Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

Construção de Preços Referenciais a partir das Técnicas de Reamostragem de *Jackknife* e *Bootstrap*

Julice Suares Souza

Orientador:

Prof. Joab de Oliveira Lima

João Pessoa, PB – Brasil

Julho/2011

Julice Suares Souza

Construção de Preços Referenciais a partir das Técnicas de Reamostragem de *Jackknife* e de *Bootstrap*

Monografia apresentada ao Departamento de Estatística da Universidade Federal da Paraíba, para obtenção do título de Bacharel em Estatística.

Orientador: **Prof. Joab de Oliveira Lima**

Monografia sob o título *Construção de Preços Referenciais a partir das Técnicas de Reamostragem de Jackknife e de Bootstrap*, defendida por Julice Suares Souza e aprovada em 08 de julho de 2011, para obtenção do título de Bacharel em Estatística.

Banca Examinadora

____

Prof. Joab de Oliveira Lima

Universidade Federal da Paraíba

____

Prof. Hemílio Fernandes Campos Coêlho

Universidade Federal da Paraíba

____

Prof. Marcelo Rodrigo Portela Ferreira

Universidade Federal da Paraíba

Dedico com muito amor a minha mãe, Ivonete Suares, meu maior tesouro e meu grande exemplo de vida.

Agradecimentos

Agradeço a Deus pelo dom da vida e pelas oportunidades que me foram dadas. Por ter me dado forças para concretizar sonhos como este, que hoje se torna realidade. E a minha intercessora, Nossa Senhora, que sempre iluminou meus passos na dura caminhada da vida.

À minha mãe Ivonete, por ser exemplo de amor, perseverança, dedicação e humildade; a qual tenho o maior orgulho de chamar de mãe e amiga. Meu agradecimento eterno por todos os momentos que esteve a meu lado, me apoiando, me escutando, me ensinando, me fazendo acreditar que sou capaz de ultrapassar qualquer barreira; e por me fornecer condições para me tornar a profissional e Mulher que sou.

Ao meu irmão Jonas Rafael, por me ajudar dia após dia a superar as dificuldades e pelo carinho de sempre.

Ao meu namorado Giovanni, por me entender e apoiar nos últimos semestres do curso; e por me proporcionar alegrias imensas demonstrando um amor único.

Aos meus familiares e amigos, em especial Ana Hermínia, Anna Paola, Everlane Suane, Natália, Sadraque e Tiê, pela convivência, companheirismo e experiências compartilhadas.

Ao meu orientador, professor Joab de Oliveira, pelos ensinamentos e conhecimentos compartilhados e, principalmente, pela compreensão que teve a mim em todos os momentos finais do curso. Sou muito grata a ele por toda dedicação dispensada para a concretização dessa monografia.

A todos os professores do departamento de Estatística da UFPB, pela paciência, dedicação e ensinamentos disponibilizados em sala; em especial, aos professores Hemílio Fernandes e Marcelo Rodrigo, membros da Banca Examinadora, pela dedicação e atenção a esta monografia.

“O futuro pertence àqueles que

acreditam na beleza de seus sonhos”

Elleanor Roosevelt

“É graça divina começa bem.

Graça maior persistir na caminha da certa.

Mas graça das graças é não desistir nunca”

Dom Hélder Câmara

Resumo

O processo de geração de preços referenciais é utilizado como base para a aquisição e contratação de serviços públicos. O atual sistema de preço referencial se baseia numa amostra (*amostra-mestre*) muito pequena e, por vezes, pouco representativa da população, sendo obtida por meio da coleta de preços praticados no mercado. As técnicas de reamostragem são propostas para a geração desse *preço referência*, tendo em vista que são alternativas mais eficientes, pois se baseiam em centenas ou milhares de reamostras, coletadas da *amostra-mestre*. É o que se gostaria de ter na prática, poder selecionar várias e várias amostras aleatórias de uma mesma população. Os resultados obtidos indicaram uma proximidade entre os valores referenciais, encontrados por meio da aplicação das técnicas de *Bootstrap* e *Jackknife*, e os preços médios propostos pelo mercado, sendo a *amostra-mestre* coletada nas farmácias do município de João Pessoa.

Palavras-chave: Preço referencial, *Bootstrap*, *Jackknife*.

Abstract

The process of generating benchmark price is used as the basis for procurement and contracting of public services. The current pricing system is based on a reference sample (*master sample*) very small and sometimes unrepresentative of the population, is obtained through the collection of market prices. Resampling techniques are proposed for the generation of this *reference price*, given that alternatives are more efficient because they are based on hundreds or thousands of resampling, the sample collected master. This is what we would like to have in practice over and you can select random samples from the same population. The results indicated a closeness between the reference values found by applying the techniques of *Bootstrap* and *Jackknife*, and the average prices offered by the market, being the *sample master* collected at pharmacies in the city of João Pessoa.

Keywords: Reference price, *Bootstrap*, *Jackknife*.

Lista de Símbolos

μ	Média da população
	Média da amostra
S	Desvio padrão da amostra
n	Tamanho da amostra
α	Nível de significância
$t_{(n-1,\, 1\, -\, \alpha/2)}$	Valor tabelado da distribuição *t* de *Student* com (n-1) graus de liberdade e nível de significância α
θ	Parâmetro de interesse
	Estimador do parâmetro de interesse
	Estimador *Jackknife*
	Estimador da variância *Jackknife*
	Distribuição estimada dos dados
B	Número de reamostras
	Estimador Bootstrap
W^*	Amostra *Bootstrap*
	Réplicas *Bootstrap* do parâmetro de interesse
	Desvio da distribuição *Bootstrap*
	Valor do parâmetro de interesse para cada reamostra *Bootstrap*
	Média das estimativas das reamostras *Bootstrap*

Lista de Tabelas

Lista de Figuras

Sumário

1. Introdução

Os processos de aquisição e contratação de serviços na administração pública são indispensáveis ao bom funcionamento das instituições governamentais. Comprar bem significa empregar corretamente os recursos oriundos dos contribuintes e devolver à sociedade serviços públicos de qualidade a um preço justo.

O Sistema de Preços Referenciais é um conjunto de tabelas de preços praticados pelo mercado e apuração de custos de serviços terceirizados, sendo utilizado como eixo para determinar o melhor preço a ser pago por determinado serviço ou produto adquirido pelo Governo. Com base numa amostra de preços coletados no mercado e por meio de procedimentos estatísticos simples, se determina o preço máximo a ser pago pelo bem ou serviço desejado; esse preço é o limite inferior de um intervalo de confiança gerado pela amostra de preços coletados e utilizado como *preço referência* para as compras públicas (CASAGRANDE; CESTARI; MOTTA, 2011).

O processo de geração do preço referência otimiza a utilização do tempo de trabalho dos servidores, apresentando resultados impactantes no processo de compra e na contratação da administração pública. O trabalho das comissões de licitação é reduzido, pois não há necessidade de se dedicar à realização de novas cotações ou orçamentos a cada novo processo, gerando economia de tempo e dinheiro. Além da agilidade e da economia financeira, a transparência é uma qualidade muito importante, pois qualquer cidadão pode consultar e fiscalizar o valor pago nas compras governamentais por meio do Diário Oficial.

O fato do preço referência ser formado com base em amostras pequenas, muitas vezes pouco representativas da população, devido ao pouco tempo para as cotações ou pela falta de empresas especializadas para determinados serviços solicitados pelo Estado, faz com que se acredite que a metodologia utilizada pode ser melhorada com a aplicação de métodos de reamostragem.

Um método de reamostragem consiste em sortear dados pertencentes a uma amostra que seja representativa da população (obtida por meios probabilísticos), de modo a construir uma nova amostra (chamada também de *pseudo-amostra*), ou seja, é baseada na retirada sucessiva de amostras repetidas com ou sem reposição (SILVA FILHO, 2011). Uma única amostra (*amostra-mestre*) pode gerar um grande número de outras amostras que podem ser empregadas para gerar a distribuição amostral empírica de um parâmetro de interesse.

As técnicas de reamostragem são úteis quando o cálculo dos estimadores por métodos analíticos for complicado, ou seja, quando não se conhece a distribuição de probabilidade dos dados. Reamostrar permite diferentes alternativas para se encontrar médias, desvios padrões e intervalos de confiança através da análise de um conjunto de dados, sem ser necessário confiar na distribuição assumida para os dados ou ser cuidadoso quanto à violação de uma das suposições inerentes aos processos de tomada de decisão (COSTA, 2006). Existem diversas técnicas de reamostragem utilizadas para estimar parâmetros de interesse quando não se conhece muito sobre o comportamento dos dados, as mais citadas na literatura são o *Bootstrap* (EFRON; TIBSHIRANI, 1993) e o *Jackknife* (EFRON, 1979).

O *Jackknife* é um método destinado a estimar a variância dos estimadores em condições teoricamente complexas ou quando não há confiança no modelo especificado, estimando assim o viés dos parâmetros de interesse, no intuito de reduzi-lo. A idéia é simples e se baseia na construção de reamostras da *amostra-mestre*, em que se retira uma observação da amostra e se estima os parâmetros de interesse com base nessa reamostra gerada e assim sucessivamente, até que cada observação tenha sido retirada da *amostra-mestre* exatamente uma vez.

Já o método *Bootstrap* é uma técnica de reamostragem amplamente utilizada na obtenção de estimativas pontuais e intervalares, bem como na avaliação da acurácia de estimativas e testes. Consiste no sorteio, com reposição, dos dados pertencentes a uma amostra retirada da população (*amostra-mestre*), afim de compor uma nova amostra, gerando a estimação dos parâmetros de interesse a cada nova amostra, sem adicionar nenhuma nova informação a *amostra-mestre*. Pode-se dizer que o *Bootstrap* é um método que procura substituir a análise estatística teórica pela força dos métodos computacionais, pois para a construção de cada nova amostra (ou reamostra) é necessário a repetição de um número muito grande de procedimentos e o cálculo de diversas estatísticas para cada uma dessas reamostras (RIZZO; CYMROT, 2006).

Nos casos em que não se conhece a distribuição de probabilidade dos dados é que o *Bootstrap* é mais útil, pois possibilita a obtenção da distribuição amostral de um parâmetro de interesse a partir da *amostra-mestre* (*Bootstrap* não-paramétrico). Porém existem casos em que a distribuição dos dados é conhecida e se utiliza o método para estimar o viés das estimativas dos parâmetros, efetuando por meio do viés as correções necessárias (*Bootstrap* paramétrico). Outra vantagem em se utilizar a técnica de reamostragem *Bootstrap* é a generalidade com que pode ser aplicada, pois requer que menos suposições sejam feitas, além de não exigir diferentes fórmulas para cada problema, podendo ser utilizada em casos gerais sem depender da distribuição original do parâmetro de interesse (CUNHA; COLOSIMO, 2003).

O uso de técnicas como o *Bootstrap* e o *Jackknife* para a construção da distribuição de probabilidade amostral do parâmetro de interesse da amostra de preços coletados é uma alternativa simples que pode melhorar as estimativas desses parâmetros e reduzir os erros, gerando intervalos de confiança mais precisos, com base numa metodologia específica para pequenas amostras. Assim, o preço referência obtido por meio dessas técnicas pode ser mais representativo da real situação do mercado para aquele produto ou serviço, evitando, assim, possíveis inconsistências numéricas provenientes de amostras pequenas e pouco representativas da população.

2. Objetivos

O objetivo desse trabalho foi construir uma metodologia de construção de preço referencial a partir das técnicas de reamostragem de *Bootstrap* e *Jackknife*.

Mais especificamente, se almeja apresentar uma formalização das técnicas de reamostragem de *Bootstrap* e *Jackknife* dentro da teoria de construção de preços referenciais, com o intuito de formar resultados estatisticamente mais confiáveis para a administração pública.

3. Materiais e Métodos

Neste capítulo será apresentada toda a fundamentação teórica dessa monografia, incluindo a técnica padrão utilizada atualmente através do Sistema de Preços Referenciais, bem como as técnicas de reamostragem sugeridas para a formação do preço referência.

3.1. A Formação do Preço Referência

3.1.1. Introdução

O Sistema de Preços Referenciais é utilizado para garantir economia, rapidez, qualidade e transparência nas compras governamentais, a fim de contribuir com a satisfação da sociedade, dando mais credibilidade aos processos licitatórios e contratos formados. Os preços são pesquisados no mercado e se tornam a referência de preços máximos a serem pagos nas licitações públicas. Tais preços máximos nada mais são do que os limites inferiores de intervalos de confiança estatísticos, construídos a partir de metodologias específicas da área (BURATTO; OLIVEIRA; PEREIRA, 2005).

Mas, de um modo geral, os benefícios que justificam a utilização desse sistema são:

- **Transparência:** Tabelas e preços pesquisados são disponibilizados na internet e publicados no Diário Oficial do Estado, possibilitando ao cidadão interessado saber os valores máximos pagos pela administração pública;
- **Agilidade:** Sucesso e redução na duração dos procedimentos de licitação, uma vez que para a formação do preço referência é necessário apenas imprimir a pesquisa disponível na internet e anexá-la ao processo;
- **Economia:** Disputas nas licitações a partir de preços de mercado e renegociação dos contratos já existentes na data da publicação das primeiras pesquisas. Além disso, a necessidade de consolidar em poucos documentos grupos de itens

utilizados por diversos órgãos permite a padronização das especificações de itens e serviços contratados pelo Governo (CASAGRANDE, 2011).

Na realização das pesquisas de preço, os pesquisadores coletam o preço do produto ou serviço em questão, praticado em uma determinada região. Para formar o preço referência são coletados no mercado no mínimo 6 preços do determinado produto e, por meio de procedimentos estatísticos, se determina o melhor preço, que não é nem o maior nem o menor valor coletado (CASAGRANDE; CESTARI; MOTTA, 2011).

Como pode ser notado, as idéias principais da construção do preço referência são muito simples e serão explicadas em detalhes na seção que segue.

3.1.2. Procedimento de Construção

Para o cálculo do preço referência de um determinado item, é necessário, primeiramente, coletar uma amostra (normalmente baseado num plano amostral não-probabilístico), de maior tamanho possível, dos preços praticados no mercado daquele item de interesse.

Com essa amostra de preços associados aquele item (produto), é construído um intervalo de confiança para a média populacional. Os cálculos são apresentados abaixo.

- **Média Aritmética**: Medida de tendência central bastante utilizada na análise exploratória de dados e que consiste no somatório dos valores coletados dividido pelo número total de observações coletadas.

———— , **[1]**

em que x_i são os preços de mercado coletados pelos pesquisadores e n é o tamanho da amostra ou o número de preços coletados;

- **Desvio Padrão**: Medida de dispersão muito utilizada que mede a variabilidade dos valores coletados em torno da média. Seu cálculo é obtido pela equação (2).

; **[2]**

- **Nível de Confiança**: É uma medida que exprime o grau de confiança ou de confiabilidade que se tem em relação a uma estimativa de um determinado parâmetro. Normalmente, o nível de confiança é expresso em percentual como sendo (1–α)100%, em que **α** é o nível de significância do teste ou a probabilidade do erro tipo I (rejeitar a hipótese nula quando essa é verdadeira), também expressa em percentual (BUSSAB; MORETTIN, 2002). Na prática, o grau de confiança mais utilizado é 95%;

- **Intervalos de Confiança (IC)**: É uma região centrada na estimativa pontual do parâmetro de interesse, sendo interpretado da seguinte maneira: utilizando-se um nível de confiança de 95%, se forem construídos um grande número de intervalos (aleatórios), todos baseados em amostras de tamanho *n*, 95% deles conterá o verdadeiro valor do parâmetro de interesse. Por se tratar, geralmente, de uma amostra pequena, o intervalo de confiança é comumente construído a partir da distribuição *t* de *Student* (SIEGEL, 1975), com nível de confiança (1–α)100% e (n–1) graus de liberdade. Assim, o intervalo de confiança é dado pela equação (3).

, **[3]**

em que é a média amostral; é o quantil de nível (1–α/2) da distribuição *t–student* com (n–1) graus de liberdade; *S* é o desvio padrão amostral e *n* é o tamanho da amostra.

Assim, com base nesse procedimento, é possível obter os limites inferior e superior do intervalo de confiança desejado, de modo que esse contenha, com um grau de confiabilidade de (1–α)100%, o valor do parâmetro populacional de interesse. O processo finaliza adotando-se o limite inferior desse intervalo de confiança como sendo o ***Preço Referência*** desejado.

3.1.3. Deficiências do Procedimento de Formação de Preços

Apesar da metodologia aplicada ser correta (média, desvio padrão amostral e IC), é muito arriscado construir um preço referência de um determinado produto com base em uma amostra que na maioria das vezes é coletada sem um plano amostral – de forma não-probabilística (intencional) – e muito pequena. Na prática, muitas pesquisas de preços referência se baseiam em amostras de tamanho 3, 4 ou 5 valores. Esses tamanhos são muito reduzidos e podem comprometer as condições de *regularidade* (distribuição com esperança e variância finitas) e de *convergência* (a distribuição de parâmetros obtidos converge para a distribuição real dos parâmetros) das funções utilizadas na construção do intervalo de confiança, mesmo empregando a distribuição *t* de *Student*.

Por isso, seria interessante que se tentasse construir a distribuição de probabilidade amostral do parâmetro de interesse, com o intuito de simular todas as situações provenientes da população e, assim, obter médias e desvios padrões amostrais mais precisos. Nota-se que a construção da distribuição de probabilidade empírica do parâmetro de interesse é uma aplicação direta das técnicas de reamostragem, tais como os métodos de *Jackknife* e de *Bootstrap*. E são justamente esses métodos que serão discutidos a seguir.

3.2. Técnicas de Reamostragem

3.2.1. Introdução

A reamostragem obtém a distribuição empírica do parâmetro – uma aproximação da verdadeira distribuição de probabilidade do parâmetro de interesse – a partir de centenas ou milhares de reamostras. Ou seja, uma única amostra (*amostra-mestre*) gera um número considerável de reamostras, de maneira que se obtenha a distribuição amostral empírica do parâmetro de interesse e possibilite a construção de testes e intervalos de confiança (CESÁRIO; BARRETO, 2003).

Existem diversas técnicas de reamostragem que visam estimar parâmetros de uma determinada distribuição de interesse, como a de *Bootstrap* e a de *Jackknife*, que diferem entre si apenas na maneira como a reamostra é obtida. O *Jackknife* exclui uma observação amostral (por vez) para gerar cada nova reamostra, enquanto que o *Bootstrap* faz inúmeras reamostragens aleatórias (com reposição) das observações amostrais (SILVA FILHO, 2011; O *JACKKNIFE*, 2011). A seguir, os referidos métodos serão descritos com mais clareza.

3.2.2. Método de *Jackknife*

A metodologia *Jackknife* foi introduzida por Quenouille em 1949 e retomada por Tukey em 1958, sendo desenvolvida para estimar parâmetros de interesse em situações onde os métodos estatísticos clássicos não podem ser aplicados e, com isso, reduzir o viés dos estimadores, além de construir intervalos de confiança em situações teoricamente complexas, tendo como base a divisão da amostra em reamostras (MILLER, 1974).

O método *Jackknife* mais simples extrai n reamostras, de uma determinada amostra observada de tamanho n, pela eliminação seqüencial de um caso por vez. Assim cada reamostra tem um tamanho de n–1 e difere da *amostra-mestre* apenas pelo caso que foi omitido. Apesar desse método ter sido substituído pelo *Bootstrap*, ele continua sendo uma medida viável, principalmente, para a detecção de observações influentes (valores extremos das variáveis) e uma opção computacionalmente mais econômica para muitos pacotes estatísticos (MARTINEZ; LOUZADA-NETO, 2001).

O procedimento para a construção do estimador de *Jackknife* ( ) de certo parâmetro de interesse θ se inicia com a obtenção de uma amostra de tamanho n, digamos A={x_1, x_2, ..., x_n}, de uma certa população X. Para essa amostra, suponha que se sabe estimar o parâmetro θ a partir de um estimador que utiliza os dados amostrais através de uma determinada função F, ou seja, . Essa função F pode representar qualquer estrutura de cálculo, como por exemplo, a média ou a variância da amostra. Então, para cada reamostra gerada a partir de A tem-se o vetor , com ; onde se exclui da amostra observada e se estima o parâmetro de interesse, . Esse procedimento é feito até que todas as n

observações amostrais tenham sido excluídas apenas uma vez, gerando assim o vetor de parâmetros estimados . Com isso, o estimador *Jackknife* de θ será:

[4]

e o estimador da variância *Jackknife* é dado por:

[5]

A FIGURA 1 mostra o algoritmo completo de construção. Já o método de *Bootstrap* é mais geral e será discutido a seguir.

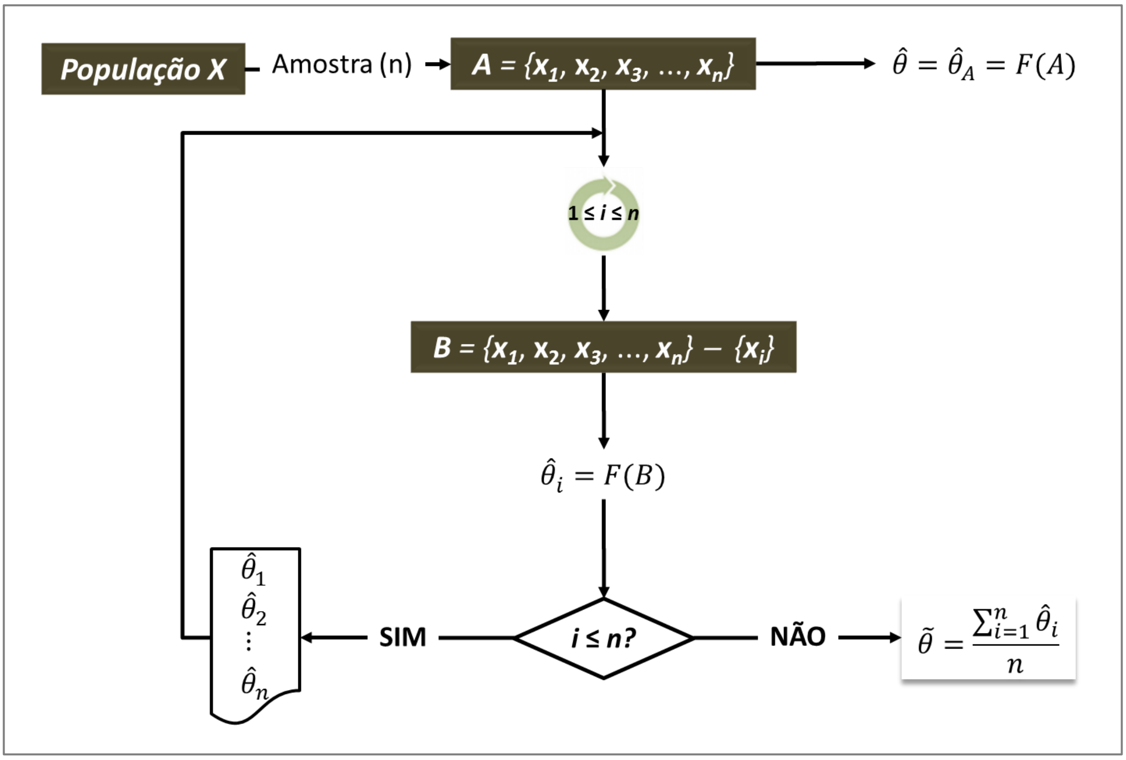

FIGURA 1. Algoritmo do método de reamostragem de *Jackknife*.

3.2.3. Método de *Bootstrap*

A metodologia *Bootstrap* foi introduzida por Efron em 1979, como uma técnica não paramétrica que procura substituir análises estatísticas teóricas, em casos onde as condições de aplicação destas metodologias não são válidas, por métodos de computação intensiva. Esse método tem o mesmo objetivo do anterior, ou seja, reduzir os vieses e prover desvios padrões menores. Mas a grande vantagem do método de *Bootstrap* é a possibilidade de construção da distribuição amostral empírica do estimador do parâmetro de interesse.

Na realização do método *Bootstrap* utiliza-se uma amostra de tamanho *n* (*amostra-mestre)*, que deverá ser coletada de maneira planejada, por meio de procedimentos probabilísticos, no intuito de gerar resultados mais precisos. As reamostras dessa *amostra-mestre* representam o que seria a obtenção de novas amostras da população original. Portanto, a distribuição *Bootstrap* do parâmetro baseada nessas reamostras representa a distribuição amostral desse parâmetro de interesse, o que propicia a construção e aplicação de técnicas de inferência estatística para a estimação de parâmetros.

A quantidade de reamostras, a partir da *amostra-mestre*, necessária para se obter resultados confiáveis com a aplicação desse método não é determinada na literatura especializada, depende do tamanho da *amostra-mestre* utilizada. Se o *n* for igual ou maior que 5, vale utilizar um número bem alto de reamostragens, algo em torno de 1000 reamostras por exemplo; porém, para tamanhos amostrais menores que 5, só vale utilizar o número máximo de combinações possíveis dentro da amostra. Outra forma de determinar essa quantidade é verificar a variação do desvio padrão para a estimativa do parâmetro de interesse calculado nas reamostras. Quando o valor do desvio se estabilizar é um indicativo de que o tamanho da reamostra *Bootstrap* estará adequado.

Uma ferramenta indispensável para a geração de reamostras são as técnicas computacionais, que economizam tempo em muitos procedimentos que manualmente seriam bastante complicados. Além disso, é importante que a reamostragem seja realizada com reposição e com seleção de valores de forma aleatória (amostra probabilista), onde todos os elementos têm chance real e conhecida de serem escolhidos (RIZZO; CYMROT, 2011).

O principal objetivo do *Bootstrap* é estimar um determinado parâmetro através da média de um vetor que contém B estimativas desse parâmetro, todas baseadas em reamostras de tamanho *n* e todas extraídas da *amostra-mestre*. O número de réplicas B é definido pelo usuário. Essa idéia também é válida para outros parâmetros além das médias amostrais.

O *Bootstrap* pode ser paramétrico ou não-paramétrico. Na forma paramétrica considera-se que a função de distribuição dos dados (*F*) pode ser estimada por , a partir de um modelo paramétrico conhecido para os dados. Já no caso não-paramétrico considera-se que a função de distribuição *F* é desconhecida, sendo estimada através da distribuição empírica (TACONELI, 2005).

A aplicação do *Bootstrap* de uma maneira geral tem base no princípio do *plug-in*, que é um método simples de estimação de parâmetros a partir de amostras. O estimador do *plug-in* de um parâmetro é definido como . Dessa forma, nota-se que o *Bootstrap* se utiliza desse princípio, pois indica que se substitua a distribuição populacional pela distribuição dos dados, e a partir daí se extraia amostras (ou reamostras) que imitem o processo de construção de uma distribuição amostral.

As definições para o caso paramétrico e não-paramétrio do *Bootstrap* são detalhadas a seguir, melhorando o entendimento das funções e parâmetros utilizados em cada caso.

- ***Bootstrap* Paramétrico**

A reamostragem *Bootstrap* paramétrica baseia-se em modelos probabilísticos com parâmetros estimados via *amostra-mestre*. Suponha *F* uma distribuição indexada por um parâmetro , com estimador , baseado na *amostra-mestre* , e considere como uma função de . Deste modo, pode-se considerar como estimador de . Sejam, então, amostras da distribuição e estimativas de para estas amostras. Assim, obtém-se B estimativas para o parâmetro de interesse . O *Bootstrap* paramétrico consiste na aproximação da distribuição por .

- ***Bootstrap* Não-paramétrico**

O *Bootstrap* não paramétrico é usado quando não se tem conhecimento a cerca da distribuição dos dados, *F*. Assim, deve-se gerar B amostras com reposição e de mesmo tamanho da *amostra-mestre*, a partir da distribuição , que corresponde à distribuição empírica dos dados:

[6]

em que #( ) representa o número de vezes em que a condição é verdadeira. assume valores no conjunto {0, 1/*n*, 2/*n*, ... , 1} e é também conhecida como freqüência relativa acumulada.

Ao possibilitar a estimação de parâmetros populacionais de interesse baseados exclusivamente na amostra, o *Bootstrap* não-paramétrico é responsável pela produção de resultados significativos, em situações onde as suposições relativas à distribuição populacional não são conhecidas ou válidas. O caso não paramétrico é o mais utilizado e o mais indicado para o estudo do preço referencial, pois esse preço se baseia numa amostra muito pequena, em que não se pode determinar a distribuição dos dados.

3.2.3.1. Procedimento de Construção

Para a construção do estimador de *Bootstrap* ( ) de certo parâmetro de interesse β, considera-se uma amostra de tamanho *n*, digamos A = $\{w_1, w_2, ..., w_n\}$, de uma certa população W. Para essa amostra, suponha que se sabe estimar o parâmetro β a partir de um estimador que utiliza os dados amostrais através de uma determinada função *F*, ou seja, . Essa função *F* pode ser qualquer parâmetro, como a média ou a variância da amostra.

Considerando que é a distribuição empírica de *w*, um vetor de amostras *Bootstrap* é , com , sendo construído escolhendo-se aleatoriamente, com reposição, *n* elementos da amostra A. A replicação *Bootstrap* do parâmetro de interesse para cada amostra *Bootstrap* é denotada por , gerando um vetor de estimativas do parâmetro de interesse . Assim, a estimação do parâmetro de interesse para a *i-ésima* reamostra é dada pela equação (7):

. **[7]**

Assim, o estimador de *Bootstrap* de um parâmetro de interesse β com base em B replicações da amostra original W é dado por:

. **[8]**

Análogo ao que foi feito para o método de *Jackknife*, a FIGURA 2 ilustra o processo completo de geração e estimação do estimador de *Bootstrap*.

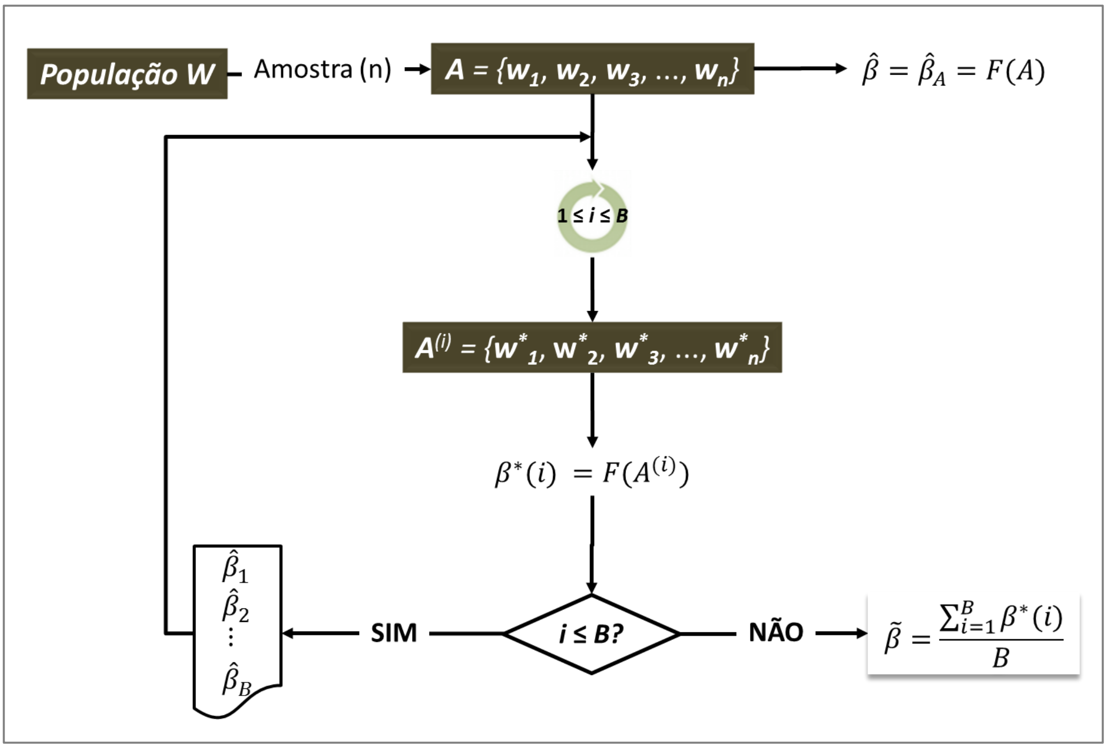

FIGURA 2. Algoritmo do método de reamostragem de *Bootstrap*.

3.2.3.2. Distribuição Amostral do Estimador de *Bootstrap*

Intervalos de confiança, testes de hipóteses e erros padrões baseiam-se na *distribuição amostral* de um parâmetro de interesse, que é a distribuição dos valores que esse parâmetro pode assumir em todas as amostras possíveis de mesmo tamanho, extraídas da mesma população. Em várias situações não se conhece nenhum modelo probabilístico para a população, e é justamente nesses casos que o *Bootstrap* não-paramétrico se mostra como uma boa alternativa, gerando uma distribuição *Bootstrap* no lugar da distribuição amostral.

A variância dos parâmetros estimados pelo *Bootstrap* ocorre de acordo com a escolha da *amostra-mestre* e do número de reamostras a serem geradas, onde a coleta da *amostra-mestre* deve ser feita por meio de técnicas de amostragem probabilistas. O desvio ou erro padrão da distribuição *Bootstrap* para a média é calculado pela equação (9):

, **[9]**

em que é o valor da estatística para cada reamostra; B é o número de reamostragens realizadas.

Em geral, a distribuição *Bootstrap* tem aproximadamente a mesma forma e dispersão da distribuição amostral, porém está centrada no valor do parâmetro estimado pela *amostra-mestre*, e não no verdadeiro valor do parâmetro de interesse. Uma forma de estimar o viés de um parâmetro é observar o quanto a distribuição amostral se afasta do verdadeiro valor do parâmetro, ou seja, é calcular a diferença entre o verdadeiro valor do parâmetro e o valor estimado. O estimador do viés da distribuição *Bootstrap* é dado por:

, **[10]**

em que é o estimador do parâmetro da *amostra-mestre*.

3.2.3.3. Intervalos de Confiança *Bootstrap*

Segundo alguns pesquisadores, uma das aplicações da metodologia *Bootstrap* é para a obtenção de intervalos de confiança mais precisos (menores), devido a grande quantidade de reamostras, havendo diversas técnicas distintas para esse cálculo, que serão abordadas a seguir.

a) Intervalo de Confiança Bootstrap t

Suponha que a distribuição *Bootstrap* de um parâmetro extraído de uma Amostra Aleatória Simples (AAS) de tamanho *n* seja aproximadamente Normal e que o viés seja pequeno. Um intervalo de confiança de nível α aproximado para o parâmetro de interesse é:

$$\_ \quad \overline{=} , \qquad \textbf{[11]}$$

em que é a estimativa do parâmetro; _ é o quantil de nível (1–α/2) de uma distribuição *t* com (*B*–1) graus de liberdade; é o desvio padrão *Bootstrap*.

Observação: Esse intervalo só funciona bem quando sabemos que a distribuição da estatística na distribuição *Bootstrap* é aproximadamente Normal e que a estatística é pouco viciada (RIZZO; CYMROT, 2011).

b) Intervalo de Confiança Bootstrap Percentil

O intervalo de confiança percentil pode ser calculado de duas maneiras. Segundo Efron (1986), para uma confiança (1–α)100%, a primeira forma é encontrar o percentil (1–α/2)100% e o percentil (α/2)100% da distribuição amostral do parâmetro estimado nas reamostras.

A segunda maneira de se obter o intervalo de confiança percentil é através do cálculo dos percentis das diferenças dos valores das estatísticas das reamostras em relação ao valor médio desta mesma estatística nas reamostras (MONTGOMERY; RUNGER, 2003). Mais claramente, para estimar um intervalo de confiança para uma estimativa , calcula-se o valor desse parâmetro para cada uma das *i* reamostras *Bootstrap* ( ) e a média dessas estimativas . Encontra-se, então, para cada reamostra *i*, a diferença (d_i) entre esses valores, isto é:

. **[12]**

Para uma confiança de 95%, encontram-se os percentis 97,5% e 2,5% destas diferenças e calcula-se o intervalo de confiança *Bootstrap* Percentil da seguinte forma:

. **[13]**

Observações:

- Para verificar se o intervalo de confiança *t* calculado é confiável, pode se fazer uma comparação entre o intervalo *t* e o percentil. Se o viés for pequeno e a distribuição *Bootstrap* for aproximadamente Normal, os dois intervalos irão apresentar valores muito próximos. Casos em que os intervalos de confiança *Bootstrap* calculados pela *t* e pelo percentil não tiverem valores próximos, nenhum destes métodos deve ser utilizado (HESTERBERG et al, 2003).
- Se o viés e a assimetria estão presentes de forma muito forte é mais recomendável que se utilize outros métodos *Bootstrap* de correção, como o Método BC (*Bias-corrected*) e o método BC_a (*Bias-corrected and accelerated*) (EFRON; TIBSHIRANI, 1993).

c) Intervalo de Confiança Bootstrap BC (Bias-corrected)

No cálculo do intervalo de confiança *Bootstrap* com correção de viés (BC) os limites do intervalo são os percentis empíricos da distribuição *Bootstrap*, mas não são necessariamente $\alpha/2$ e $1–\alpha/2$, pois são ajustados para corrigir o viés e a assimetria desta distribuição; são mais indicados quando a distribuição dos dados não é simétrica e possui um viés acentuado. Os percentis usados para o cálculo dos limites inferiores e superiores dos intervalos de confiança correção do viés dependem do número k_0 chamado *Bias-corrected*, ou corretor do viés, que é definido pela expressão (14):

– , **[14]**

em que *I* é uma função indicadora que recebe valor 1 se ( ) e valor 0 caso contrário; B é o número de reamostras *Bootstrap* independentes; é a estimativa do parâmetro da *amostra-mestre*; é a estimativa do parâmetro para cada uma reamostra *Bootstrap*; é a função de distribuição da Normal padronizada.

Os limites do intervalo *Bootstrap* com correção de viés são dados por:

, **[15]**

sendo _ e _ – em que Z_x é o *x*-ésimo percentil da distribuição normal padrão – e é igual ao $B(p1)^o$ valor ordenado das replicações .

d) Intervalo de Confiança *Bootstrap* BC$_a$ (*Bias-corrected and accelerated)*

Quando a assimetria estiver presente de maneira muito expressiva, o método mais indicado é o intervalo de confiança *Bootstrap* de correção de viés acelerado, que não difere muito do anterior, a não ser pela constante de aceleração *a*, que ajusta o intervalo de confiança em relação à assimetria apresentada na distribuição dos dados.

O intervalo de confiança BC$_a$ é obtido da mesma forma que o intervalo BC ( ), porém utiliza-se um ajuste por meio de uma constante de aceleração *a*. O cálculo dos limites inferiores e superiores é dado pela equação (16):

,

[16]

,

em que **Φ** é a função de distribuição da Normal Padrão e $Z_{\alpha/2}$ é o α-ésimo percentil da Normal Padrão.

Para calcular *a* e z_0 utiliza-se as expressões abaixo:

,

[17]

,

em que representa o valor das estimativas do parâmetro estudado para cada amostra *i* que consiste na *amostra-mestre* sem a observação *i* da mesma, com e o valor da média das estimativas .

No geral, as técnicas de reamostragem possibilitam a análise da real distribuição dos dados, provendo estimadores mais precisos. Técnicas como essas evitam a total confiança em métodos estatísticos baseados em aproximações e fazem o pesquisador raciocinar um pouco mais a respeito dos dados.

Como a suposição de normalidade (ou simetria) é um fator determinante na escolha e aplicação do intervalo de confiança mais apropriado, a utilização de testes não-paramétricos que verificam a aderência da distribuição dos dados à distribuição Normal se faz necessária. No tópico seguinte serão descritos alguns testes de normalidade.

3.3. Testes de Normalidade

Uma variável aleatória assume uma distribuição de frequências específica, que pode apresentar formas variadas. Essa distribuição de frequência nada mais é do que uma distribuição de probabilidade, onde para um evento tem-se uma probabilidade de ocorrência associada. Porém, a mais importante e mais utilizada distribuição de probabilidade é a distribuição Normal. Sabe-se que a suposição de normalidade dos dados é exigida para a realização de muitos métodos estatísticos, assim como a suposição de independência entre as observações.

Na aplicação de um teste de normalidade, a hipótese nula é que os dados amostrais provêm de uma distribuição Normal. Então, sejam uma amostra aleatória de uma população *X* com distribuição *F* desconhecida. Para testar se *X* tem distribuição Normal – *N*(μ,) – utiliza-se os dados amostrais estandardizados, usando as estimativas de μ e σ.

$$\frac{\quad}{\quad}, \qquad \textbf{[18]}$$

em que e *S* são as estimativas de μ e σ, respectivamente.

Assim, as hipóteses a testar são:

[19]

sendo *N(0,1)* a normal Padrão.

Na literatura existem vários métodos para testar a suposição de normalidade dos dados, como o Shapiro-Wilk e o Lilliefors; além de recursos gráficos, como histograma e normal plot (ou Q-Q plot). Essas estatísticas têm metodologia diferente e serão descritas a seguir.

3.3.1. Teste de *Shapiro-Wilk*

Proposto por Shapiro & Wilk (1965) utiliza a estatística W (equação 18):

, **[20]**

em que são as observações da amostra; a média das observações; as constantes são calculadas como a solução da equação (19):

, **[21]**

sendo o vetor dos valores esperados das estatísticas de ordem da amostra e V a matriz de covariâncias dessas estatísticas.

Esse teste é utilizado quando o número de observações da amostra é reduzido, como no caso da amostra coletada para formar o preço referência. O cálculo no R é feito através da função *shapiro.test* no pacote *nortest*, onde a hipótese de normalidade é aceita quando o *valor-p* é maior que o nível de significância adotado (α).

3.3.2. Teste de *Lilliefors*

Proposto por Lilliefors (1965), esse teste é uma modificação do teste *Kolmogorov-Smirnov* (K-S), que utiliza a estatística D de K-S que mede a diferença máxima absoluta entre a função de distribuição acumulada empírica e a teórica.

A estatística de teste é dada pela equação (20):

, **[22]**

onde – ; ; ,

onde é a função de distribuição Normal Padrão e, e são a média e o desvio padrão amostrais.

A hipótese H_0 é rejeitada, para um nível de significância α, se o valor observado for superior ou igual ao ponto crítico , com tal que,

. **[23]**

O teste *Lilliefors* é o mais famoso teste para verificar normalidade, sendo calculado pelo R através da função *lillie.test* no pacote *nortest*, onde a hipótese de normalidade é aceita quando o *valor-p* é maior que o nível de significância adotado (α).

No capítulo que se segue serão discutidas algumas aplicações dos métodos de *Jackknife* e *Bootstrap* na precificação de alguns medicamentos e materiais de consumo hospitalares. O *software* utilizado para a geração dos resultados foi o R.

4. Resultados

Neste capítulo serão expostos os resultados obtidos com a aplicação das metodologias descrita anteriormente para a formação do preço referência. Para possibilitar a utilização das técnicas de reamostragem foi feita uma pesquisa de preço de mercado de medicamentos e equipamentos utilizados na área de saúde, por meio da coleta de preços nas farmácias de João Pessoa, em diferentes bairros da cidade.

Os medicamentos utilizados na pesquisa foram o Captopril de 25mg e a Ranitidina de 150mg (ampola); já o equipamento pesquisado foi o Equipo para Soro. Esses itens foram escolhidos por serem bastante utilizados nos hospitais de emergência, sendo itens que não podem faltar de forma alguma no atendimento de saúde pública.

4.1. Aplicação 1

O Captopril é um anti-hipertensivo utilizado em casos de hipertensão, insuficiência cardíaca congestiva, infarto do miocárdio e nefropatia diabética, atuando na diminuição imediata da pressão arterial. Para a formação do preço referência, foram coletados preços unitários de comprimidos do medicamento gerando uma amostra de tamanho $n = 11$ (TABELA 1).

TABELA 1: Preços (R$) coletados para o comprimido de Captopril de 25mg.

0,07	0,29	0,40	0,52	0,52	0,58
0,63	0,79	0,99	1,06	1,64	

Com base nos preços coletados, realizou-se uma análise descritiva a fim de determinar o preço referência a ser utilizado, onde se obteve o preço médio de R$0,6809 para cada unidade de comprimido de Captopril, com desvio de R$ 0,4291.

Utilizando a amostra coletada como *amostra-mestre*, foram aplicadas as duas técnicas de reamostragem para a formação do preço referência e os resultados serão mostrados a seguir e separadamente.

4.1.1. Preço Referência pelo Método de *Jackknife*

Para a geração dos novos dados pelo método *Jackknife* foi utilizada 11 reamostragens, em que a cada nova amostra gerada era retirada uma observação, resultando num vetor de dados de tamanho $n_i = 10$ (com $1 \leq i \leq 11$). É possível constatar que a média e o desvio das reamostras foram respectivamente 0,6809 e 0,0429. A FIGURA 3 ilustra o histograma e o gráfico de probabilidade normal (Q-Q Plot), em que se observou a aderência dos dados à distribuição Normal pelo teste de normalidade *Shapiro-Wilk* (p = 0,502).

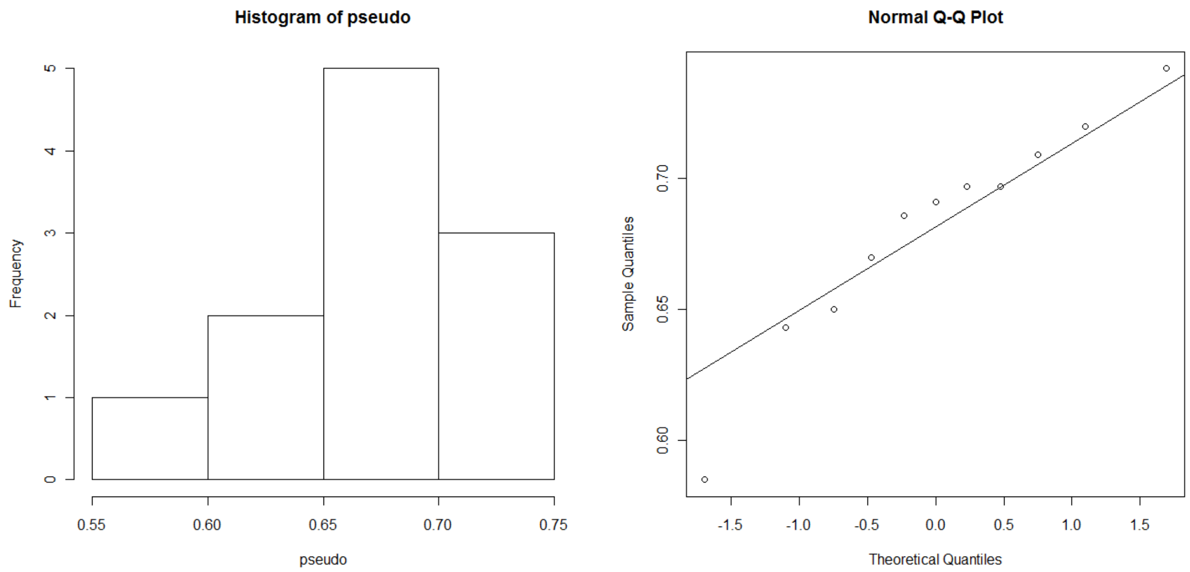

FIGURA 3. Histograma e *qq-plot* das médias das reamostras *Jackknife* para a Aplicação 1.

O intervalo de confiança encontrado foi (0,6502; 0,7116), ao nível de 95% de confiança, indicando que o preço referência a ser utilizado nas compras públicas é R$0,6502.

4.1.2. Preço Referência pelo Método de *Bootstrap*

Foram realizadas mil reamostragens com reposição a partir da *amostra-mestre*. A FIGURA 4 mostra o histograma e o Q-Q plot para as médias das reamostras. O teste de *Lilliefors* atestou a normalidade do vetor de *Bootstrap* (p = 0,396).

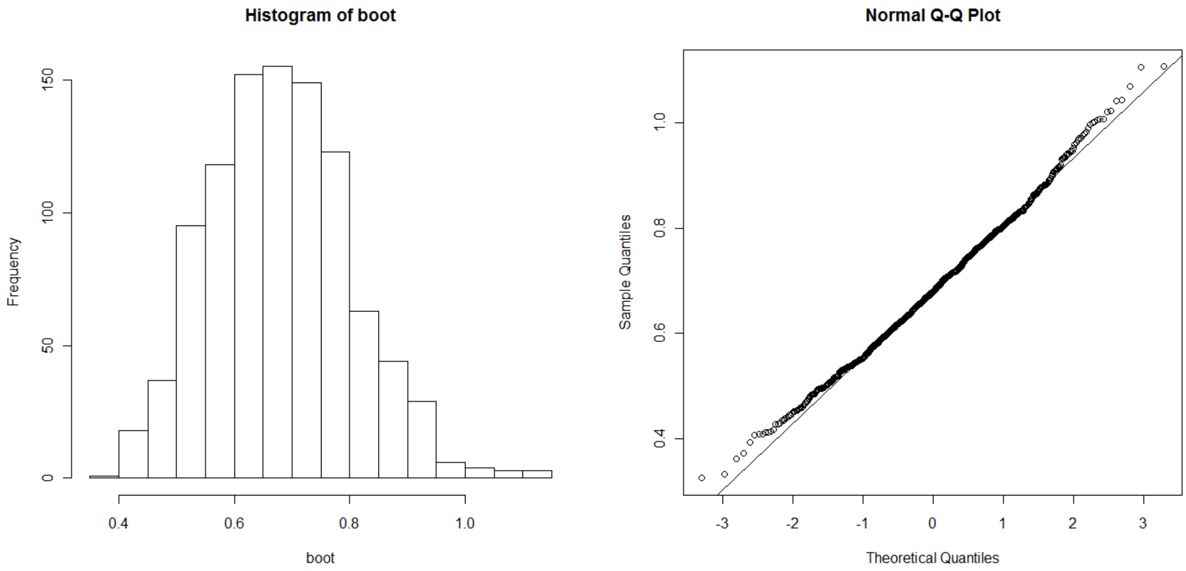

FIGURA 4. Histograma e *qq-plot* das médias das reamostras *Bootstrap* para a Aplicação 1.

Os valores de média e desvio das reamostras *Bootstrap* foram respectivamente iguais a 0,6831 e 0,1241, apresentando um viés para a média de 0,002, considerado muito pequeno. Nestas condições, foram calculados os intervalos de confiança para as médias utilizando o método *Bootstrap* t, resultando no preço referência de R$ 0,6732 por unidade, ao nível de confiança de 95%.

4.1.3. Comparação entre os Métodos

A TABELA 2 mostra os preços referência de acordo com 3 níveis de confiança diferentes (90%, 95% e 99%), em que se observa valores muito próximos da média da *amostra-mestre*, R$ 0,68. Por considerar um número bem maior de reamostras, a técnica de *Bootstrap* apresenta uma variação bem menor que a do *Jackknife*.

TABELA 2: Preços referência (R$) para diferentes níveis de confiança.

Método	**Níveis de Confiança**		
	90%	**95%**	**99%**
Bootstrap	0,6745	0,6732	0,6708
Jackknife	0,6560	0,6502	0,6368

4.2. Aplicação 2

A Ranitidina é um medicamento anti-ulceroso indicado no tratamento da úlcera péptica gástrica e duodenal, esofagite de refluxo, gastrite e duodenites. Esse medicamento promove uma diminuição na produção de ácido e pepsina no estômago, favorecendo a cicatrização da gastrite e/ou das úlceras pépticas do estômago e do duodeno e prevenindo suas complicações. Também pode ser usado no tratamento de doenças relacionadas à hipersecreção ou hipersensibilidade à secreção gástrica, tais como esofagite de refluxo. Para a formação do preço referência foram coletados preços unitários de ampolas do medicamento, gerando uma amostra de tamanho $n = 5$ (TABELA 3).

TABELA 3: Preços coletados (R$) para a ampola de Ranitidina de 150mg.

0,99	1,21	1,27	1,90	2,29

O preço médio encontrado com base na amostra coletada foi de R$ 1,532 para cada unidade de ampola de Ranitidina, com desvio de R$ 0,5422.

Utilizando a amostra coletada como *amostra-mestre*, foram aplicadas as duas técnicas de reamostragem para a formação do preço referencial e os resultados serão discutidos a seguir e separadamente.

4.2.1. Preço Referência pelo Método de *Jackknife*

Na utilização do método *Jackknife* foram realizadas 5 reamostragens, em que a cada nova amostra gerada era retirada uma observação, resultando num vetor de dados de tamanho $n_i = 4$ (com $1 \leq i \leq 5$). Verificou-se que a média e o desvio das reamostras foram respectivamente 1,532 e 0,1356. A FIGURA 5 ilustra o histograma e o gráfico de probabilidade normal (Q-Q Plot), cujo teste de *Shapiro-Wilk* (p = 0,427) confirmou a normalidade dos dados.

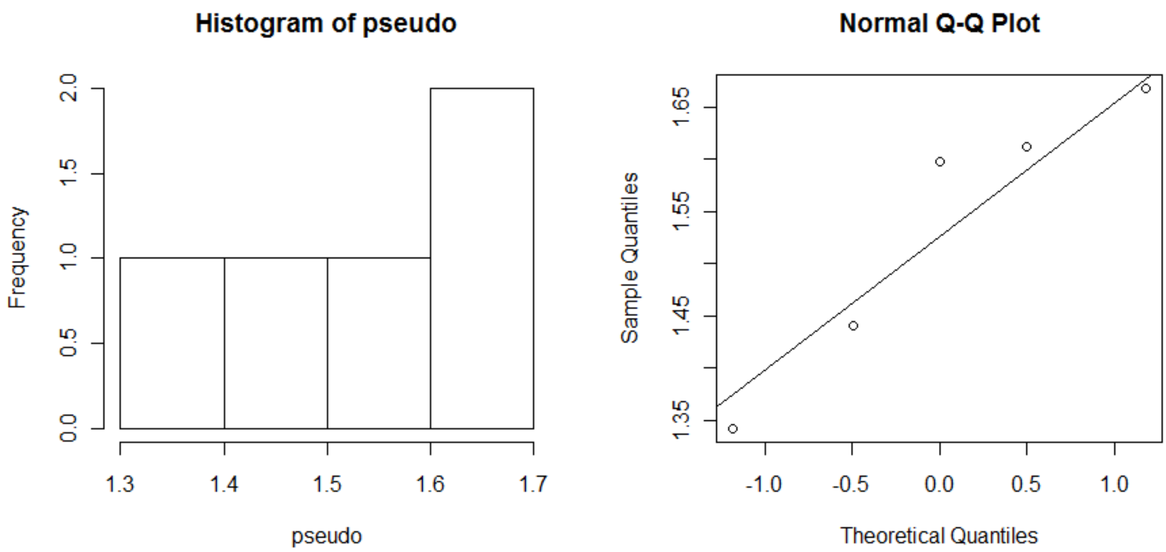

FIGURA 5. Histograma e *qq-plot* das médias das reamostras *Jackknife* para a Aplicação 2.

O intervalo de confiança encontrado foi (1,3163; 1,7477), ao nível de 95% de confiança, indicando que o preço referência a ser utilizado nas compras públicas é R$1,3163.

4.2.2. Preço Referência pelo Método de *Bootstrap*

Assim como foi realizado para a Aplicação 1, foram realizadas mil reamostragens com reposição a partir da *amostra-mestre*. A FIGURA 6 mostra o histograma e o Q-Q plot, cuja normalidade dos dados não foi confirmada pelo teste de *Lilliefors* (p = 0,0007).

Nota-se que o teste de normalidade de *Lilliefors* foi bastante rigoroso nesse caso, pois observando o histograma exposto na FIGURA 6, há evidências estatísticas suficientes para concluir que a distribuição amostral do estimador do parâmetro é normal. O Teorema do Limite Central (TLC) mostra que à medida que o tamanho da amostra aumenta, a distribuição das médias amostrais tende a uma distribuição normal (MEYER, 1969).

Ainda assim, com base nos valores de média e desvio das reamostras *Bootstrap* foram respectivamente iguais a 1,5318 e 0,2248, resolveu-se construir o intervalo de confiança Bootstrap através do método do Percentil (item *b* do capítulo 3.2.3.3), o que resultou num preço referência de R$ 1,146.

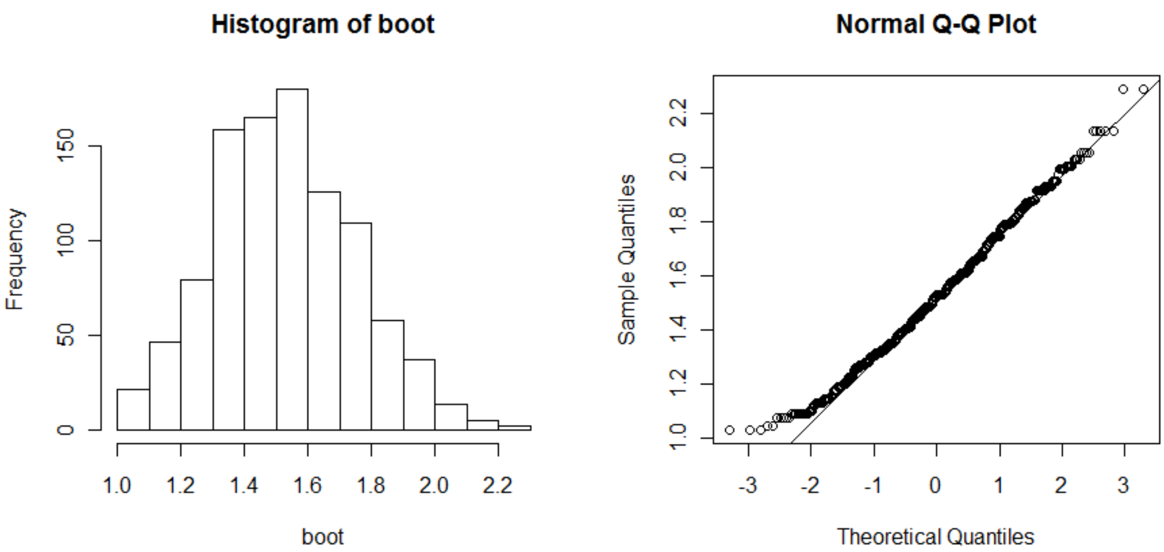

FIGURA 6. Histograma e *qq-plot* das médias das reamostras *Bootstrap* para a Aplicação 2.

4.2.3. Comparação entre os Métodos

A TABELA 4 mostra os preços referência de acordo com vários níveis de confiança, onde se observa que os valores são próximos da média da *amostra-mestre*, R$ 1,532, além da variação entre os diferentes níveis de significância ser um pouco maior que a aplicação anterior, tanto para o *Bootstrap* como para o *Jackknife*.

TABELA 4: Preços referência (R$) para diferentes níveis de confiança.

Método	**Níveis de Confiança**		
	90%	**95%**	**99%**
Bootstrap	R$ 1,1900	R$ 1,1460	R$ 1,0460
Jackknife	R$ 1,3725	R$ 1,3163	R$ 1,1361

4.3. Aplicação 3

O Equipo é um aparelho descartável utilizado como acessório encaixado na agulha que está no paciente, tendo a finalidade de aplicar soluções parenterais e soro. Ele possibilita a locomoção e os movimentos sem precisar interromper a aplicação do soro no paciente. Para a formação do preço referência foram coletados preços unitários de comprimidos do medicamento, gerando uma amostra de tamanho n = 11 (TABELA 5).

TABELA 5: Preços (R$) coletados para equipo de soro.

0,71	0,79	0,83	0,99	1,09	1,18
1,18	1,24	1,38	1,45	1,51	

O preço médio encontrado para o equipo foi de R$ 1,1227 para cada unidade, com desvio de R$ 0,27. Esses valores servirão de balizamento para a construção dos intervalos de confiança através dos métodos de *Jackknife* e *Bootstrap* apresentados a seguir.

4.3.1. Preço Referência pelo Método de *Jackknife*

Pelo método *Jackknife* foram utilizadas um número de 11 reamostragens, onde a cada nova amostra gerada era retirada uma observação, resultando num vetor de dados de tamanho $n_i = 10$ (com $1 \leq i \leq 11$). Pode-se verificar que a média e o desvio das reamostras foram respectivamente 1,1227 e 0,027. A FIGURA 7 ilustra o histograma e o gráfico de probabilidade normal (Q-Q Plot), em que o *valor-p* = 0,674 do teste de *Shapiro-Wilk* atestou a normalidade dos dados, mesmo que, aparentemente, a distribuição dos pontos no histograma não leve a essa conclusão.

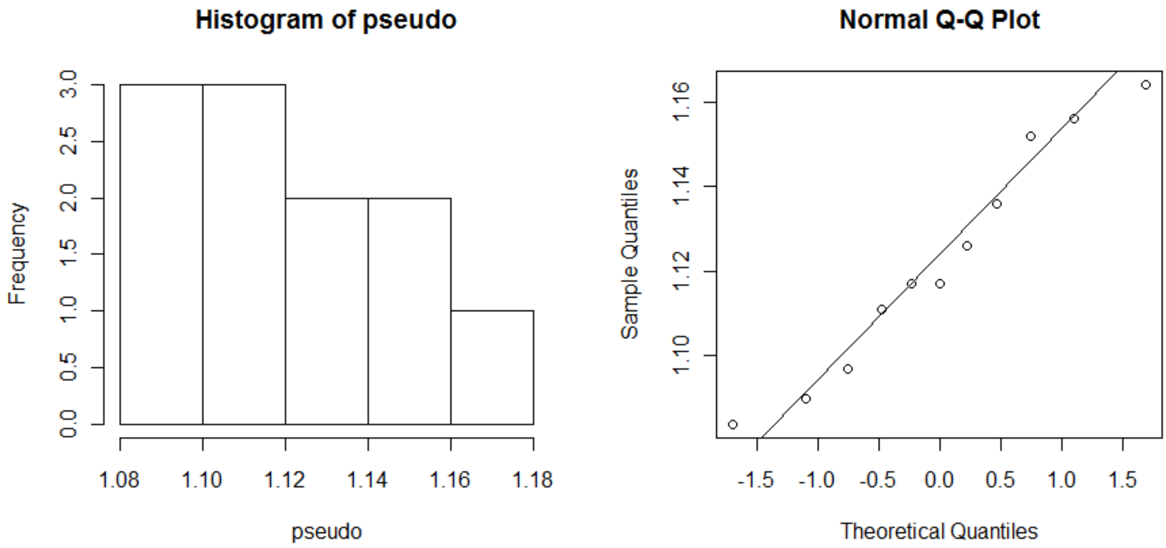

FIGURA 7. Histograma e *qq-plot* das médias das reamostras *Jackknife* para a Aplicação 3.

O intervalo de confiança encontrado foi (1,1034; 1,142), ao nível de 95% de confiança, indicando que o preço referência a ser utilizado nas compras públicas é R$1,1034.

4.3.2. Preço Referência pelo Método de *Bootstrap*

Pelo método *Bootstrap* foram realizadas mil reamostragens com reposição a partir da *amostra-mestre*. A FIGURA 8 mostra o histograma e o Q-Q plot, utilizando o método de *Lilliefors*, para as médias das reamostras, em que se verificou a aderência dos dados à distribuição Normal (p = 0,7163).

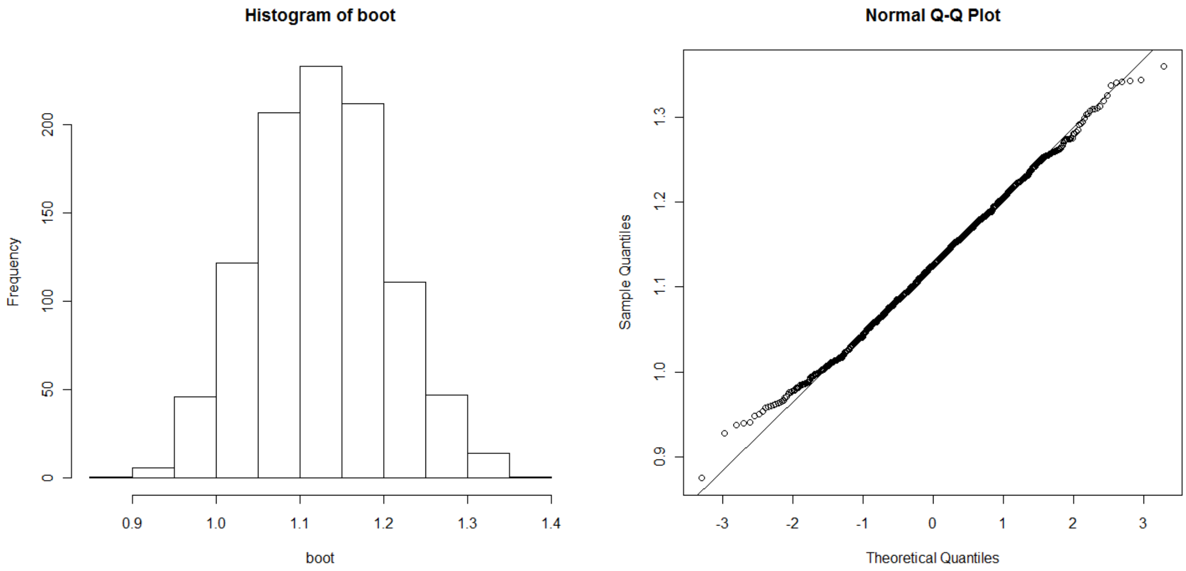

FIGURA 8. Histograma e *qq-plot* das médias das reamostras *Bootstrap* para a Aplicação 3.

Os valores de média e desvio das reamostras *Bootstrap* foram respectivamente iguais a 1,1256 e 0,0785, apresentando um viés para a média de – 0,007. Nestas condições, foram calculados os intervalos de confiança para as médias utilizando o método *Bootstrap* t, resultando o intervalo (1,1179; 1,1276). Dessa forma, o preço referência utilizado seria R$ 1,1179, ao nível de confiança de 95%.

4.3.3. Comparação entre os Métodos

A TABELA 4 mostra os preços referência de acordo com vários níveis de confiança, onde se observa que os valores são próximos da média da *amostra-mestre*, R$ 1,1227, não havendo muita variabilidade de valores.

TABELA 6: Preços referência (R$) para diferentes níveis de confiança.

Método	Níveis de Confiança		
	90%	**95%**	**99%**
Bootstrap	R$ 1,1186	R$ 1,1179	R$ 1,1163
Jackknife	R$ 1,1071	R$ 1,1034	R$ 1,0949

5. Considerações Finais e Sugestões de Trabalhos Futuros

A aplicação de técnicas de reamostragem na construção do preço referência de medicamentos e equipamentos utilizados na área de saúde se mostra uma alternativa mais robusta do ponto de vista estatístico e de fácil acesso, onde se obtém resultados extremamente próximos da média de preço do mercado obtidos pela *amostra-mestre*.

Os métodos de *Bootstrap* e de *Jackknife* apresentaram resultados muito próximos, para os diferentes níveis de confiança adotados; porém, devido à quantidade superior de reamostras, o método *Bootstrap* gerou intervalos de confiança menores (mais precisos) e mais próximos da realidade de mercado, com variação a partir da segunda casa decimal. Em relação à suposição de normalidade, os histogramas ilustraram curvas bastante simétricas para o caso *Bootstrap*, já pelo *Jackknife* não se pode chegar a nenhuma conclusão; mesmo se sabendo que conclusões baseadas em gráficos não são seguras.

Os testes de normalidade aplicados, *Shapiro-Wilk* e *Lilliefors*, aceitaram a aderência dos dados a distribuição Normal na maioria dos casos, sendo observada a rejeição da hipótese de normalidade apenas na aplicação *Bootstrap* onde se utilizou preços do medicamento Ranitidina.

Com base nos resultados obtidos, uma opção de continuidade na aplicação dessas técnicas é trabalhar com o preço referência de produtos e serviços em geral, não apenas na área de saúde. Uma excelente opção para atender as necessidades dos governantes, com decisões tomadas baseadas em técnicas estatísticas. Outra opção é trabalhar com uma técnica híbrida de reamostragem, que envolva o melhor de cada um dos métodos de *Jackknife* e *Bootstrap*, sendo necessário um estudo mais aprofundado dessa idéia inicial.

6. Referências Bibliográficas

BURATTO, A. L.; OLIVEIRA, P. J. R.; PEREIRA, W. S. Preços de referência de obras públicas – Planilha orçamentária padrão. In: SIMPÓSIO NACIONAL DE AUDITORIA DE OBRAS PÚBLICAS, 10, 2005, Recife. *Anais eletrônicos...* Recife, 2005.

BUSSAB, W. O.; MORETTIN, P. A. *Estatística básica*. 5 ed. São Paulo: Saraiva, 2002.

CASAGRANDE, M. L. *Sistema de preços referenciais do Estado do Espírito Santo*. Disponível em <sistemas3.seplag.ce.gov.br/download/CONGRESSO/.../Leila_ Casagrande.ppt>. Acesso em: 20 maio 2011.

CASAGRANDE, M. L.; CESTARI, A. N.; MOTTA, A. P. P. D. *Preços Referenciais: economia, rapidez e qualidade nas compras governamentais*. Disponível em <www.seplag.rs.gov.br/download.asp?nomeArq=Painel_33_Angeliki...>. Acesso em: 30 março 2011.

CESARIO, 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. *Revista de Matemática e Estatística*, São Paulo, v. 21, n. 3, p. 7-20, 2003.

CUNHA, W. J; COLOSIMO, E. A. Intervalos de confiança *Bootstrap* para modelos de regressão com erros de medida. *Revista de Matemática e Estatística*, São Paulo, v. 21, n. 2, p. 25-45, 2003.

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

EFRON, B.; TIBSHIRANI, R. *Bootstrap* methods for standard errors, confidence intervals and others measures of statistical accuracy. *Statistical Science*, v. 1, n. 1, p. 55-77, 1986.

EFRON, B.; TIBSHIRANI, R. *An introduction to the Bootstrap*. Chapman and Hall, 1993.

HESTERBERG, T.; MOORE, D. S.; MONAGHAN, S.; CLIPSON, A.; EPSTEIN, R. Bootstrap methods and permutation testes. In: *The practice of business statistics: using data for decisions*. New York: W. H. Freeman, 2003. cap. 18.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, n. 62, p. 399-402, 1967.

MARTINEZ, E. Z.; F. Estimação intervalar via *Bootstrap*. *Revista de Matemática e Estatística*, São Paulo, v. 19, p. 217-251, 2001.

MEYER, P. L. *Probabilidade: aplicações à estatística*. 2 ed. Rio de Janeiro: LTC, 1983.

MILLER, R. G. The *Jackknife* – a review. *Biometrika*, v. 61, n. 1, p. 1-15, 1974.

MONTGOMERY, D. C.; RUNGER, G. C. *Estatística aplicada e probabilidade para engenheiros*. 2 ed. Rio de Janeiro: LTC, 2003.

O *JACKKNIFE* e o *Bootstrap*. Disponível em <http://www.isa.utl.pt/dm/mestrado/mmacb/UCs/ta/boot_jackk_apres.pdf >. Acesso em: 20 abril 2011.

R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2008. ISBN 3-900051-07-0. Disponível em: <http://www.Rproject.org>. Acesso em: 28 abril 2011.

RIZZO, A. L. T.; CYMROT, R. Estudo e aplicações da técnica *Bootstrap*. In: II Jornada de Iniciação Científica PIBIC e PIVIC, 2006. São Paulo. *Anais*... São Paulo, Universidade Presbiteriana Mackenzie, 2006.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete sample). *Biometrika*, Great Britain, v. 52, n. 3, p. 591-611, 1965.

SIEGEL, S. *Estatística não-paramétrica para ciências do comportamento*. São Paulo: McGraw-Hill do Brasil, 1975.

SILVA FILHO, A. S. *Inferência em amostras pequenas: métodos Bootstrap*. Disponível em < http://geocities.ws/augustofilho/*Bootstrap*.pdf>. Acesso em: 15 abril 2011.

TACONELI, C. A. *Reamostragem Bootstrap em amostragem por conjuntos ordenados e intervalos de confiança não-paramétricos para a média*. 2005. 196 f. Tese (Mestrado em Estatística) – Programa de Pós-Graduação em Estatística, Universidade de São Carlos, São Paulo, 2005.

7. Anexo I

Código-fonte em R para a Construção dos Preços Referênciais através dos Métodos de Jackknife e de Bootstrap

```
## DADOS

dados1 = c(0.40, 0.58, 1.64, 0.63, 0.52, 0.29, 0.79, 0.07, 0.99, 1.06, 0.52)
dados2 = c(1.90, 0.99, 1.21, 1.27, 2.29)
dados3 = c(0.79, 0.83, 1.18, 1.24, 1.51, 1.45, 0.71, 1.18, 1.38, 1.09, 0.99)

## ESTATISTICA DESCRITIVA

# Captopril_25mg
media_1 <- mean(dados1)
var_1 <- var(dados1)
desvio_1 <- sd(dados1)

IC_1.1 = media_1 + 2.228*(desvio_1/sqrt(length(dados1)))
IC_1.2 = media_1 - 2.228*(desvio_1/sqrt(length(dados1)))

# Ranitidina_Ampola
media_2 <- mean(dados2)
var_2 <- var(dados2)
desvio_2 <- sd(dados2)

IC_2.1 = media_2 + 2.776*(desvio_2/sqrt(length(dados2)))
IC_2.2 = media_2 - 2.776*(desvio_2/sqrt(length(dados2)))

# Equipo para Soro
media_3 <- mean(dados3)
var_3 <- var(dados3)
desvio_3 <- sd(dados3)

IC_3.1 = media_3 + 2.228*(desvio_3/sqrt(length(dados3)))
IC_3.2 = media_3 - 2.228*(desvio_3/sqrt(length(dados3)))


## BOOTSTRAP – EXEMPLO PARA O BANCO DADOS 3

x <- dados3
boot <- numeric(1000)
  for (i in 1:1000) boot[i] <- mean(sample(x, replace=T))
media_boot <- mean(boot)
```

```
var_boot <- var(boot)
desvio_boot <- sqrt(var_boot)

hist(boot)
qqnorm(boot)
qqline(boot)

lillie.test(boot)

# INTERVALOS DE CONFIANÇA BOOTSTRAP
IC_boot_1 = media_3 - qt(0.95, length(boot)-1)*(desvio_boot/sqrt(length(boot)))
IC_boot_2 = media_3 + qt(0.95, length(boot)-1)*(desvio_boot/sqrt(length(boot)))

IC_boot_1 = media_3 - qt(0.975, length(boot)-1)*(desvio_boot/sqrt(length(boot)))
IC_boot_2 = media_3 + qt(0.975, length(boot)-1)*(desvio_boot/sqrt(length(boot)))

IC_boot_1 = media_3 - qt(0.995, length(boot)-1)*(desvio_boot/sqrt(length(boot)))
IC_boot_2 = media_3 + qt(0.995, length(boot)-1)*(desvio_boot/sqrt(length(boot)))

# INTERVALOS DE CONFIANÇA PERCENTIS
IC_perct_1 = quantile(boot, 0.05)
IC_perct_2 = quantile(boot, 0.95)

IC_perct_1 = quantile(boot, 0.025)
IC_perct_2 = quantile(boot, 0.975)

IC_perct_1 = quantile(boot, 0.005)
IC_perct_2 = quantile(boot, 0.995)


## JACKKNIFE – EXEMPLO PARA O BANCO DADOS 3

w <- dados3
jack <- numeric(length(w)-1)
pseudo <- numeric(length(w))
  for (i in 1:length(w))
  { for (j in 1:length(w))
    { if(j<i) jack[j] <- w[j] else if(j>i) jack[j-1] <- w[j] }
  pseudo[i] <- mean(jack)
  }
pseudo
media_jack <- mean(pseudo)
var_jack <- var(pseudo)
desvio_jack <- sqrt(var_jack)
```

```
hist(pseudo)
qqnorm(pseudo)
qqline(pseudo)

shapiro.test(pseudo)

# INTERVALOS DE CONFIANÇA JACKKNIFE
IC_jack_1 = media_3 - qt(0.95, length(jack)-1)*sqrt(var_jack/length(jack))
IC_jack_2 = media_3 + qt(0.95, length(jack)-1)*sqrt(var_jack/length(jack))

IC_jack_1 = media_3 - qt(0.975, length(jack)-1)*sqrt(var_jack/length(jack))
IC_jack_2 = media_3 + qt(0.975, length(jack)-1)*sqrt(var_jack/length(jack))

IC_jack_1 = media_3 - qt(0.995, length(jack)-1)*sqrt(var_jack/length(jack))
IC_jack_2 = media_3 + qt(0.995, length(jack)-1)*sqrt(var_jack/length(jack))
```