	0,83 (0,18)	0,99 (0,25)
	95%	0,88 (0,21)	0,89 (0,21)	0,97 (0,23)	1,24 (0,34)	0,91 (0,22)	0,92 (0,20)	0,98 (0,21)	1,18 (0,30)
	99%	1,15 (0,28)	1,15 (0,27)	1,25 (0,29)	1,60 (0,44)	1,19 (0,28)	1,19 (0,27)	1,27 (0,27)	1,59 (0,40)
Percentil	90%	0,74 (0,18)	0,74 (0,17)	0,82 (0,19)	1,04 (0,29)	0,76 (0,18)	0,77 (0,17)	0,83 (0,18)	1,00 (0,25)
	95%	0,88 (0,21)	0,89 (0,21)	0,97 (0,23)	1,23 (0,34)	0,90 (0,22)	0,92 (0,20)	0,98 (0,21)	1,19 (0,30)
	99%	1,15 (0,28)	1,15 (0,27)	1,25 (0,29)	1,60 (0,45)	1,19 (0,29)	1,19 (0,27)	1,27 (0,27)	1,56 (0,40)
BCa	90%	0,73 (0,18)	0,76 (0,18)	0,83 (0,19)	1,07 (0,29)	0,76 (0,18)	0,77 (0,17)	0,84 (0,19)	1,01 (0,25)
	95%	0,87 (0,22)	0,90 (0,22)	0,99 (0,23)	1,27 (0,35)	0,91 (0,21)	0,92 (0,21)	1,00 (0,23)	1,20 (0,29)
	99%	1,13 (0,28)	1,16 (0,28)	1,27 (0,30)	1,63 (0,45)	1,17 (0,28)	1,19 (0,27)	1,28 (0,30)	1,54 (0,39)
Studentizado	90%	0,86 (0,28)	0,88 (0,27)	1,01 (0,30)	1,38 (0,44)	0,88 (0,27)	0,90 (0,25)	0,99 (0,29)	1,25 (0,35)
	95%	1,08 (0,36)	1,15 (0,35)	1,27 (0,39)	1,77 (0,59)	1,09 (0,34)	1,11 (0,31)	1,22 (0,35)	1,57 (0,45)
	99%	1,49 (0,56)	1,53 (0,56)	1,80 (0,71)	2,66 (1,11)	1,51 (0,55)	1,56 (0,61)	1,71 (0,56)	2,33 (0,38)

Tabela 5 (cont)- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição normal.

		<i>N=15; m=3, n=5</i>				<i>N=15; m=5, n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	0,50 (0,08)	0,52 (0,08)	0,61 (0,09)	0,79 (0,15)	0,59 (0,10)	0,60 (0,11)	0,67 (0,11)	0,79 (0,15)
	95%	0,60 (0,10)	0,62 (0,10)	0,73 (0,11)	0,94 (0,18)	0,71 (0,12)	0,72 (0,13)	0,80 (0,13)	0,94 (0,18)
	99%	0,79 (0,13)	0,81 (0,13)	0,96 (0,15)	1,24 (0,24)	0,93 (0,16)	0,95 (0,17)	1,05 (0,17)	1,24 (0,24)
Básico	90%	0,50 (0,08)	0,52 (0,08)	0,61 (0,09)	0,79 (0,14)	0,60 (0,11)	0,60 (0,11)	0,61 (0,11)	0,79 (0,14)
	95%	0,60 (0,10)	0,62 (0,10)	0,72 (0,11)	0,94 (0,17)	0,71 (0,13)	0,72 (0,13)	0,79 (0,13)	0,94 (0,17)
	99%	0,79 (0,14)	0,81 (0,13)	0,94 (0,15)	1,24 (0,23)	0,93 (0,17)	0,94 (0,17)	1,03 (0,17)	1,24 (0,23)
Percentil	90%	0,50 (0,09)	0,52 (0,08)	0,61 (0,09)	0,80 (0,14)	0,59 (0,11)	0,60 (0,11)	0,61 (0,11)	0,80 (0,14)
	95%	0,60 (0,10)	0,62 (0,10)	0,72 (0,11)	0,96 (0,17)	0,70 (0,13)	0,72 (0,13)	0,79 (0,13)	0,96 (0,17)
	99%	0,79 (0,14)	0,81 (0,13)	0,94 (0,15)	1,27 (0,23)	0,93 (0,17)	0,94 (0,17)	1,03 (0,17)	1,27 (0,23)
BCa	90%	0,50 (0,08)	0,51 (0,08)	0,61 (0,10)	0,80 (0,15)	0,60 (0,10)	0,61 (0,11)	0,66 (0,11)	0,80 (0,15)
	95%	0,60 (0,10)	0,61 (0,10)	0,73 (0,12)	0,96 (0,18)	0,71 (0,13)	0,73 (0,13)	0,79 (0,13)	0,96 (0,18)
	99%	0,78 (0,13)	0,80 (0,13)	0,94 (0,16)	1,24 (0,23)	0,92 (0,17)	0,94 (0,17)	1,02 (0,18)	1,24 (0,23)
Studentizado	90%	0,52 (0,11)	0,56 (0,11)	0,66 (0,13)	0,90 (0,19)	0,64 (0,14)	0,65 (0,14)	0,73 (0,16)	0,90 (0,19)
	95%	0,62 (0,14)	0,67 (0,14)	0,80 (0,16)	1,09 (0,24)	0,77 (0,18)	0,79 (0,18)	0,89 (0,19)	1,09 (0,24)
	99%	0,81 (0,19)	0,88 (0,20)	1,05 (0,23)	1,45 (0,36)	1,02 (0,25)	1,04 (0,25)	1,18 (0,28)	1,45 (0,36)

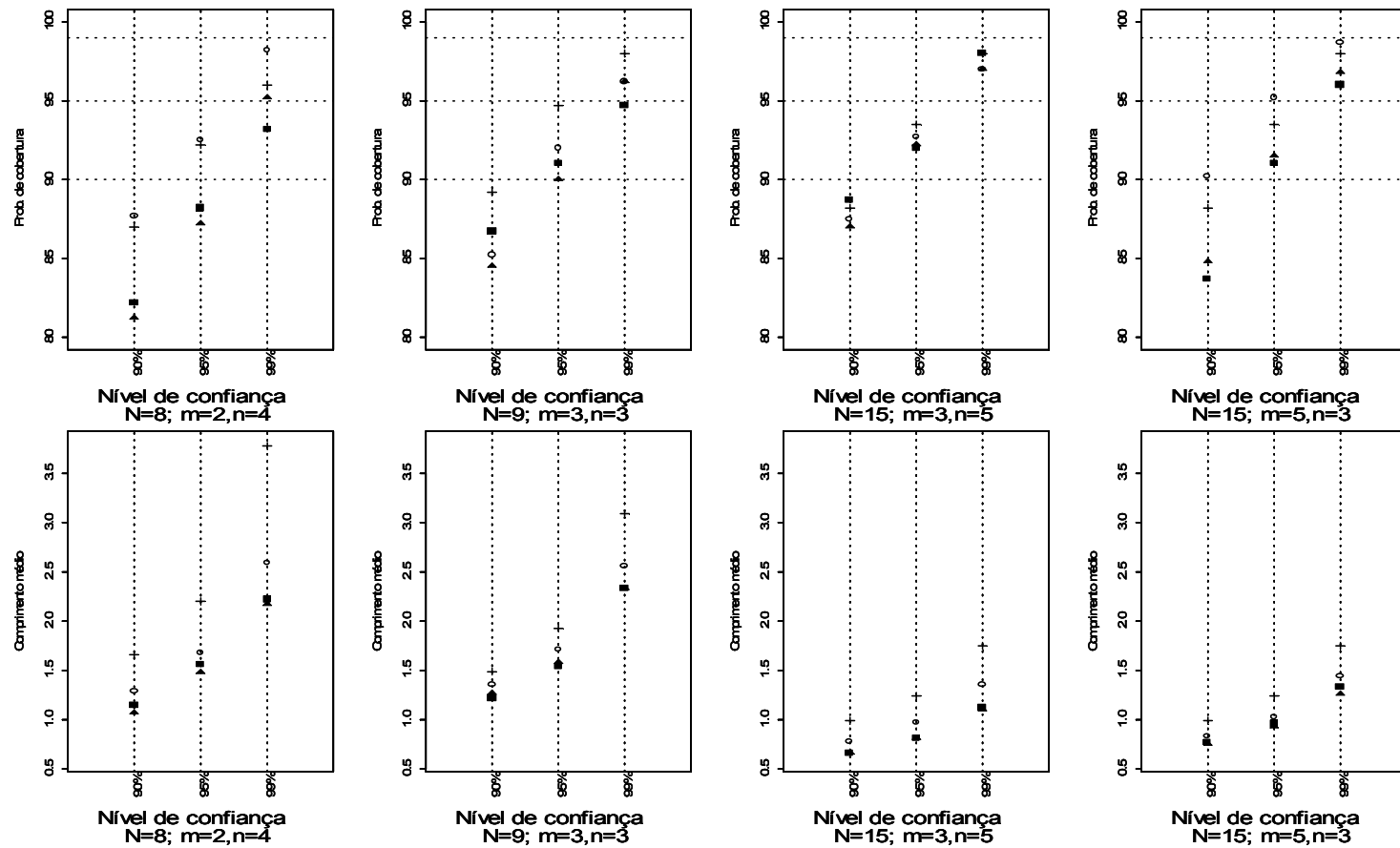
6.3 Estimação da média de uma população com distribuição exponencial

Os resultados obtidos para esta distribuição encontram-se sumarizados nas Figuras 9 e 10 e nas Tabelas 6 e 7. Outra vez pode-se destacar a maior precisão dos intervalos propostos, o desempenho semelhante dos intervalos bootstrap percentil, básico, normal e BCa, além da recorrência, desta vez um pouco mais acentuada, dos problemas detectados para as distribuições anteriores quanto à acurácia dos métodos bootstrap não paramétricos. Neste quesito, mais uma vez os intervalos do tipo studentizado se diferenciaram por sua maior acurácia.



(+) SRS (■) RSS (Ordenação perfeita) (▲) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.05$)
 (○) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 9- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição exponencial, via SRS e RSS com ordenação perfeita e imperfeita.



(+) SRS (■) RSS (Ordenação perfeita) (▲) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.05$)
 (○) RSS (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 10- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição exponencial, via SRS e RSS com ordenação perfeita e imperfeita.

Tabela 6- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição exponencial.

		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	80,9	82,4	79,3	77,9	80,7	78,8	80,9	80,7
	95%	86,0	87,1	86,7	82,8	85,7	84,8	87,5	84,8
	99%	93,3	91,2	93,0	88,8	91,7	91,3	93,2	90,6
Básico	90%	83,0	82,0	79,3	77,1	78,2	78,6	80,2	80,5
	95%	86,7	86,1	85,5	81,4	83,3	83,4	86,3	84,8
	99%	92,3	90,1	91,4	85,8	88,6	88,9	91,5	88,9
Percentil	90%	81,6	81,8	79,7	80,8	80,5	78,6	81,7	78,5
	95%	86,9	87,0	87,0	85,1	86,0	84,7	87,3	84,4
	99%	92,5	91,7	92,7	90,2	91,9	90,6	93,4	90,7
BCa	90%	81,6	80,3	80,6	83,3	82,1	82,2	83,2	83,2
	95%	87,3	86,5	87,2	87,9	87,0	87,7	87,9	88,3
	99%	92,5	92,3	94,2	93,3	92,5	93,1	93,8	92,8
Studentizado	90%	82,2	81,2	87,7	87,0*	86,7	84,5	85,2	89,2*
	95%	88,2	87,2	92,5	92,2	91,0	90,0	92,0	94,7*
	99%	93,2	95,2	98,2	96,0	94,7	96,2	96,2	98,0*

Tabela 6 (cont)- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição exponencial.

		<i>N=15; m=3,n=5</i>				<i>N=15; m=5,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	84,3	84,5	84,6	85,2	83,2	84,6	84,0	85,2
	95%	89,8	89,3	89,5	89,2	88,1	89,1	87,4	89,2
	99%	94,6	94,9	96,3	94,4	94,4	95,2	93,1	94,4
Básico	90%	83,2	84,1	84,4	83,3	82,9	84,2	83,2	83,3
	95%	88,7	88,9	89,0	87,4	87,5	88,5	87,2	87,4
	99%	94,0	94,0	95,1	92,2	93,2	94,3	91,7	92,2
Percentil	90%	83,7	84,6	84,4	84,9	85,5	84,4	84,5	84,9
	95%	89,2	89,4	89,5	89,9	89,8	89,5	87,9	89,9
	99%	95,6	94,6	96,3	95,3	94,8	95,2	93,5	95,3
BCa	90%	84,3	83,4	85,1	84,4	85,8	85,8	85,5	84,4
	95%	90,9	89,2	90,8	89,7	91,2	91,3	91,1	89,7
	99%	96,1	95,3	96,0	94,9	96,2	95,4	96,7	94,9
Studentizado	90%	88,7*	87,0*	87,5*	88,2*	83,7	84,8	90,2*	88,2*
	95%	92,0	92,2	92,7*	93,5*	91,0	91,5	95,2*	93,5*
	99%	98,0*	97,0	97,0	98,0*	96,0	96,8	98,7*	98,0*

Tabela 7- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição exponencial.

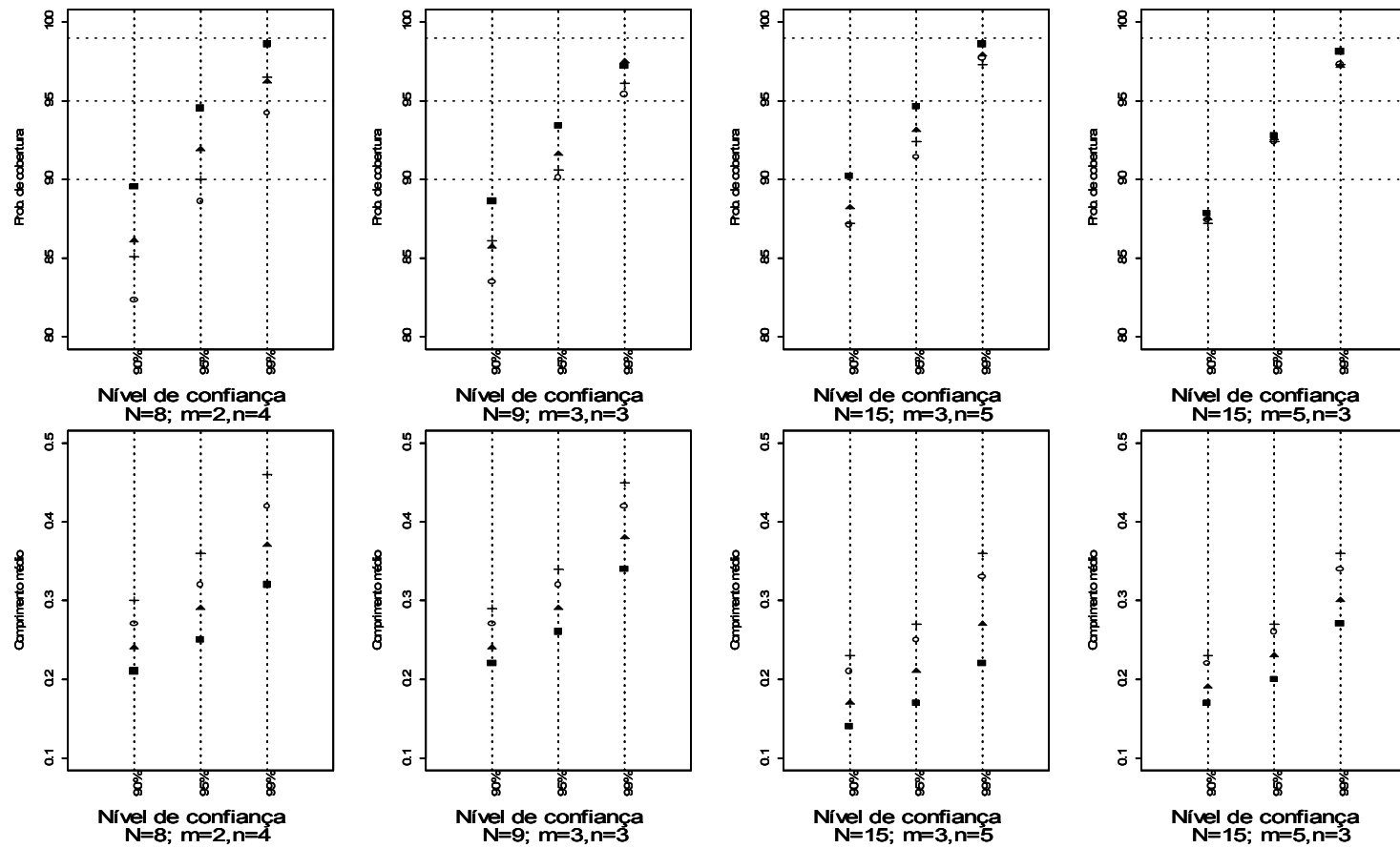
		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	0,73 (0,33)	0,76 (0,34)	0,83 (0,34)	0,97 (0,44)	0,74 (0,32)	0,74 (0,30)	0,81 (0,31)	0,95 (0,41)
	95%	0,88 (0,40)	0,90 (0,41)	0,99 (0,40)	1,16 (0,52)	0,89 (0,38)	0,89 (0,36)	0,97 (0,37)	1,14 (0,49)
	99%	1,15 (0,53)	1,19 (0,54)	1,31 (0,53)	1,52 (0,69)	1,17 (0,50)	1,17 (0,48)	1,27 (0,49)	1,50 (0,65)
Básico	90%	0,76 (0,34)	0,74 (0,32)	0,82 (0,32)	0,99 (0,47)	0,75 (0,34)	0,74 (0,30)	0,81 (0,31)	0,95 (0,40)
	95%	0,91 (0,42)	0,90 (0,40)	0,98 (0,40)	1,15 (0,54)	0,88 (0,39)	0,88 (0,35)	0,96 (0,36)	1,11 (0,46)
	99%	1,15 (0,50)	1,12 (0,47)	1,25 (0,48)	1,47 (0,69)	1,12 (0,48)	1,11 (0,43)	1,22 (0,45)	1,42 (0,58)
Percentil	90%	0,72 (0,30)	0,74 (0,32)	0,82 (0,32)	0,98 (0,44)	0,75 (0,33)	0,74 (0,30)	0,81 (0,31)	0,95 (0,42)
	95%	0,87 (0,38)	0,90 (0,40)	0,98 (0,40)	1,16 (0,51)	0,89 (0,38)	0,88 (0,35)	0,96 (0,36)	1,12 (0,49)
	99%	1,11 (0,46)	1,12 (0,47)	1,25 (0,48)	1,49 (0,65)	1,15 (0,48)	1,11 (0,43)	1,22 (0,45)	1,45 (0,63)
BCa	90%	0,79 (0,36)	0,77 (0,38)	0,84 (0,39)	1,06 (0,50)	0,78 (0,36)	0,78 (0,35)	0,85 (0,37)	0,98 (0,46)
	95%	0,88 (0,41)	0,91 (0,43)	0,99 (0,40)	1,26 (0,59)	0,93 (0,44)	0,93 (0,42)	1,02 (0,45)	1,17 (0,55)
	99%	1,12 (0,52)	1,17 (0,54)	1,26 (0,50)	1,59 (0,72)	1,18 (0,52)	1,19 (0,52)	1,29 (0,54)	1,48 (0,67)
Studentizado	90%	1,15 (1,19)	1,07 (0,98)	1,29 (1,52)	1,66 (1,18)	1,22 (1,28)	1,27 (1,09)	1,35 (1,18)	1,49 (1,01)
	95%	1,56 (1,68)	1,48 (1,66)	1,68 (1,89)	2,20 (1,73)	1,54 (1,56)	1,58 (1,33)	1,71 (1,40)	1,93 (1,30)
	99%	2,22 (2,34)	2,17 (3,22)	2,59 (3,60)	3,78 (3,47)	2,33 (2,72)	2,33 (2,16)	2,56 (2,28)	3,09 (2,39)

Tabela 7 (cont)- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição exponencial.

		<i>N=15; m=3, n=5</i>				<i>N=15; m=5, n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	0,52 (0,18)	0,54 (0,18)	0,61 (0,18)	0,77 (0,26)	0,60 (0,21)	0,61 (0,21)	0,66 (0,21)	0,77 (0,26)
	95%	0,62 (0,21)	0,65 (0,22)	0,73 (0,22)	0,92 (0,31)	0,72 (0,25)	0,73 (0,25)	0,79 (0,25)	0,92 (0,31)
	99%	0,81 (0,28)	0,85 (0,29)	0,96 (0,29)	1,21 (0,41)	0,95 (0,34)	0,97 (0,33)	1,04 (0,33)	1,21 (0,41)
Básico	90%	0,52 (0,19)	0,54 (0,18)	0,61 (0,18)	0,75 (0,25)	0,60 (0,20)	0,61 (0,21)	0,66 (0,21)	0,75 (0,25)
	95%	0,62 (0,22)	0,64 (0,22)	0,72 (0,21)	0,89 (0,29)	0,71 (0,23)	0,72 (0,24)	0,78 (0,24)	0,89 (0,29)
	99%	0,80 (0,27)	0,82 (0,27)	0,94 (0,27)	1,16 (0,38)	0,92 (0,30)	0,93 (0,31)	1,01 (0,31)	1,16 (0,38)
Percentil	90%	0,52 (0,18)	0,54 (0,18)	0,61 (0,18)	0,77 (0,24)	0,61 (0,21)	0,61 (0,21)	0,66 (0,21)	0,77 (0,24)
	95%	0,62 (0,21)	0,64 (0,22)	0,72 (0,21)	0,92 (0,29)	0,73 (0,25)	0,72 (0,24)	0,78 (0,24)	0,92 (0,29)
	99%	0,81 (0,26)	0,82 (0,27)	0,94 (0,27)	1,20 (0,38)	0,95 (0,32)	0,93 (0,31)	1,01 (0,31)	1,20 (0,38)
BCa	90%	0,54 (0,22)	0,56 (0,22)	0,63 (0,20)	0,80 (0,30)	0,62 (0,24)	0,63 (0,22)	0,69 (0,24)	0,80 (0,30)
	95%	0,64 (0,26)	0,67 (0,26)	0,76 (0,24)	0,97 (0,37)	0,74 (0,29)	0,75 (0,27)	0,83 (0,30)	0,97 (0,37)
	99%	0,82 (0,32)	0,86 (0,32)	0,97 (0,31)	1,23 (0,45)	0,95 (0,35)	0,97 (0,33)	1,06 (0,36)	1,23 (0,45)
Studentizado	90%	0,66 (0,44)	0,66 (0,37)	0,78 (0,37)	0,99 (0,46)	0,77 (0,52)	0,75 (0,39)	0,83 (0,35)	0,99 (0,46)
	95%	0,81 (0,52)	0,81 (0,45)	0,97 (0,48)	1,24 (0,61)	0,96 (0,69)	0,93 (0,51)	1,03 (0,47)	1,24 (0,61)
	99%	1,12 (0,75)	1,10 (0,64)	1,35 (0,71)	1,75 (0,92)	1,33 (0,96)	1,26 (0,70)	1,44 (0,69)	1,75 (0,92)

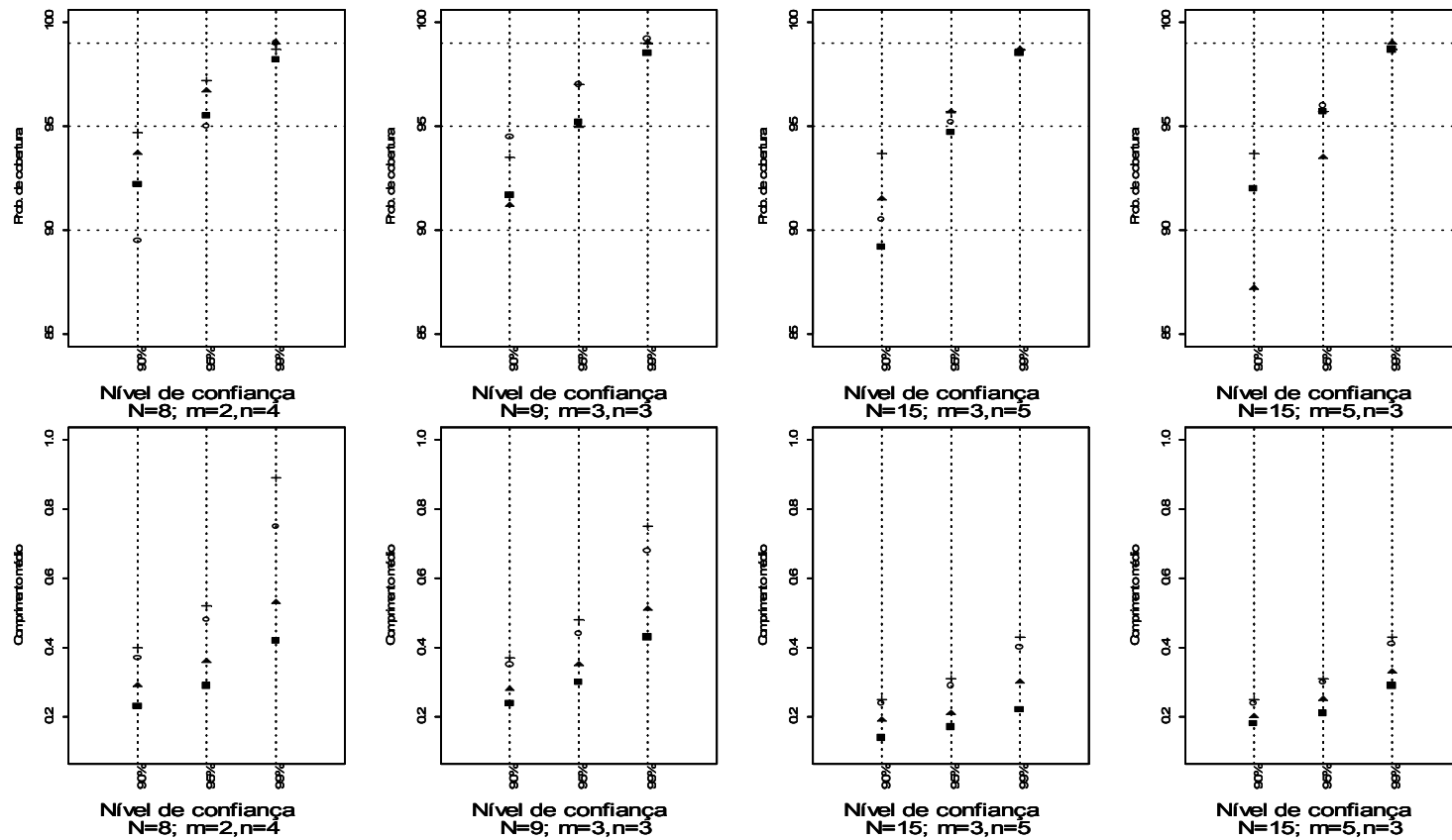
6.4 Estimação da média de uma população com distribuição uniforme

As Figuras 11 e 12 e as Tabelas 8 e 9 possibilitam a análise dos resultados alcançados na estimação da média de uma população com distribuição uniforme. Para esta distribuição verificou-se os melhores resultados, no que diz respeito à precisão e à acurácia dos intervalos. No mais, outra vez os intervalos de confiança obtidos via *RSS* são mais precisos, o que fica nítido através das tabelas e gráficos presentes nesta seção.



(+) *SRS* (■) *RSS* (Ordenação perfeita) (▲) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.05$)
 (○) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 11- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição uniforme, via *SRS* e *RSS* com ordenação perfeita e imperfeita.



(+) SRS (■) RSS (Ordenação perfeita) (▲) RSS (Ordenação imperfeita, $\sigma_c^2 = 0.05$)
 (○) RSS (Ordenação imperfeita, $\sigma_c^2 = 0.50$)

Figura 12- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição uniforme, via SRS e RSS com ordenação perfeita e imperfeita.

Tabela 8- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição uniforme.

		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	88,7*	85,6	81,4	82,8	87,6	86,0	83,2	83,4
	95%	93,4	91,8	87,2	87,9	93,3	91,4	89,3	90,3
	99%	97,8	96,3	93,7	94,8	97,9	97,4	95,0	96,0
Básico	90%	86,2	84,9	80,8	82,4	87,7	85,2	82,6	83,5
	95%	91,2	90,5	85,7	87,3	92,8	90,6	87,3	88,5
	99%	96,0	95,7	91,7	93,1	96,6	96,1	92,9	94,1
Percentil	90%	89,5*	86,1	82,3	85,1	88,6*	85,7	83,5	86,1
	95%	94,5*	91,9	88,6	90,0	93,4	91,6	90,1	90,6
	99%	98,6*	96,2	94,2	96,5	97,2	97,5	95,4	96,1
BCa	90%	87,6	88,7	87,6	86,8	88,3*	88,5*	86,7	86,8
	95%	92,1	93,3	92,0	90,9	92,9	92,8	90,9	92,9
	99%	96,6	97,0	96,6	94,6	96,9	97,5	96,4	96,6
Studentizado	90%	92,2*	93,7	89,5*	94,7	91,7*	91,2	94,5	93,5
	95%	95,5*	96,7*	95,0*	97,2	95,2*	95,0	97,0*	97,0*
	99%	98,2*	99,0*	99,0*	98,7*	98,5*	99,0	99,2*	99,0*

Tabela 8 (cont)- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição uniforme.

		<i>N=15; m=3,n=5</i>				<i>N=15; m=5,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	91,0*	87,3	86,4	85,6	87,5	86,8	87,1	85,6
	95%	95,6*	93,0	91,4	91,9	93,2	92,7	92,2	91,9
	99%	99,0*	98,2	97,0	96,9	97,8	97,6	96,6	96,9
Básico	90%	90,0*	87,1	85,7	85,5	88,7*	86,8	85,8	85,5
	95%	94,6*	92,5	90,8	90,6	92,5	92,2	90,8	90,6
	99%	98,4*	97,9	96,2	95,8	97,0	96,9	95,9	95,8
Percentil	90%	90,2*	88,2*	87,1	87,2	87,8	87,5	87,4	87,2
	95%	94,6*	93,1	91,4	92,4	92,8	92,6	92,4	92,4
	99%	98,6*	97,9	97,7	97,3	98,1	97,2	97,3	97,3
BCa	90%	90,0*	88,8*	87,5	88,1*	90,5*	89,5*	87,8	88,1*
	95%	95,1*	93,6*	92,4	94,4*	96,0*	94,4*	91,6	94,4*
	99%	98,9*	98,0	97,8	98,6*	98,5*	97,1	97,6	98,6*
Studentizado	90%	89,2*	91,5*	90,5*	93,7	92,0*	87,2*	92,0*	93,7
	95%	94,7*	95,7*	95,2*	95,7*	95,7*	93,5*	96,0*	95,7*
	99%	98,5*	98,7*	98,5*	98,7*	98,7*	99,0*	98,7*	98,7*

Tabela 9- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição uniforme.

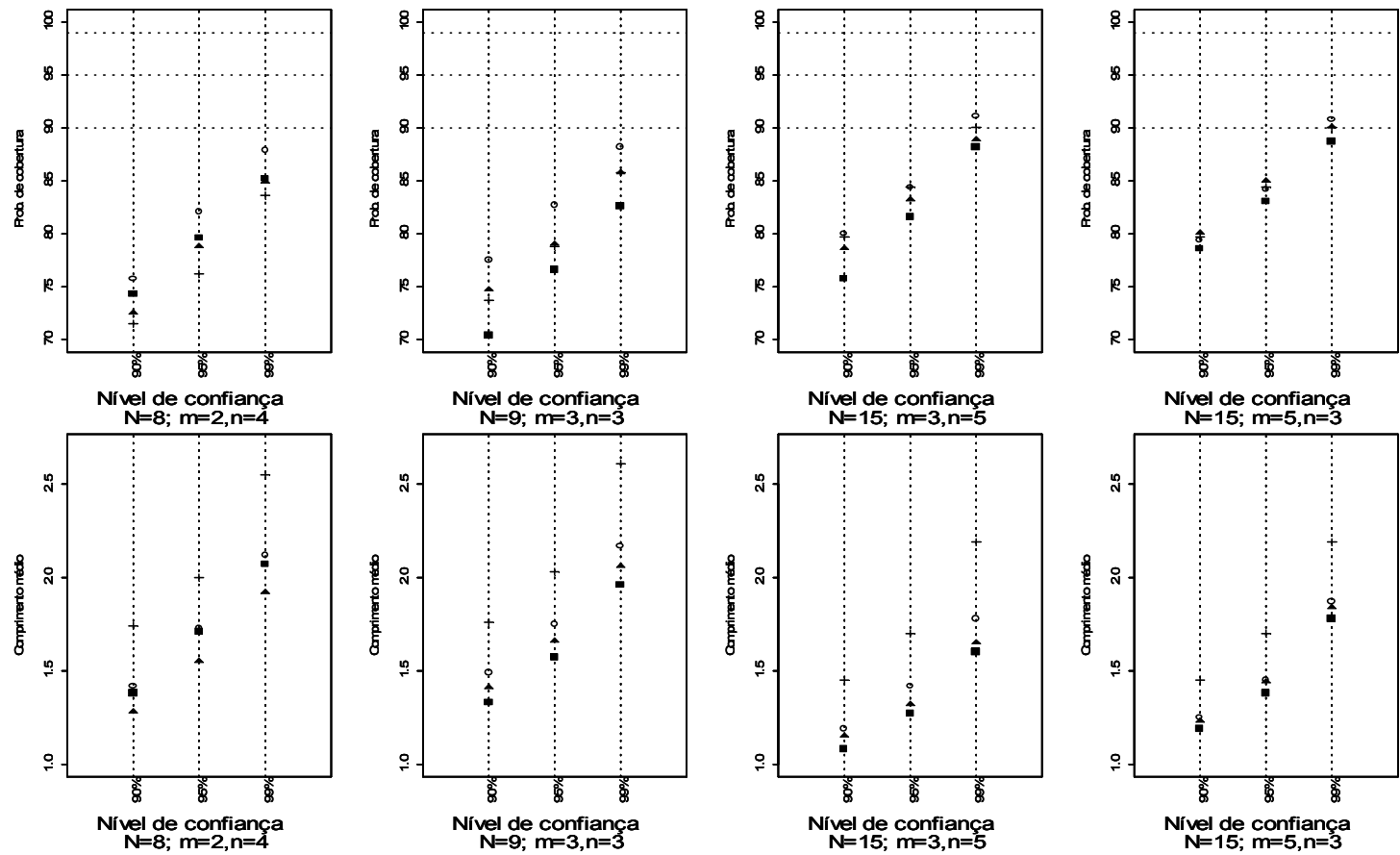
		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	0,21 (0,03)	0,24 (0,04)	0,27 (0,05)	0,30 (0,05)	0,22 (0,03)	0,24 (0,03)	0,27 (0,04)	0,29 (0,05)
	95%	0,25 (0,04)	0,29 (0,04)	0,33 (0,06)	0,36 (0,07)	0,26 (0,04)	0,29 (0,04)	0,32 (0,05)	0,34 (0,06)
	99%	0,33 (0,05)	0,38 (0,06)	0,43 (0,08)	0,48 (0,09)	0,34 (0,05)	0,38 (0,05)	0,43 (0,07)	0,45 (0,08)
Básico	90%	0,21 (0,03)	0,24 (0,04)	0,27 (0,05)	0,30 (0,06)	0,22 (0,03)	0,24 (0,03)	0,27 (0,04)	0,29 (0,05)
	95%	0,25 (0,04)	0,29 (0,04)	0,32 (0,06)	0,36 (0,07)	0,26 (0,04)	0,29 (0,04)	0,32 (0,05)	0,34 (0,06)
	99%	0,32 (0,05)	0,37 (0,06)	0,42 (0,08)	0,46 (0,09)	0,34 (0,05)	0,38 (0,06)	0,42 (0,07)	0,44 (0,08)
Percentil	90%	0,21 (0,03)	0,24 (0,04)	0,27 (0,05)	0,30 (0,06)	0,22 (0,03)	0,24 (0,03)	0,27 (0,04)	0,29 (0,05)
	95%	0,25 (0,04)	0,29 (0,04)	0,32 (0,06)	0,36 (0,07)	0,26 (0,04)	0,29 (0,04)	0,32 (0,05)	0,34 (0,06)
	99%	0,32 (0,05)	0,37 (0,06)	0,42 (0,08)	0,46 (0,09)	0,34 (0,05)	0,38 (0,06)	0,42 (0,07)	0,45 (0,08)
BCa	90%	0,21 (0,03)	0,24 (0,04)	0,28 (0,05)	0,30 (0,06)	0,21 (0,03)	0,24 (0,03)	0,27 (0,05)	0,29 (0,05)
	95%	0,25 (0,04)	0,29 (0,04)	0,33 (0,06)	0,36 (0,07)	0,25 (0,04)	0,29 (0,04)	0,32 (0,06)	0,35 (0,06)
	99%	0,32 (0,05)	0,37 (0,06)	0,43 (0,08)	0,46 (0,09)	0,33 (0,05)	0,38 (0,05)	0,42 (0,07)	0,44 (0,07)
Studentizado	90%	0,23 (0,05)	0,29 (0,07)	0,37 (0,10)	0,40 (0,11)	0,24 (0,05)	0,28 (0,06)	0,35 (0,10)	0,37 (0,08)
	95%	0,29 (0,07)	0,36 (0,10)	0,48 (0,16)	0,52 (0,11)	0,30 (0,07)	0,35 (0,08)	0,44 (0,13)	0,48 (0,13)
	99%	0,42 (0,15)	0,53 (0,13)	0,75 (0,36)	0,89 (0,52)	0,43 (0,13)	0,51 (0,16)	0,68 (0,29)	0,75 (0,32)

Tabela 9 (cont)- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição uniforme.

		<i>N=15; m=3, n=5</i>				<i>N=15; m=5, n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	0,14 (0,01)	0,17 (0,02)	0,21 (0,02)	0,23 (0,02)	0,17 (0,02)	0,19 (0,02)	0,22 (0,02)	0,23 (0,02)
	95%	0,17 (0,01)	0,21 (0,02)	0,25 (0,03)	0,27 (0,03)	0,20 (0,02)	0,23 (0,02)	0,26 (0,03)	0,27 (0,03)
	99%	0,22 (0,02)	0,27 (0,03)	0,34 (0,04)	0,36 (0,04)	0,26 (0,03)	0,30 (0,03)	0,34 (0,04)	0,36 (0,04)
Básico	90%	0,14 (0,01)	0,17 (0,02)	0,21 (0,02)	0,23 (0,03)	0,17 (0,02)	0,19 (0,02)	0,22 (0,02)	0,23 (0,03)
	95%	0,17 (0,01)	0,21 (0,02)	0,25 (0,03)	0,27 (0,03)	0,20 (0,02)	0,23 (0,02)	0,26 (0,03)	0,27 (0,03)
	99%	0,22 (0,02)	0,27 (0,03)	0,33 (0,04)	0,36 (0,04)	0,27 (0,03)	0,30 (0,03)	0,34 (0,04)	0,36 (0,04)
Percentil	90%	0,14 (0,01)	0,17 (0,02)	0,21 (0,02)	0,23 (0,03)	0,17 (0,02)	0,19 (0,02)	0,22 (0,02)	0,23 (0,03)
	95%	0,17 (0,01)	0,21 (0,02)	0,25 (0,03)	0,27 (0,03)	0,20 (0,02)	0,23 (0,02)	0,26 (0,03)	0,27 (0,03)
	99%	0,22 (0,02)	0,27 (0,03)	0,33 (0,04)	0,36 (0,04)	0,27 (0,03)	0,30 (0,03)	0,34 (0,04)	0,36 (0,04)
BCa	90%	0,14 (0,01)	0,17 (0,02)	0,21 (0,02)	0,23 (0,03)	0,17 (0,02)	0,19 (0,02)	0,22 (0,02)	0,23 (0,03)
	95%	0,17 (0,01)	0,21 (0,02)	0,25 (0,03)	0,27 (0,03)	0,20 (0,02)	0,23 (0,02)	0,26 (0,03)	0,27 (0,03)
	99%	0,22 (0,02)	0,27 (0,03)	0,33 (0,04)	0,35 (0,04)	0,26 (0,03)	0,29 (0,03)	0,34 (0,04)	0,35 (0,04)
Studentizado	90%	0,14 (0,02)	0,19 (0,03)	0,24 (0,04)	0,25 (0,04)	0,18 (0,02)	0,20 (0,03)	0,24 (0,04)	0,25 (0,04)
	95%	0,17 (0,02)	0,21 (0,04)	0,29 (0,05)	0,31 (0,05)	0,21 (0,03)	0,25 (0,04)	0,30 (0,05)	0,31 (0,05)
	99%	0,22 (0,03)	0,30 (0,06)	0,40 (0,09)	0,43 (0,08)	0,29 (0,05)	0,33 (0,06)	0,41 (0,08)	0,43 (0,08)

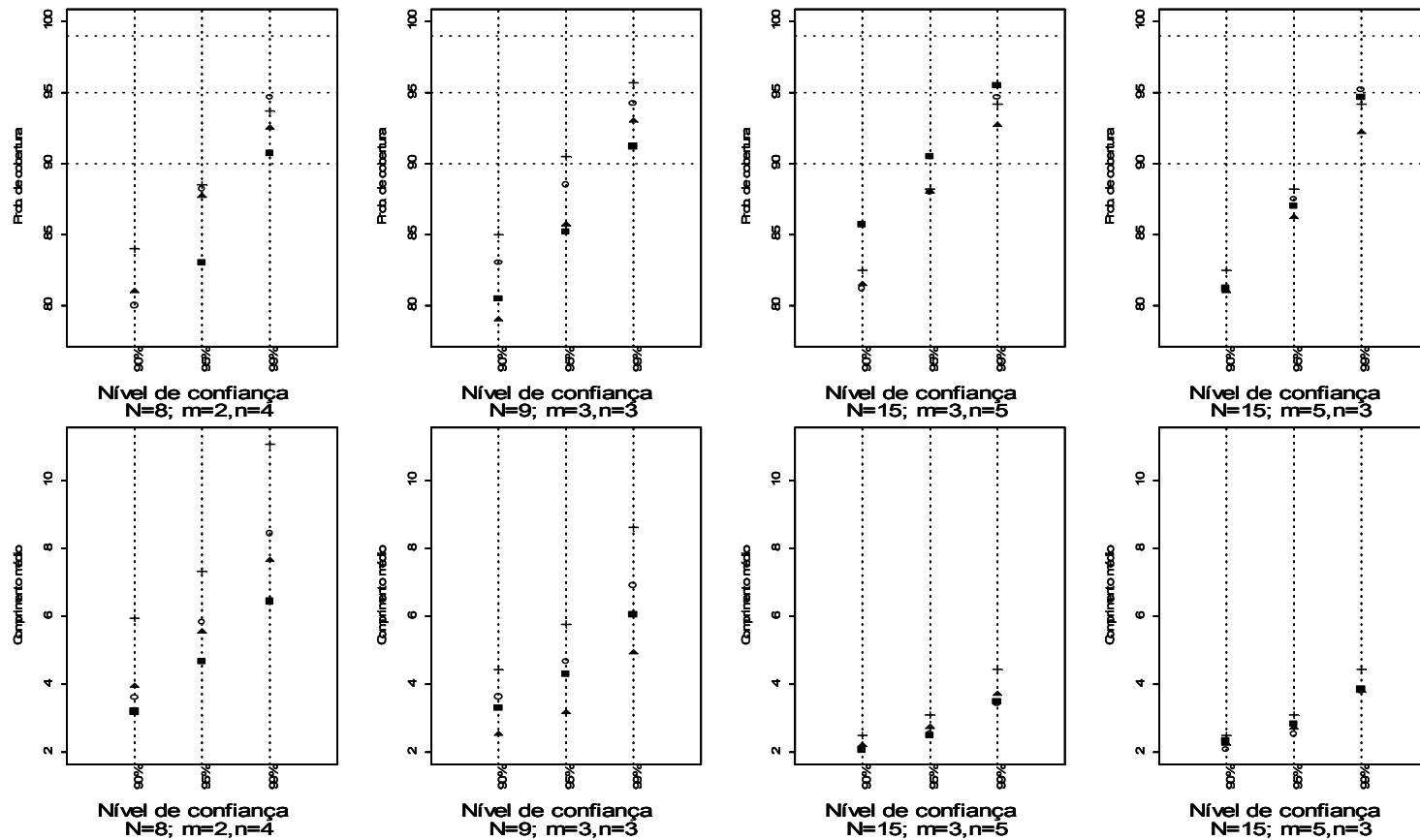
6.5 Estimação da média de uma população com distribuição lognormal

Seguem as Figuras 13 e 14 e as Tabelas 10 e 11, com os resultados relativos à estimação da média de uma população com distribuição lognormal. A distribuição lognormal ocasionou o maior distanciamento das probabilidades de cobertura simulada e teórica, tanto em relação aos intervalos propostos, como para aqueles baseados em amostragem aleatória simples. As acurácias verificadas para as estimativas produzidas pelos dois tipos de amostragem são bastante semelhantes. Os intervalos de confiança bootstrap do tipo studentizado mais uma vez proporcionaram estimativas mais acuradas. Os intervalos originários do esquema de reamostragem proposto se destacam pelo acentuado aumento quanto à precisão das estimativas, o que é perceptível pelos gráficos dos comprimentos médios dos intervalos.



(+) *SRS* (■) *RSS* (Ordenação perfeita) (▲) *RSS* (Ordenação imperfeita, $\sigma_\epsilon^2 = 0.05$)
 (○) *RSS* (Ordenação imperfeita, $\sigma_\epsilon^2 = 0.50$)

Figura 13- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição lognormal, via *SRS* e *RSS* com ordenação perfeita e imperfeita.



(+) *SRS* (■) *RSS* (Ordenação perfeita) (▲) *RSS* (Ordenação imperfeita, $\sigma_{\epsilon}^2 = 0.05$)

(○) *RSS* (Ordenação imperfeita, $\sigma_{\epsilon}^2 = 0.50$)

Figura 14- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição lognormal, via *SRS* e *RSS* com ordenação perfeita e imperfeita.

Tabela 10- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição lognormal.

		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	74,0	72,7	76,0	70,0	70,1	74,2	76,3	72,4
	95%	79,8	78,3	82,5	75,0	76,3	79,5	82,2	78,1
	99%	84,9	85,0	87,8	82,2	82,2	85,1	88,1	83,7
Básico	90%	74,0	72,3	75,2	69,0	69,7	73,8	75,9	71,6
	95%	78,3	77,1	81,6	72,4	74,8	78,3	80,6	75,2
	99%	83,6	83,4	85,9	77,9	79,9	82,9	85,8	79,6
Percentil	90%	74,3	72,5	75,7	71,5	70,4	74,7	77,5	73,7
	95%	79,6	78,8	82,1	76,2	76,6	79,0	82,7	78,8
	99%	85,2	84,9	87,9	83,6	82,6	85,8	88,2	85,8
BCa	90%	74,1	76,4	74,5	70,7	74,0	75,4	75,3	75,0
	95%	79,8	80,7	80,3	77,5	80,8	83,1	82,8	81,4
	99%	86,8	87,3	88,3	84,6	87,6	88,8	88,5	87,7
Studentizado	90%	73,0	81,0	80,0	84,0	80,5	79,0	83,0	85,0
	95%	83,0	87,7	88,2	88,5	85,2	85,7	88,5	90,5
	99%	90,7	92,5	94,7	93,7	91,2	93,0	94,2	95,7

Tabela 10 (cont)- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição lognormal.

		<i>N=15; m=3,n=5</i>				<i>N=15; m=5,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	75,5	78,8	79,8	79,2	78,8	79,1	78,9	79,2
	95%	80,2	82,7	84,6	84,2	82,5	84,4	83,4	84,2
	99%	87,5	88,6	90,8	88,9	88,0	90,0	90,2	88,9
Básico	90%	75,0	78,0	79,7	78,2	76,8	78,5	78,1	78,2
	95%	79,7	81,8	83,9	82,1	81,4	82,5	82,4	82,1
	99%	86,3	87,3	89,5	86,3	86,3	87,8	88,3	86,3
Percentil	90%	75,7	78,6	80,0	79,7	78,6	80,1	79,4	79,7
	95%	81,6	83,2	84,4	84,4	83,1	85,0	84,2	84,4
	99%	88,2	88,9	91,1	90,1	88,7	90,1	90,8	90,1
BCa	90%	79,4	78,0	80,3	79,2	77,7	78,0	78,1	79,2
	95%	85,8	83,6	85,9	84,9	84,3	84,7	84,7	84,9
	99%	90,8	90,9	92,1	92,1	90,9	91,1	92,0	92,1
Studentizado	90%	85,7	81,5	81,2	82,5	81,2	81,0	81,2	82,5
	95%	90,5	88,0	88,0	88,2	87,0	86,2	87,5	88,2
	99%	95,5	92,7	94,7	94,2	94,7	92,2	95,2	94,2

Tabela 11- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição lognormal.

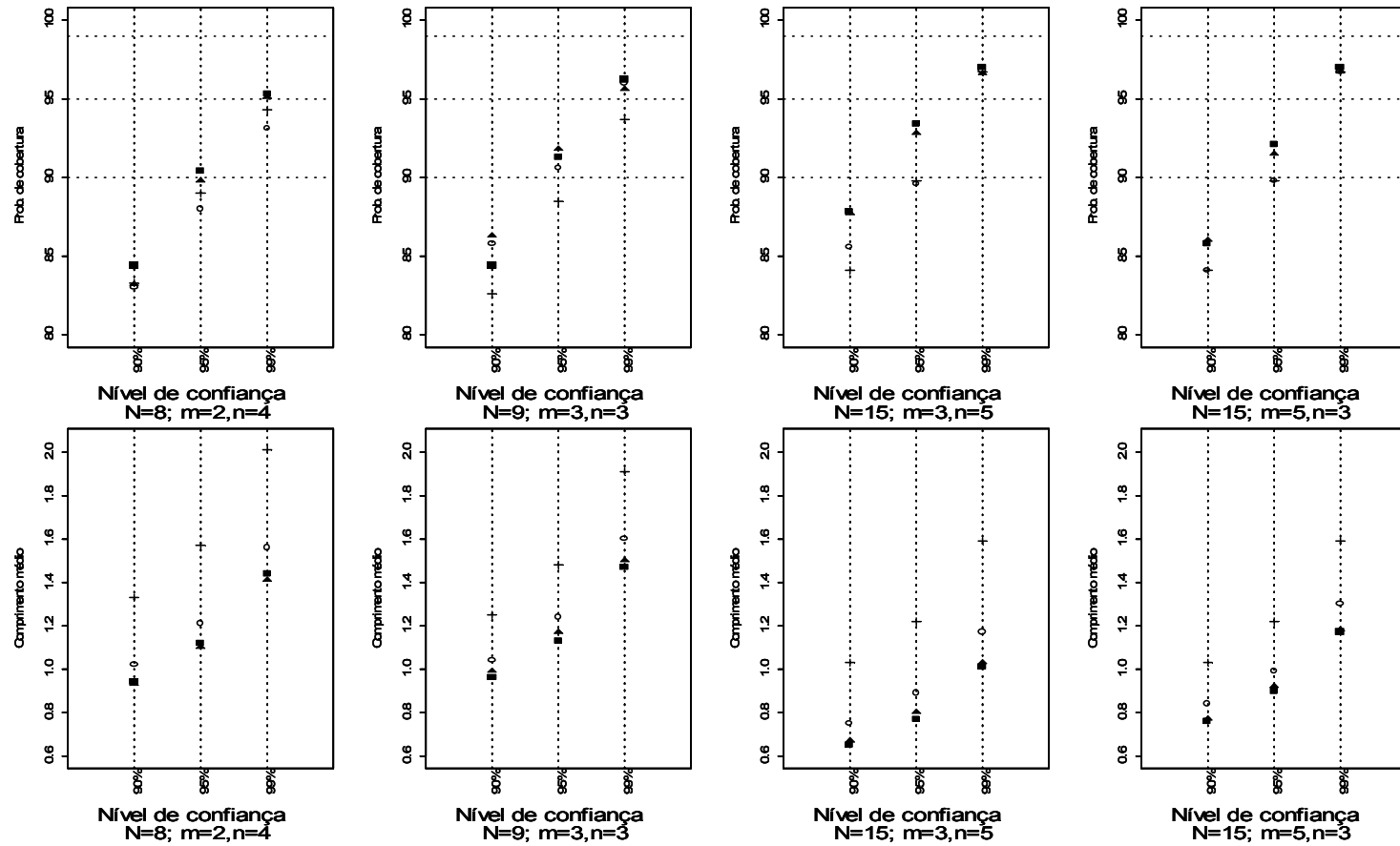
		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	1,45 (1,29)	1,32 (1,01)	1,47 (1,14)	1,75 (1,47)	1,35 (1,01)	1,42 (1,27)	1,50 (1,14)	1,77 (1,54)
	95%	1,73 (1,54)	1,58 (1,21)	1,76 (1,36)	2,09 (1,68)	1,61 (1,21)	1,70 (1,52)	1,79 (1,36)	2,12 (1,84)
	99%	2,28 (2,02)	2,07 (1,59)	2,31 (1,79)	2,75 (2,21)	2,12 (1,59)	2,24 (1,99)	2,36 (1,79)	2,79 (2,42)
Básico	90%	1,38 (1,10)	1,28 (0,91)	1,42 (0,99)	1,74 (1,39)	1,33 (0,99)	1,41 (1,18)	1,49 (1,13)	1,76 (1,52)
	95%	1,71 (1,50)	1,55 (1,17)	1,73 (1,32)	2,00 (1,51)	1,57 (1,14)	1,66 (1,42)	1,75 (1,29)	2,03 (1,64)
	99%	2,07 (1,61)	1,92 (1,31)	2,12 (1,46)	2,55 (1,93)	1,96 (1,32)	2,06 (1,57)	2,17 (1,50)	2,61 (2,14)
Percentil	90%	1,38 (1,10)	1,28 (0,91)	1,42 (0,99)	1,74 (1,39)	1,33 (0,99)	1,41 (1,18)	1,49 (1,13)	1,76 (1,52)
	95%	1,71 (1,50)	1,55 (1,17)	1,73 (1,32)	2,00 (1,51)	1,57 (1,14)	1,66 (1,42)	1,75 (1,29)	2,03 (1,64)
	99%	2,07 (1,61)	1,92 (1,31)	2,12 (1,46)	2,55 (1,93)	1,96 (1,32)	2,06 (1,57)	2,17 (1,50)	2,61 (2,14)
BCa	90%	1,52 (1,28)	1,55 (1,34)	1,51 (1,26)	1,87 (1,52)	1,51 (1,50)	1,54 (1,44)	1,57 (1,26)	2,00 (1,77)
	95%	1,77 (1,44)	1,80 (1,48)	1,76 (1,39)	2,23 (1,78)	1,84 (1,91)	1,86 (1,84)	1,87 (1,50)	2,38 (2,08)
	99%	2,21 (1,74)	2,27 (1,82)	2,20 (1,63)	2,74 (2,10)	2,25 (2,08)	2,30 (2,01)	2,30 (1,72)	2,92 (2,40)
Studentizado	90%	3,19 (8,56)	3,94 (6,62)	3,61 (8,32)	5,95 (17,85)	3,30 (6,34)	2,52 (5,77)	3,63 (8,85)	4,42 (8,87)
	95%	4,67 (11,45)	5,54 (9,21)	5,83 (19,40)	7,32 (20,14)	4,29 (7,58)	3,16 (7,03)	4,66 (12,42)	5,77 (12,93)
	99%	6,43 (14,92)	7,64 (12,88)	8,44 (24,60)	11,07 (26,59)	6,05 (9,97)	4,92 (12,27)	6,90 (17,96)	8,62 (18,44)

Tabela 11 (cont)- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição lognormal.

		<i>N=15; m=3, n=5</i>				<i>N=15; m=5, n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	1,09 (0,76)	1,13 (0,85)	1,21 (0,97)	1,46 (0,87)	1,19 (0,87)	1,24 (0,86)	1,25 (0,75)	1,46 (0,87)
	95%	1,30 (0,90)	1,35 (1,01)	1,45 (1,16)	1,74 (1,04)	1,42 (1,05)	1,48 (1,03)	1,49 (0,89)	1,74 (1,04)
	99%	1,71 (1,19)	1,77 (1,33)	1,91 (1,52)	2,29 (1,37)	1,87 (1,38)	1,94 (1,35)	1,96 (1,18)	2,29 (1,37)
Básico	90%	1,08 (0,74)	1,15 (0,77)	1,19 (0,86)	1,45 (0,86)	1,19 (0,87)	1,23 (0,85)	1,25 (0,75)	1,45 (0,86)
	95%	1,27 (0,85)	1,32 (0,95)	1,42 (1,19)	1,70 (0,96)	1,38 (0,94)	1,44 (0,95)	1,45 (0,83)	1,70 (0,96)
	99%	1,60 (1,02)	1,65 (0,58)	1,78 (1,22)	2,19 (1,24)	1,78 (1,21)	1,84 (1,19)	1,87 (1,06)	2,19 (1,24)
Percentil	90%	1,08 (0,74)	1,15 (0,77)	1,19 (0,86)	1,45 (0,86)	1,19 (0,87)	1,23 (0,85)	1,25 (0,75)	1,45 (0,86)
	95%	1,27 (0,85)	1,32 (0,95)	1,42 (1,19)	1,70 (0,96)	1,38 (0,94)	1,44 (0,95)	1,45 (0,83)	1,70 (0,96)
	99%	1,60 (1,02)	1,65 (0,58)	1,78 (1,22)	2,19 (1,24)	1,78 (1,21)	1,84 (1,19)	1,87 (1,06)	2,19 (1,24)
BCa	90%	1,16 (0,98)	1,15 (0,79)	1,25 (0,89)	1,66 (1,25)	1,31 (1,04)	1,37 (1,10)	1,34 (0,86)	1,66 (1,25)
	95%	1,40 (1,19)	1,38 (0,95)	1,51 (1,12)	2,00 (1,51)	1,58 (1,27)	1,65 (1,30)	1,60 (1,02)	2,00 (1,51)
	99%	1,73 (1,35)	1,72 (1,07)	1,87 (1,24)	2,48 (1,70)	1,95 (1,47)	2,03 (1,47)	1,99 (1,19)	2,48 (1,70)
Studentizado	90%	2,06 (3,82)	2,20 (4,32)	2,07 (3,50)	2,49 (3,30)	2,31 (3,41)	2,23 (3,35)	2,08 (3,46)	2,49 (3,30)
	95%	2,48 (4,43)	2,73 (5,53)	2,54 (4,27)	3,09 (3,32)	2,82 (3,96)	2,71 (4,07)	2,53 (4,02)	3,09 (3,32)
	99%	3,48 (7,54)	3,70 (6,63)	3,43 (5,24)	4,43 (5,75)	3,84 (4,96)	3,80 (5,76)	3,79 (7,99)	4,43 (5,75)

6.6 Estimação da média de uma população com distribuição Gumbell

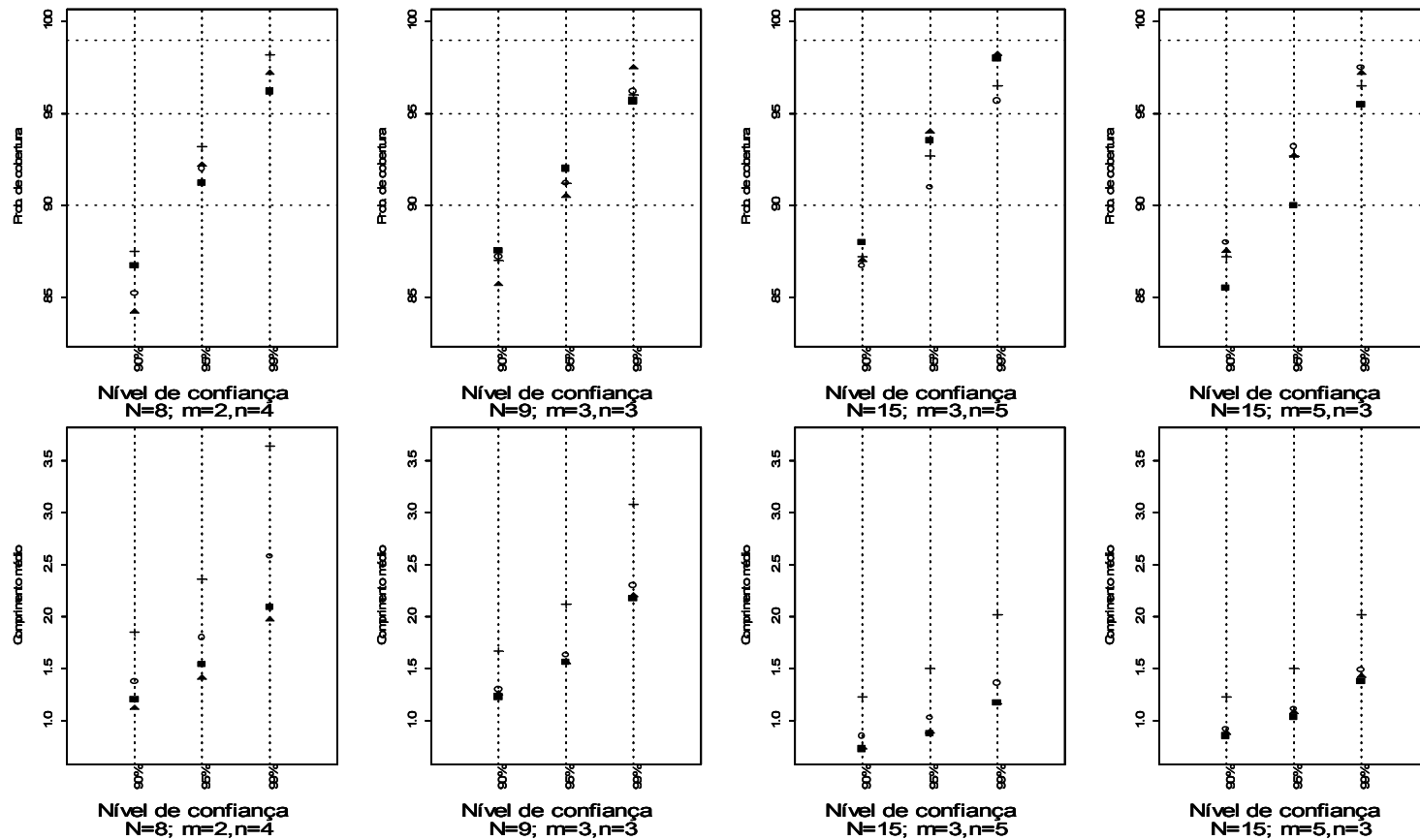
Finalmente, utilizou-se os intervalos propostos na estimação da média de uma variável com distribuição gumbell. Os resultados encontram-se nas Figuras 15 e 16 e nas Tabelas 12 e 13. Os resultados obtidos para esta distribuição apenas ratificam as conclusões anteriores: os estimadores propostos são mais precisos e tão acurados quanto os estimadores convencionais.



(+) *SRS* (■) *RSS* (Ordenação perfeita) (▲) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.05$)

(○) *RSS* (Ordenação imperfeita, $\sigma_e^2 = 0.50$)

Figura 15- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap percentil para a média de uma variável com distribuição Gumbell, via *SRS* e *RSS* com ordenação perfeita e imperfeita.



(+) *SRS* (■) *RSS* (Ordenação perfeita) (▲) *RSS* (Ordenação imperfeita, $\sigma_\epsilon^2 = 0.05$)

(○) *RSS* (Ordenação imperfeita, $\sigma_\epsilon^2 = 0.50$)

Figura 16- Probabilidades de cobertura e comprimentos médios de intervalos de confiança bootstrap studentizado para a média de uma variável com distribuição Gumbell, via *SRS* e *RSS* com ordenação perfeita e imperfeita.

Tabela 12- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição Gumbell.

		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	84,5	84,1	83,0	83,5	84,3	86,8	85,8	82,5
	95%	90,6	90,3	87,9	89,6	91,3	92,3	91,3	88,9
	99%	95,5	95,6	94,0	94,3	96,4	96,4	96,5	94,4
Básico	90%	84,5	83,8	83,3	83,9	84,3	86,8	86,1	82,3
	95%	89,7	90,2	86,8	88,7	91,7	91,8	91,3	87,9
	99%	95,1	94,7	93,3	93,8	95,9	95,9	96,1	94,1
Percentil	90%	84,4	83,2	83,0	83,3	84,4	86,3	85,8	82,6
	95%	90,4	89,8	88,0	89,0	91,3	91,8	90,6	88,5
	99%	95,3	95,1	93,1	94,3	96,2	95,6	96,0	93,7
BCa	90%	84,4	83,0	82,2	80,5	83,8	82,6	83,2	83,4
	95%	90,1	89,6	88,0	86,5	89,5	88,5	88,9	88,7
	99%	94,9	94,7	94,7	92,8	95,5	93,6	94,9	94,1
Studentizado	90%	86,7	84,2	85,2	87,5*	87,5*	85,7	87,2*	87,0*
	95%	91,2	92,2	92,0	93,2*	92,0	90,5	91,2	91,2
	99%	96,2	97,2	96,2	98,2*	95,7	97,5	96,2	96,0

Tabela 12 (cont)- Probabilidade de cobertura de intervalos de confiança bootstrap para a média populacional, obtida através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição Gumbell.

		<i>N=15; m=3, n=5</i>				<i>N=15; m=5, n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	87,9	88,1	85,4	84,1	85,8	85,9	85,9	84,1
	95%	93,3	93,0	90,0	89,8	92,8	91,0	91,4	89,8
	99%	97,0	97,2	96,5	96,7	96,8	96,1	96,7	96,7
Básico	90%	88,5	88,0	85,3	84,1	88,3	85,7	86,4	84,1
	95%	93,6	93,1	90,3	88,8	92,3	90,8	91,4	88,8
	99%	96,9	96,4	96,4	96,1	96,5	95,5	96,2	96,1
Percentil	90%	87,8	87,7	85,6	84,1	85,8	86,1	86,0	84,1
	95%	93,4	92,8	89,6	89,8	92,1	91,2	91,5	89,8
	99%	97,0	96,6	96,8	96,7	97,0	96,1	96,7	96,7
BCa	90%	85,4	84,1	86,7	86,3	87,1	83,8	86,4	86,3
	95%	91,1	89,7	91,5	90,9	92,1	89,7	91,6	90,9
	99%	96,9	95,8	97,0	96,5	97,8	96,2	97,1	96,5
Studentizado	90%	88,0*	87,0	86,7	87,2*	85,5	87,5	88,0	87,2*
	95%	93,5*	94,0*	91,0	92,7*	90,0	92,7	93,2	92,7*
	99%	98,0*	98,2*	95,7	96,5	95,5	97,2	97,5	96,5

Tabela 13- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição Gumbell.

		<i>N=8; m=2,n=4</i>				<i>N=9; m=3,n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS
Normal	90%	0,94 (0,33)	0,93 (0,31)	1,02 (0,34)	1,33 (0,47)	0,95 (0,30)	0,99 (0,33)	1,04 (0,33)	1,25 (0,42)
	95%	1,13 (0,40)	1,11 (0,37)	1,22 (0,41)	1,59 (0,57)	1,14 (0,36)	1,18 (0,39)	1,25 (0,39)	1,50 (0,50)
	99%	1,48 (0,53)	1,46 (0,49)	1,61 (0,54)	2,09 (0,74)	1,50 (0,48)	1,55 (0,52)	1,64 (0,52)	1,97 (0,66)
Básico	90%	0,94 (0,33)	0,93 (0,30)	1,02 (0,33)	1,33 (0,47)	0,96 (0,30)	0,99 (0,33)	1,04 (0,32)	1,25 (0,42)
	95%	1,12 (0,39)	1,10 (0,37)	1,21 (0,40)	1,57 (0,55)	1,13 (0,36)	1,17 (0,38)	1,24 (0,38)	1,48 (0,49)
	99%	1,44 (0,49)	1,41 (0,45)	1,56 (0,50)	2,01 (0,70)	1,47 (0,45)	1,50 (0,48)	1,60 (0,48)	1,91 (0,62)
Percentil	90%	0,94 (0,33)	0,93 (0,30)	1,02 (0,33)	1,33 (0,47)	0,96 (0,30)	0,99 (0,33)	1,04 (0,32)	1,25 (0,42)
	95%	1,12 (0,39)	1,10 (0,37)	1,21 (0,40)	1,57 (0,55)	1,13 (0,36)	1,17 (0,38)	1,24 (0,38)	1,48 (0,49)
	99%	1,44 (0,49)	1,41 (0,45)	1,56 (0,50)	2,01 (0,70)	1,47 (0,45)	1,50 (0,48)	1,60 (0,48)	1,91 (0,62)
BCa	90%	0,96 (0,34)	0,98 (0,35)	1,02 (0,34)	1,32 (0,48)	1,00 (0,36)	0,99 (0,34)	1,07 (0,36)	1,31 (0,47)
	95%	1,14 (0,40)	1,17 (0,41)	1,21 (0,40)	1,57 (0,58)	1,19 (0,42)	1,17 (0,40)	1,27 (0,42)	1,56 (0,56)
	99%	1,46 (0,50)	1,50 (0,53)	1,55 (0,50)	2,00 (0,73)	1,53 (0,54)	1,51 (0,51)	1,64 (0,55)	1,99 (0,70)
Studentizado	90%	1,20 (0,71)	1,12 (0,49)	1,38 (0,93)	1,85 (0,93)	1,23 (0,58)	1,26 (0,64)	1,30 (0,69)	1,67 (0,81)
	95%	1,54 (1,02)	1,41 (0,66)	1,80 (1,74)	2,36 (1,30)	1,56 (0,79)	1,56 (0,83)	1,63 (0,86)	2,12 (1,19)
	99%	2,09 (1,33)	1,97 (1,03)	2,58 (2,28)	3,64 (2,42)	2,17 (1,20)	2,20 (1,33)	2,30 (1,31)	3,08 (1,82)

Tabela 13 (cont)- Comprimentos médios e desvio padrão dos comprimentos (entre parêntesis) de intervalos de confiança bootstrap para a média populacional, obtidos através de simulação, via amostragem aleatória simples (SRS) e amostragem por conjuntos ordenados (RSS), considerando ordenação perfeita e imperfeita, de uma população com distribuição Gumbell.

		<i>N=15; m=3, n=5</i>					<i>N=15; m=5, n=3</i>			
	Confiança	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	RSS O.P.	RSS- O.I. ($\sigma_e^2 = 0,05$)	RSS- O.I. ($\sigma_e^2 = 0,50$)	SRS	
Normal	90%	0,65 (0,16)	0,67 (0,17)	0,75 (0,17)	1,02 (0,27)	0,75 (0,18)	0,77 (0,20)	0,84 (0,20)	1,02 (0,27)	
	95%	0,78 (0,20)	0,81 (0,20)	0,90 (0,21)	1,23 (0,32)	0,90 (0,22)	0,92 (0,23)	1,00 (0,24)	1,23 (0,32)	
	99%	1,02 (0,26)	1,06 (0,27)	1,18 (0,27)	1,61 (0,42)	1,19 (0,29)	1,22 (0,31)	1,31 (0,32)	1,61 (0,42)	
Básico	90%	0,65 (0,16)	0,67 (0,17)	0,75 (0,17)	1,03 (0,27)	0,76 (0,18)	0,77 (0,20)	0,84 (0,20)	1,03 (0,27)	
	95%	0,77 (0,19)	0,80 (0,20)	0,89 (0,20)	1,22 (0,32)	0,90 (0,22)	0,92 (0,23)	0,99 (0,24)	1,22 (0,32)	
	99%	1,01 (0,24)	1,03 (0,25)	1,17 (0,26)	1,59 (0,41)	1,17 (0,28)	1,18 (0,30)	1,30 (0,31)	1,59 (0,41)	
Percentil	90%	0,65 (0,16)	0,67 (0,17)	0,75 (0,17)	1,03 (0,27)	0,76 (0,18)	0,77 (0,20)	0,84 (0,20)	1,03 (0,27)	
	95%	0,77 (0,19)	0,80 (0,20)	0,89 (0,20)	1,22 (0,32)	0,90 (0,22)	0,92 (0,23)	0,99 (0,24)	1,22 (0,32)	
	99%	1,01 (0,24)	1,03 (0,25)	1,17 (0,26)	1,59 (0,41)	1,17 (0,28)	1,18 (0,30)	1,30 (0,31)	1,59 (0,41)	
BCa	90%	0,65 (0,18)	0,69 (0,20)	0,76 (0,18)	1,04 (0,30)	0,78 (0,22)	0,79 (0,22)	0,84 (0,21)	1,04 (0,30)	
	95%	0,78 (0,22)	0,82 (0,25)	0,90 (0,22)	1,24 (0,36)	0,93 (0,26)	0,94 (0,26)	1,01 (0,26)	1,24 (0,36)	
	99%	1,01 (0,29)	1,07 (0,32)	1,17 (0,28)	1,61 (0,47)	1,20 (0,33)	1,21 (0,34)	1,30 (0,33)	1,61 (0,47)	
Studentizado	90%	0,73 (0,36)	0,74 (0,28)	0,85 (0,30)	1,23 (0,47)	0,85 (0,29)	0,88 (0,34)	0,92 (0,31)	1,23 (0,47)	
	95%	0,88 (0,43)	0,89 (0,34)	1,03 (0,36)	1,50 (0,58)	1,03 (0,36)	1,08 (0,42)	1,11 (0,38)	1,50 (0,58)	
	99%	1,17 (0,56)	1,17 (0,45)	1,36 (0,49)	2,02 (0,78)	1,38 (0,53)	1,43 (0,56)	1,49 (0,55)	2,02 (0,78)	

Capítulo 7

Conclusão e propostas futuras

Realizou-se, neste trabalho, um amplo estudo via simulação buscando avaliar a performance de diferentes intervalos de confiança não paramétricos, determinados mediante a metodologia bootstrap, via *RSS*, na estimação da média populacional. A alteração do delineamento amostral na obtenção de reamostras bootstrap tem como motivação a possibilidade de determinação de estimativas não paramétricas mais precisas considerando tamanhos de amostras reduzidos.

Em determinados levantamentos amostrais, algumas dificuldades na obtenção das medidas de interesse (como, por exemplo, custos elevados) inviabilizam a seleção de uma grande amostra. Dada a imprecisão de estimativas originárias de pequenas amostras, deve-se atentar à necessidade da utilização de um esquema amostral capaz de contornar tal dificuldade. Nesta situação, a amostragem por conjuntos ordenados, em particular, é uma alternativa altamente eficaz, uma vez que é responsável pela produção de estimativas mais precisas frente a restrições na seleção de amostras numerosas.

Desta forma, a conjunção de um procedimento não paramétrico de reamostragem e um delineamento amostral adequado para amostras reduzidas que, baseado em estudos anteriores, é capaz de produzir estimadores significativamente mais precisos, pode ser de extrema relevância e notável aplicabilidade para um grande número de estudos que envolvem amostragem.

Verificou-se, inicialmente, que os intervalos de confiança obtidos através dos esquemas de reamostragem propostos neste trabalho apresentam uma forte similaridade com os intervalos de confiança bootstrap baseados em amostragem aleatória simples no que se refere a acurácia, se destacando, no entanto, por um acentuado ganho em termos de precisão das estimativas geradas. Este ganho decresce à medida que aumenta o grau de imperfeição na ordenação das unidades

amostrais, embora haja um relevante ganho em precisão ainda na situação mais crítica, em que se considerou o maior valor para a variância dos erros de ordenação. Isto foi constatado para as seis diferentes distribuições sob análise. Pôde-se observar também que os intervalos de confiança bootstrap não paramétricos, para ambos os delineamentos amostrais considerados, apresentam probabilidades de cobertura distantes do nível de confiança estabelecido, tendência esta que é acentuada no procedimento de inferência da média de distribuições assimétricas. Os intervalos studentizados, no entanto, produziram estimativas mais acuradas na estimação da média para todas as distribuições estudadas. Os desempenhos dos outros quatro intervalos de confiança bootstrap são bastante análogos, não apresentando diferenças significativas quanto a acurácia e precisão.

Ratificada a viabilidade do bootstrap baseado em amostragem por conjuntos ordenados, são apresentadas na seqüência algumas sugestões a serem estudadas futuramente. Uma primeira proposta seria incorporar às estimativas uma estrutura de custos, visando otimizar a relação custo/precisão e, baseado nesta análise, definir a conveniência da utilização dos intervalos propostos sob circunstâncias diversas.

Poderia-se também buscar alternativas que possibilitassem atenuar os problemas verificados quanto à acurácia dos intervalos de confiança bootstrap não paramétricos.

Outra proposta seria estender a aplicação dos intervalos de confiança não paramétricos bootstrap via *RSS* à estimação de outros parâmetros, como a variância e quantis populacionais, verificando o impacto, em termos de precisão, da utilização do referido delineamento amostral em contraposição a *SRS*.

Apêndice

Apêndice A. Programa para gerar intervalos de confiança bootstrap não paramétricos dos tipos normal, básico e percentil, via RSS com ordenação perfeita, para a média de MP10 na estação Ibirapuera em 2000/2001

```
ar.data<-read.table("a:\\dados.txt",col.names="med.po")
# ar.data armazena os dados
repl<-# número de replicações (m)
amst<-# tamanho da amostra (n)
vetor<-matrix(0,amst,amst)
vetor2<-matrix(0,amst,amst)
ams<-rep(0,amst*repl)
ams2<-rep(0,amst*repl)
PC<- # número de intervalos de confiança gerados
B<- # número de re-amostras geradas na obtenção de cada intervalo de confiança.
param<-mean(ar.data[, "méd.po"]) # média populacional

# contadores utilizados no cálculo das probabilidades de cobertura:
bnormal.cob90<-0
bnormal.cob95<-0
bnormal.cob99<-0
percent.cob90<-0
percent.cob95<-0
percent.cob99<-0
```

```
basic.cob90<-0
```

```
basic.cob95<-0
```

```
basic.cob99<-0
```

```
# vetores utilizados na obtenção dos comprimentos dos intervalos
```

```
comp.bnnormal90<-rep(0,PC)
```

```
comp.bnnormal95<-rep(0,PC)
```

```
comp.bnnormal99<-rep(0,PC)
```

```
comp.percent90<-rep(0,PC)
```

```
comp.percent95<-rep(0,PC)
```

```
comp.percent99<-rep(0,PC)
```

```
comp.basic90<-rep(0,PC)
```

```
comp.basic95<-rep(0,PC)
```

```
comp.basic99<-rep(0,PC)
```

```
lim.inf90bnormal<-rep(0,PC)
```

```
lim.sup90bnormal<-rep(0,PC)
```

```
lim.inf95bnormal<-rep(0,PC)
```

```
lim.sup95bnormal<-rep(0,PC)
```

```
lim.inf99bnormal<-rep(0,PC)
```

```
lim.sup99bnormal<-rep(0,PC)
```

```
lim.inf90percent<-rep(0,PC)
```

```
lim.sup90percent<-rep(0,PC)
```

```
lim.inf95percent<-rep(0,PC)
```

```
lim.sup95percent<-rep(0,PC)
```

```
lim.inf99percent<-rep(0,PC)
```

```
lim.sup99percent<-rep(0,PC)
```

```
lim.inf90basic<-rep(0,PC)
```

```
lim.sup90basic<-rep(0,PC)
```

```
lim.inf95basic<-rep(0,PC)
```

```
lim.sup95basic<-rep(0,PC)
```

```

lim.inf99basic<-rep(0,PC)
lim.sup99basic<-rep(0,PC)
logar.data<-ar.data[,"med.po"]

# extração da amostra por conjuntos ordenados:
for(j in 1:PC){

  for (t in 1:repl){
    for (p in 1:amst){
      vetor[,p]<-sample(logar.data,amst)
      vetor2[,p]<-sort(vetor[,p])
      ams[t*p+(t-1)*(amst-p)]<-vetor2[p,p]
    }
  }

# extração de re-amostras bootstrap por conjuntos ordenados
x.star<-matrix(0,amst*repl,B)
teta.star<-rep(0,B)
for (i in 1:B){
  for (t in 1:repl){
    for (p in 1:amst){
      vetor[,p]<-sample(ams,amst,replace=T)
      vetor2[,p]<-sort(vetor[,p])
      ams2[t*p+(t-1)*(amst-p)]<-vetor2[p,p]
    }
  }
  x.star[,i]<-ams2
  teta.star[i]<-mean(x.star[,i])
}

```

```

# bootstrap normal
teta.star<-sort(teta.star)
vicio.boot<-sum(teta.star-mean(ams))*1/B
var.boot<-sum((teta.star-mean(teta.star))^2)*1/(B-1)
lim.inf95bnormal[j]<-mean(ams)-vicio.boot-1.96*sqrt(var.boot)
lim.sup95bnormal[j]<-mean(ams)-vicio.boot+1.96*sqrt(var.boot)

lim.inf90bnormal[j]<-mean(ams)-vicio.boot-1.64*sqrt(var.boot)
lim.sup90bnormal[j]<-mean(ams)-vicio.boot+1.64*sqrt(var.boot)

lim.inf99bnormal[j]<-mean(ams)-vicio.boot-2.575*sqrt(var.boot)
lim.sup99bnormal[j]<-mean(ams)-vicio.boot+2.575*sqrt(var.boot)

# cálculo do número de intervalos que contém o parâmetro.

if ((lim.sup95bnormal[j]>=param)&&
    (lim.inf95bnormal[j]<=param))
  bnormal.cob95<-bnormal.cob95+1

if ((lim.sup90bnormal[j]>=param)&&
    (lim.inf90bnormal[j]<=param))
  bnormal.cob90<-bnormal.cob90+1

if ((lim.sup99bnormal[j]>=param)&&
    (lim.inf99bnormal[j]<=param))
  bnormal.cob99<-bnormal.cob99+1

# cálculo dos comprimentos dos intervalos.
comp.bnornal95[j]<-lim.sup95bnormal[j]-lim.inf95bnormal[j]
comp.bnornal90[j]<-lim.sup90bnormal[j]-lim.inf90bnormal[j]
comp.bnornal99[j]<-lim.sup99bnormal[j]-lim.inf99bnormal[j]

```

```

# bootstrap percentil

# cálculo do número de intervalos que contém o parâmetro.

if (((quantile(teta.star,0.025))<=param)&&
     quantile(teta.star,0.975))>=param))
percent.cob95<-percent.cob95+1
if (((quantile(teta.star,0.050))<=param)&&
     ((quantile(teta.star,0.950))>=param))
percent.cob90<-percent.cob90+1
if (((quantile(teta.star,0.005))<=param)&&
     ((quantile(teta.star,0.995))>=param))
percent.cob99<-percent.cob99+1

# cálculo dos comprimentos dos intervalos.

comp.percent90[j]<- quantile(teta.star,0.950)- quantile(teta.star,0.050)
comp.percent95[j]<- quantile(teta.star,0.975)- quantile(teta.star,0.025)
comp.percent99[j]<- quantile(teta.star,0.995)- quantile(teta.star,0.005)

# bootstrap básico

# cálculo dos limites dos intervalos
lim.inf90basic[j]<-(2*mean(ams)-( quantile(teta.star,0.950)))
lim.sup90basic[j]<-(2*mean(ams)-( quantile(teta.star,0.050)))

lim.inf95basic[j]<-(2*mean(ams)-( quantile(teta.star,0.975)))
lim.sup95basic[j]<-(2*mean(ams)-( quantile(teta.star,0.025)))

lim.inf99basic[j]<-(2*mean(ams)-( quantile(teta.star,0.995)))
lim.sup99basic[j]<-(2*mean(ams)-( quantile(teta.star,0.005)))

```

```
# cálculo do número de intervalos que contém o parâmetro.
```

```
if (lim.sup95basic[j]>=param&&  
    lim.inf95basic[j]<=param)  
basic.cob95<-basic.cob95+1
```

```
if (lim.sup90basic[j]>=param&&  
    lim.inf90basic[j]<=param)  
basic.cob90<-basic.cob90+1
```

```
if (lim.sup99basic[j]>=param&  
    lim.inf99basic[j]<=param)  
basic.cob99<-basic.cob99+1
```

```
# cálculo dos comprimentos dos intervalos.
```

```
comp.basic90[j]<-lim.sup90basic[j]-lim.inf90basic[j]  
comp.basic95[j]<-lim.sup95basic[j]-lim.inf95basic[j]  
comp.basic99[j]<-lim.sup99basic[j]-lim.inf99basic[j]  
}
```

```
# resultados
```

```
bnormal.cob90/PC  
mean(comp.bnormal90)  
sqrt(var(comp.bnormal90))  
bnormal.cob95/PC  
mean(comp.bnormal95)  
sqrt(var(comp.bnormal95))  
bnormal.cob99/PC  
mean(comp.bnormal99)  
sqrt(var(comp.bnormal99))
```

```

basic.cob90/PC
mean(comp.basic90)
sqrt(var(comp.basic90))
basic.cob95/PC
mean(comp.basic95)
sqrt(var(comp.basic95))
basic.cob99/PC
mean(comp.basic99)
sqrt(var(comp.basic99))
percent.cob90/PC
mean(comp.percent90)
sqrt(var(comp.percent90))
percent.cob95/PC
mean(comp.percent95)
sqrt(var(comp.percent95))
percent.cob99/PC
mean(comp.percent99)
sqrt(var(comp.percent99))

```

Apêndice B. Programa para gerar intervalos de confiança bootstrap não paramétricos do tipo BCa, via RSS com ordenação imperfeita, para a média de uma variável com distribuição normal (0,1)

```

# contadores utilizados no cálculo das probabilidades de cobertura:
bca.cob95<-0
bca.cob90<-0
bca.cob99<-0

PC<- # número de intervalos de confiança gerados.
B<- # número de re-amostras geradas para a obtenção de cada intervalo.

```

vetores que armazenam limites e comprimentos dos intervalos de confiança:

```
tetamin95<-rep(0,PC)
```

```
tetamax95<-rep(0,PC)
```

```
comp.bca95<-rep(0,PC)
```

```
tetamin90<-rep(0,PC)
```

```
tetamax90<-rep(0,PC)
```

```
comp.bca90<-rep(0,PC)
```

```
tetamin99<-rep(0,PC)
```

```
tetamax99<-rep(0,PC)
```

```
comp.bca99<-rep(0,PC)
```

```
param<-0
```

```
repl<- # número de replicações
```

```
amst<- # tamanho de amostra
```

```
s2e<- # variância do erro aleatório responsável pela imperfeição na ordenação das amostras.
```

```
ams<-rep(0,amst*repl)
```

```
ams3<-rep(0,amst*repl)
```

```
vetor<-matrix(0,amst,amst)
```

```
vetor2<-matrix(0,amst,amst)
```

```
vetmed<-rep(0,amst*repl-1)
```

```
jack<-rep(0,amst*repl-1)
```

```
for(j in 1:PC){
```

```
vetorreal<-numeric()
```

```
vetorjulgado<-numeric()
```



```

# geração de uma amostra por conjuntos ordenados com ordenação imperfeita
for (k in 1:repl){
  julgado<-numeric()
  real<-numeric()
  xx<-matrix(0,amst,amst)
  x<-matrix(0,amst,amst)
  # x armazena os valores reais das unidades amostrais selecionadas
  e<-matrix(0,amst,amst)
  # e armazena os erros aleatórios com distribuição normal, média 0 e variância s2e
  y<-matrix(0,amst,amst)
  # y armazena a soma dos valores de “x” e “e”
  w<-matrix(0,amst,amst)
  for (i in 1:amst){
    for (t in 1:amst){
      x[i,t]<-rnorm(1,0,1)
      e[i,t]<-rnorm(1,0,s2e)
      y[i,t]<-x[i,t]+e[i,t]
    }
    for (t in 1:amst){
      w[,t]<-sort(y[,t])
    }
    julgado<-diag(w)
    for (i in 1:amst){
      for (t in 1:amst){
        if(y[t,i]==w[i,i])
          real[i]<-x[t,i]}
    }

vetorjulgado<-c(vetorjulgado,julgado)
vetorreal<-c(vetorreal,real)
#vetorreal armazena os valores reais da amostra por conjuntos ordenados com erros de ordenação

```

```

#vetorjulgado armazena os valores julgados (acrescidos do erro) da amostra por conjuntos
ordenados com erros de ordenação
}

ams<-vetorreal
# cálculo da função influência via Jackknife
for (p in 1:(amst*repl)){
  c1<-1
  for (k in 1:(amst*repl)){

    if (p!=k){
      jack[c1]<-ams[k]
      c1<-c1+1}

    }
  vetmed[p]<-mean(jack)

}
n<-seq(1,amst*repl)
# ams é a amostra, que será re-amostrada

a<-(1/6)*sum((mean(vetmed)-vetmed)^3)/((sum((mean(vetmed)-vetmed)^2))^1.5)
contador<-0
med<-rep(0,B)
#med vai armazenar os estimadores prov. das amostras bootstrapeadas.

# geração das re-amostras bootstrap
for(u in 1:B){

```

```

vetorreal2<-numeric()
vetorjulgado2<-numeric()

for (k in 1:repl){
  julgado2<-numeric()
  real2<-numeric()
  x1<-matrix(0,amst,amst)
  xx<-matrix(0,amst,amst)
  y1<-matrix(0,amst,amst)
  w1<-matrix(0,amst,amst)

  for (i in 1:amst){
    for (t in 1:amst){
      xx[t,i]<-sample(n,1)
      y1[t,i]<-vetorjulgado[xx[t,i]]
      x1[t,i]<-vetorreal[xx[t,i]]
    }

    for (t in 1:amst){
      w1[,t]<-sort(y1[,t])}
    julgado2<-diag(w1)

    for (i in 1:amst){
      for (t in 1:amst){
        if(y1[t,i]==w1[i,i])
          real2[i]<-x1[t,i]}
    }

vetorjulgado2<-c(vetorjulgado2,julgado2)
vetorreal2<-c(vetorreal2,real2)
}

med[u]<-mean(vetorreal2)

```

```

    if (med[u] <= mean(ams))
      contador <- contador + 1
  }

w <- qnorm(contador/R, mean=0, sd=1)
zalfa195 <- -1.96 + w
zalfa295 <- -1.96 + w
zalfa190 <- -1.64 + w
zalfa290 <- -1.64 + w
zalfa199 <- -2.575 + w
zalfa299 <- -2.575 + w
alfatio195 <- pnorm(w + (zalfa195)/(1 - a*(zalfa195)), 0, 1)
alfatio295 <- pnorm(w + (zalfa295)/(1 - a*(zalfa295)), 0, 1)
alfatio190 <- pnorm(w + (zalfa190)/(1 - a*(zalfa190)), 0, 1)
alfatio290 <- pnorm(w + (zalfa290)/(1 - a*(zalfa290)), 0, 1)
alfatio199 <- pnorm(w + (zalfa199)/(1 - a*(zalfa199)), 0, 1)
alfatio299 <- pnorm(w + (zalfa299)/(1 - a*(zalfa299)), 0, 1)

# cálculo dos comprimentos dos intervalos e número de intervalos que contém o valor do
# parâmetro

tetamin90[j] <- quantile(med, alfatio190)
tetamax90[j] <- quantile(med, alfatio290)
comp.bca90[j] <- tetamax90[j] - tetamin90[j]
if (tetamax90[j] >= param &&
    tetamin90[j] <= param)
  bca.cob90 <- bca.cob90 + 1

tetamin95[j] <- quantile(med, alfatio195)
tetamax95[j] <- quantile(med, alfatio295)
comp.bca95[j] <- tetamax95[j] - tetamin95[j]

```

```

if (tetamax95[j]>=param&&
    tetamin95[j]<=param)
bca.cob95<-bca.cob95+1

tetamin99[j]<-quantile(med,alfatio199)
tetamax99[j]<-quantile(med,alfatio299)
comp.bca99[j]<-tetamax99[j]-tetamin99[j]
if (tetamax99[j]>=param&&
    tetamin99[j]<=param)
bca.cob99<-bca.cob99+1

cat("j=",j,"bca.cob95=",bca.cob95,"\n")
}

# resultados
bca.cob90/PC
mean(comp.bca90)
sqrt(var(comp.bca90))
bca.cob95/PC
mean(comp.bca95)
sqrt(var(comp.bca95))
bca.cob99/PC
mean(comp.bca99)
sqrt(var(comp.bca99))

```

Apêndice C. Programa para gerar intervalos de confiança bootstrap não paramétricos do tipo studentizado, via RSS com ordenação perfeita, para a média de uma variável com distribuição normal (0,1)

PC<- # número de intervalos de confiança gerados

```
B<- # número de re-amostras gerada para a obtenção de cada intervalo de confiança
R<- # número de re-amostras geradas a partir de cada re-amostra (duplo bootstrap)
```

```
#contador do numero de intervalos que contém o parâmetro.
```

```
t.cob95<-0
```

```
t.cob90<-0
```

```
t.cob99<-0
```

```
repl<- # número de replicações
```

```
amst<- # tamanho de amostra
```

```
# vetores que armazenam os comprimentos dos intervalos
```

```
comp90<-rep(0,PC)
```

```
comp95<-rep(0,PC)
```

```
comp99<-rep(0,PC)
```

```
ams<-rep(0,repl*amst)
```

```
ams4<-rep(0,repl*amst)
```

```
ams1<-rep(0,repl*amst)
```

```
vetor<-matrix(0,amst,amst)
```

```
vetor2<-matrix(0,amst,amst)
```

```
param<-sqrt(exp(1))
```

```
# vetores que armazenam os limites dos intervalos
```

```
lim.inf90<-rep(0,PC)
```

```
lim.sup90<-rep(0,PC)
```

```
lim.inf95<-rep(0,PC)
```

```
lim.sup95<-rep(0,PC)
```

```
lim.inf99<-rep(0,PC)
```

```
lim.sup99<-rep(0,PC)
```

```

for(j in 1:PC){
  # geração de uma amostra por conjuntos ordenados com ordenação perfeita

  for (t in 1:repl){
    for (p in 1:amst){
      vetor[,p]<-rlnorm(1,0,1)
      vetor2[,p]<-sort(vetor[,p])
      ams[t*p+(t-1)*(amst-p)]<-vetor2[p,p]
    }
  }
  teta.chap<-mean(ams)
  z.star<-rep(0,B)

  x.star<-matrix(0,repl*amst,B)
  #x.star vai armazenar as re-amostras bootstrap
  teta.star<-rep(0,B)
  #teta.star vai armazenar os estimadores provenientes das re-amostras bootstrap
  .
  v.star<-rep(0,B)

  # geração das re-amostras
  for(i in 1:B){
    for (t in 1:repl){
      for (p in 1:amst){
        vetor[,p]<-sample(ams,amst,replace=T)
        vetor2[,p]<-sort(vetor[,p])
        ams1[t*p+(t-1)*(amst-p)]<-vetor2[p,p]
      }
    }

    x.star[,i]<-ams1
  }
}

```

```

teta.star[i]<-mean(x.star[,i])

# geração de re-amostras a partir de cada re-amostra (duplo bootstrap)

x.mstar<-matrix(0,repl*amst,R)
teta.mstar<-rep(0,R)
for (k in 1:R){

    for (t in 1:repl){
    for (p in 1:amst){
    vetor[,p]<-sample(ams1,amst,replace=T)
    vetor2[,p]<-sort(vetor[,p])
    ams4[t*p+(t-1)*(amst-p)]<-vetor2[p,p]
    }
    }
    x.mstar[,k]<-ams4
    teta.mstar[k]<-mean(x.mstar[,k])
}
v.star[i]<-(1/(R-1))*sum((teta.mstar-mean(teta.mstar))^2)
z.star[i]<-(teta.star[i]-teta.chap)/sqrt(v.star[i])
}
z.star<-sort(z.star)
vezinho<-(1/(B-1))*sum((teta.star-mean(teta.star))^2)

#calculo dos limites dos intervalos.
lim.inf90[j]<-((teta.chap-quantile(z.star,0.95)*sqrt(vezinho)))
lim.sup90[j]<-((teta.chap-quantile(z.star,0.05)*sqrt(vezinho)))
lim.inf95[j]<-((teta.chap-quantile(z.star,0.975)*sqrt(vezinho)))
lim.sup95[j]<-((teta.chap-quantile(z.star,0.025)*sqrt(vezinho)))
lim.inf99[j]<-((teta.chap-quantile(z.star,0.995)*sqrt(vezinho)))
lim.sup99[j]<-((teta.chap-quantile(z.star,0.005)*sqrt(vezinho)))

```



```

#calculo dos comprimentos dos intervalos.
comp90[j]<-lim.sup90[j]-lim.inf90[j]
comp95[j]<-lim.sup95[j]-lim.inf95[j]
comp99[j]<-lim.sup99[j]-lim.inf99[j]

# cálculo do número de intervalos que contém o parâmetro
if (lim.sup90[j]>=param&&
    lim.inf90[j]<=param)
  t.cob90<-t.cob90+1
if (lim.sup95[j]>=param&&
    lim.inf95[j]<=param)
  t.cob95<-t.cob95+1
if (lim.sup99[j]>=param&&
    lim.inf99[j]<=param)
  t.cob99<-t.cob99+1
  cat("j = ", j,"t.cob90=",t.cob90,"t.cob95=",t.cob95,"t.cob99=",t.cob99, "\n")
}
# resultados
t.cob90/PC
mean(comp90)
sqrt(var(comp90))
t.cob95/PC
mean(comp95)
sqrt(var(comp95))
t.cob99/PC
mean(comp99)
sqrt(var(comp99))

```

Referências

BARNETT, V.; MOORE, K. Best linear unbiased estimates in ranked set sampling with particular reference to imperfect ordering. **J. Appl. Stat.**, 24, p.697-710, 1997.

BARNETT, V.; BROWN, M. Best Linear unbiased quantile estimators for environmental standards. **Environmetrics**, 13, p.295-310, 2002.

BARNETT, V. Ranked set sampling design for environmental investigations. **Environ. Ecol. Stat.**, v.6, p. 59-74, 1999.

BRANDÃO, J.S. Intervalos de confiança assintóticos para o parâmetro de locação em distribuições assimétricas na família locação-escala usando amostragem em conjuntos ordenados. **Relatório de Iniciação Científica**. Des – Universidade Federal de São Carlos. FAPESP, 2003.

CANTY, A. The bootstrap and confidence intervals. Disponível em www.mathstat.concordia.ca/canty/teaching/mast679t.html. Acesso em 25 ago 2004.

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.**, 21, p. 7-20, 2003.

CHEN Z. Density estimation using ranked-set sampling data. **Environ. and Ecol. Stat.**, 6, p.135-146, 1999.

CHEN Z. On ranked-set sampling quantiles and their applications. **J. Stat. Plann. Inf.**, 83, p.125-135, 2000.

DAVID, H. **Order statistics, second edition**, Wiley, 1981, 360p.

DAVISON, A.C.; HINKLEY, D.V. **Bootstrap methods and their application**. New York: Cambridge University Press, 1997. 582p.

DELL, T.R.; CUTTLER, J.L. Ranked set sampling theory with order statistics background. **Biometrics**, 28, p.545-555, 1972.

EFRON, B. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, p.1-26, 1979.

EFRON, B.; TIBSHIRANI, R.J. **An Introduction to the Bootstrap**. New 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**, 127f. Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, São Carlos, 2004.

HASSANEIN, K. M.; SALEH, A. K. M. E.; BROWN, E. F. Best linear unbiased estimators for normal distribution quantiles for sample sizes up to 20. **IEEE Transactions on Reliability**, 3, p.327-329, 1986.

IHAKA, R.; GENTLEMAN, R. A language for data analysis and graphics. **Journal of Computational and Graphical Statistics**, 5, p. 299-314, 1996.

KAUR, A., PATIL, G.P., TAILLIE, C. Unequal allocation models for ranked set sampling with skew distributions, **Biometrics**, 53, p. 123-130, 1997.

MARTINEZ, E.Z.; LOUZADA-NETO, F. **Uma introdução à estimativa intervalar via bootstrap**. Relatório técnico do DES/UFSCar, série A: teoria e métodos, set 2000, 37p.

MCINTYRE, G.A. A method for unbiased selective sampling, using ranked sets. **Australian Journal of Agricultural Research**, 3, p.385-390. 1952.

MUTTLAK, H. A. Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. **Envirometrics**, p.255-267. 1998.

NAHHAS, R.W., WOLFE, D.A., CHEN, H. Ranked set sampling: cost and optimal set size. **Biometrics**, 58, 964-971, 2003.

PEARSON, E.S.; HARTLEY, H.O. **Biometrika tables for Stat.**, 2. London: Griffin 1976.

PIGATTO, F.J.; BARRETO, M.C.M. Desempenho de estimadores da média populacional de distribuições assimétricas baseadas em amostragem por conjuntos ordenados. **Rev. Mat. Estat.**, 21, p. 19-29, 2004.

RAMINELLI, J. A. **Procedimentos de estimação intervalar de um quantil populacional em amostragem por conjuntos ordenados**. 81f. Dissertação (Mestrado em Estatística) – Universidade Federal de São Carlos, São Carlos, 2003.

REVISTA MEIO AMBIENTE INDUSTRIAL. Ano IV ed 23, n.º 22, 2000.

STOKES, S.L. Ranked set sampling with concomitant variables. **Commun. Statist. – Theory & Meth.**, 36, p. 35-42, 1977.

STOKES, S.L. Estimation of variance using judgment ordered ranked set sampling. **Biometrics**, 36, 35-42, 1980.

STOKES, S.L., SAGER, T.W. Characterization of a ranked-set sample with application to estimating distribution functions. **J. Am. Stat. Assoc.**, 83, p. 374-381, 1988.

STOKES, L. Parametric ranked set sampling. **Ann. Inst. Stat. Math.**, 47, p. 465-82, 1995.

TACONELI, C.A.; BARRETO, M.C.M. Intervalos de confiança para a média populacional usando amostragem em conjuntos ordenados. **Rev. Mat. Estat.**, 21, p. 41-66, 2003.

TAKAHASHI, K., WAKIMOTO, K. On biased estimates of population mean based on sample stratified by means of ordering. **Ann. Inst. Stat. Math.**, v.20, p.1-31, 1968.

YU, P. L. H.; LAM, K.; SINHA, B.K. Estimation of normal variance based on balanced and unbalanced ranked set samples. **Enviromental and ecological statistics**, 6, p. 23-46, 1999.