REVISTA DE CIÊNCIAS EXATAS E TECNOLOGIA

INFERÊNCIA EM AMOSTRAS PEQUENAS: MÉTODOS BOOTSTRAP

Augusto Sousa da Silva Filho – *– Faculdade Anhanguera de Belo Horizonte - unidade Centro*

RESUMO: A amostra original representa a população da qual foi extraída. Dessa maneira, as reamostras obtidas a partir dessa amostra representam o que obteríamos se retirássemos diversas amostras da população. A distribuição Bootstrap de uma estatística, baseada em um grande número de reamostras, representa a distribuição da estatística, com base em um grande número de amostras. A importância de seu uso e as técnicas utilizadas para encontrar os seus parâmetros está descrita neste artigo.

PALAVRAS-CHAVE:
Reamostragem. Intervalo de Confiança. Erro Padrão.

ABSTRACT: The original sample represents the population from which it was extracted. Thus, the resampling obtained from this sample represent what we would have to take out several samples of the population. The Bootstrap distribution of a statistic, based on a large number of resampling, represents the distribution of statistics based on a large number of sample. The importance of its use and the techniques used to find its parameters is described in this article.

KEYWORDS:
Resampling. Confidence Interval. Standard Errors.

Informe Técnico
Recebido em: 27/11/2009
Avaliado em: 15/07/2010
Publicado em: 24/02/2014

Publicação
Anhanguera Educacional Ltda.

Coordenação
Instituto de Pesquisas Aplicadas e Desenvolvimento Educacional - IPADE

Correspondência
Sistema Anhanguera de Revistas Eletronicas - SARE
rc.ipade@anhanguera.com

v.5 • n.5 • 2010 • p.115 - 126

1. INTRODUÇÃO

Existem métodos de estimação e testes de significância que produzem estimadores e testes estatísticos com propriedades desejáveis em amostras grandes. Em amostras pequenas é interessante o estudo do desempenho dos estimadores ou dos testes estatísticos para determinar quão confiável é a inferência assintótica obtida.

Neste artigo, será visto alternativas de reamostragem – métodos baseados em retirar sucessivamente amostras repetidas e sua análise através do método Bootstrap.

Os métodos de reamostragem permitem quantificar incertezas calculando erros padrões e intervalos de confiança, bem como realizar testes de significância. Eles requerem menos suposições e geralmente fornecem respostas mais precisas do que os métodos tradicionais (MOORE, McCABE, DUCKWORTH, SCLOVE, 1996).

Segundo (MADDALA, 2003), os testes de Razão de Verossimilhança, Wald e o Multiplicador Lagrangeano têm distribuições assintóticas normal ou x^2. Na prática, porém, não se sabe como é a performance desses testes em amostras pequenas. Ainda segundo (MADDALA, 2003), muitos apresentam distorções de tamanho substanciais, isto é, pode-se testar ao nível de 5% de significância usando-se as distribuições assintóticas normal ou x^2 , sendo que o verdadeiro nível de significância é 25%. Além disso, as performances de dois estimadores que têm a mesma distribuição assintótica normal podem ser diferentes em amostras pequenas.

Para examinar esses problemas, discute-se o método de reamostragem, ou métodos que dependem da retirada de amostras repetidas. Para isso, será apresentado o Métodos Bootstrap que resolve diferentes aspectos de inferência em amostras pequenas. Para (SEBER, 2004), a sua utilização visa reduzir desvios e prover desvios padrões mais confiáveis.

2. VANTAGEM DA REAMOSTRAGEM

Segundo Moore, McCabe, Duckworth e Sclove (2006), os métodos de reamostragem (método Bootstrap, Monte Carlo, etc), permite quantificar a incerteza calculando os erros padrões e intervalos de confiança, bem como realizar testes de significância. A sua utilização exige menos suposições e geralmente fornecem respostas mais precisas do que os métodos tradicionais. A reamostragem possui diversas vantagens, entre elas

- Menos suposições: os métodos de reamostragem não requerem que as distribuições sejam normais, nem que as amostras sejam grandes;
- Maior precisão: são mais precisos, na prática, que os métodos clássicos;
- Generalidade: os métodos de reamostragem são bastante similares para um grande número de estatísticas e não exigem novas formulas para cada estatística;
- Função pedagógica: os procedimentos Bootstrap aprimoram nossa intuição, fornecendo-nos analogias concretas com os conceitos teóricos;

3. MÉTODO DE REAMOSTRAGEM: BOOTSTRAP

Segundo (MADDALA, 2003), o método Bootstrap é uma técnica de reamostragem com o seguinte propósito: reduzir desvios e prover desvios padrão mais confiáveis. O seu funcionamento é dado da seguinte maneira:$(y_{1,} y_2, ..., y_n)$ seja a amostra dada. Retira-se dessa amostra uma amostra de tamanho n com reposição. Chama-se essa amostra de $B_j = \left(y_1^*, y_2^*, \ldots, y_n^*\right)$. Essa é amostra Bootstrap. Cada y_i^* é uma escolha aleatória de $\left(y_1, y_2, \ldots, y_n\right)$ E faz-se isso para $j = 1,2,\ldots,m$ e calcula-se de cada uma das amostras Bootstrap β_j. A distribuição de $\hat{\theta}_j$ é a distribuição Bootstrap do estimador θ .As estimativas Bootstrap do desvio e da variância de θ são derivadas dessa distribuição Bootstrap.

Portanto, na prática do Bootstrap, é essencial o uso de um programa de computador, pois a estimação dos parâmetros Bootstrap requer algum esforço computacional.

Para Efron e Tibshirani (1986), a reamostragem não acrescenta nenhuma informação nova à amostra original. Desta forma, a grande vantagem dos métodos como o Bootstrap é o resultado da maneira pela qual a informação amostral é processada. Em se tratando da distribuição normal, toda informação sobre a média amostral é concentrada na média amostral e na variância amostral. Logo, outras maneiras de processar a informação amostral não produzem melhores resultados nesse caso. São nestes casos em que não há distribuição amostral finita das estatísticas prontamente disponível que o Bootstrap é extremamente útil.

3.1. Por que o Bootstrap funciona

Moore, McCabe, Duckworth e Sclove (2006), afirmam que pode parecer que o Bootstrap crie dados a partir do nada. Isso parece suspeito. Entretanto, não se está utilizando as informações das reamostras como se fossem dados reais - o Bootstrap não é um substituto para o acréscimo de dados com vistas ao aumento da precisão. Em vez disso, a idéia do Bootstrap é de empregarem as médias das reamostras para se estimar como a média amostral de uma amostra de tamanho N, extraída dessa população, varia em decorrência da amostragem aleatória.

Ao se utilizar os dados duas vezes - uma vez para se estimar a média populacional (μ) e outra, para se estimar a variação das médias amostrais - é um procedimento perfeitamente legítimo. Visto que, em inferência, isso foi feito inúmeras vezes: como por exemplo, no calculo de $\bar{x}$ ou de $s/\sqrt{n}$ a partir do mesmo conjunto de dados. O que se tem de diferente agora é que:

1) Calcula-se um erro padrão utilizando a reamostragem, em vez da fórmula.

   $s/\sqrt{n}$

2) Utiliza-se a distribuição Bootstrap para verificar se a distribuição amostral é, ou não, aproximadamente Normal, em vez de simplesmente esperar que a amostra seja grande o suficiente para que o teorema central do limite se aplique;

A idéia do Bootstrap também é válida para outras estatísticas além das médias amostrais. Para utilizar o Bootstrap de maneira mais geral, utiliza-se o Princípio do Plug-In. Este principio consiste em estimar um parâmetro, uma quantidade que descreve a população, utilizando a estatística que é a quantidade correspondente para a amostra.

Para Seber e Wild (2004), o principio do plug-in sugere que a média populacional μ seja estimada por meio da média amostral $\bar{x}$, e que naturalmente o desvio padrão populacional σ seja estimado pelo desvio padrão amostral s. Conseqüentemente, pode-se estimar a mediana populacional pela mediana amostral. Para estimar o desvio padrão $\sigma/\sqrt{n}$ da média amostral para uma amostra aleatória simples, aplica-se o principio do *plug-in,* empregando s na fórmula para obter $s/\sqrt{n}$ Logo, a idéia do Bootstrap em si consiste no principio do *plug-in.*

4. DISTRIBUIÇÃO AMOSTRAL E DISTRIBUIÇÃO DE BOOTSTRAP

Segundo (HOEL, 1971), a distribuição amostral de uma estimativa é a função de densidade ou a função de probabilidade que descreve o comportamento probabilístico da estatística em amostragem repetida do mesmo universo ou do mesmo modelo de associação de variável do processo.

Para Moore, McCabe, Duckworth e Sclove (2006), na prática, não se podem tomar um número muito grande de amostras aleatórias para construir a distribuição amostral. Em vez disso, utiliza-se um atalho: as leis da probabilidade nos dizem em algumas situações qual é a distribuição amostral. Se a população tem uma distribuição Normal, então a distribuição amostral de x também é Normal.

Em situações em que não se está definido um modelo para a população e não é possível a extração de uma quantidade muito grande de amostras, o Bootstrap encontra o cenário ideal para a sua utilização. Neste contexto, utiliza-se a única amostra disponível como se fosse a população e dela extraí-se diversas reamostras, para construir-se a distribuição Bootstrap. Usa-se a distribuição Bootstrap no lugar da distribuição amostral.

Entretanto na prática, não se costuma ser exeqüível extraírem-se todas as reamostras possíveis. Desta forma, realiza-se o Bootstrap utilizando cerca de 1000 reamostras escolhidas aleatoriamente. Pode-se estimar diretamente a distribuição amostral escolhendo aleatoriamente 1000 amostras de mesmo tamanho a partir da população original. Entretanto, é muito mais rápido e barato fazer o computador obter as reamostras a partir da amostra original, do que se selecionar diversas amostras da população.

Na maioria dos casos, a distribuição Bootstrap aproxima-se da mesma forma e dispersão da distribuição amostral, no entanto está centrada no valor da estatística original, e não no valor do parâmetro de interesse. Como o uso do Bootstrap é possível o cálculo dos erros padrões originais das estatísticas para as quais não se dispõem de fórmulas, bem como chegar a Normalidade para estatísticas que não podem ser manipuladas facilmente pela teoria.

5. INTERVALOS DE CONFIANÇA BOOTSTRAP

Hall (1992) descreve dois métodos para a obtenção de intervalos de confiança "bootstrap" – método percentil e método percentil t. Os intervalos obtidos via o método percentil tem por base unicamente os quantis e outras medidas da distribuição "bootrstap" do estimador de interesse $\hat{\gamma}$. Os intervalos gerados via o método percentil t tem a forma:

$$\hat{\gamma} - t_1 s(\hat{\gamma}) \leq \gamma \leq \hat{\gamma} - t_2 s(\hat{\gamma})$$

onde $s(\hat{\gamma})$ é o desvio padrão estimado de $(\hat{\gamma})$. As quantidades t_i são determinadas com base na distribuição "Bootstrap" de $\hat{\gamma}$.

Souza (1998) descreve formalmente os dois métodos de determinação de intervalos de confiança. Seja $A = (x_1, x_2, ..., x_n)$ uma amostra aleatória de tamanho n de uma população com função de distribuição $F(x)$, média μ e variância finita σ2. Pelo Teorema Central do Limite:

$$\sqrt{n}\left(\frac{\overline{x} - \mu}{s}\right) \xrightarrow{\upsilon} N(0,1)$$

Pode-se obter um intervalo de confiança para μ resolvendo em t a equação quantílica:

$$P_F\left\{\sqrt{n}\left(\frac{\overline{x} - \mu}{s}\right) \leq t\right\} = \beta$$

sendo $\beta \in (0,1)$ O Teorema Central do Limite produz a solução aproximada $t = z_\beta$ onde:

$$P\{N(0,1) \leq z_\beta\} = \beta$$

Desta forma, obtém-se o intervalo:

$$\left(\overline{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

ao nível $100(1-\alpha)\%$ para μ. Ainda segundo Souza (1998), o método percentil t consiste na substituição de z_β por uma quantidade t_β derivada da distribuição "Bootstrap". Representando-se $\overline{x}^*$ e s* a média e o desvio padrão calculados para a amostra "Bootstrap". A solução aproximada para a equação quantílica populacional se obtém resolvendo a equação amostral:

Logo, seja $t_\beta^{(1)}$ a solução. O intervalo "Bootstrap" percentil t ao nível de $100(1-\alpha)\%$ para μ é:

$$\left(\overline{x} - t_{1-\alpha/2}^{(1)}\frac{s}{\sqrt{n}}, \overline{x} - t_{\alpha/2}^{(2)}\frac{s}{\sqrt{n}}\right)$$

Para Souza (1998), os intervalos de confiança obtidos via o método percentil aparecem em três tipos. Percentil simples, percentil viés corrigido e percentil viés corrigido acelerado. Desta forma, tem-se o intervalo de confiança percentil simples para θ

$$\left(H^{-1}(\alpha/2), H^{-1}(1-\alpha/2)\right)$$

onde H(u) é a função de distribuição de

Souza (1998), ainda define o intervalo de confiança percentil viés corrigido para o parâmetro real θ, ao nível de :

$$\left[H^{-1}\left(\Phi\left(2z_0 - z_{\alpha/2}\right)\right), H^{-1}\left(\Phi\left(2z_0 + z_{\alpha/2}\right)\right)\right]$$

com $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$

O intervalo percentil viés corrigido acelerado para θ ao nível de $100(1-\alpha)\%$ é dado a seguir: $\left[H^{-1}(\Phi(z(\alpha/2))), H^{-1}(\Phi(z(1-\alpha/2)))\right]$

sendo $z(\beta) = z_0 + \frac{z_0 + z_\beta}{1 - a(z_0 + z_\beta)}$ e $z_0 = \Phi^{-1}\left(H(\hat{\theta})\right)$, $\Phi(z_\beta) = \beta$

e é uma constante denominada constante de aceleração.

5.1. Estimativa Bootstrap do Erro Padrão

Moore, McCabe, Duckworth e Sclove (2006), afirmam que há situações em que o erro padrão do estimador é desconhecido. Geralmente, esses são os casos em que a forma de $\hat{\theta}$ é complicada e os operadores padrões do valor esperado e da variância são difíceis de aplicar. Nestes casos o Bootstrap é utilizado.

Suponha que se esteja amostrando a partir de uma população que possa ser modelada pela distribuição de probabilidades $f(x;\theta)$. A amostra aleatória resulta em valores $x_1, x_2, \ldots, x_n$ e obtêm-se $\hat{E}$ como uma estimativa de θ . Agora, usa-se um computador para obter amostras Bootstrap provenientes da distribuição $f(x;\theta)$ e, para cada uma dessas amostras, calcula-se a estimativa Bootstrap $\hat{\theta}*$ de θ .

Geralmente, são consideradas $B = 100$ ou 200 dessas amostras Bootstrap. Faça-se $\bar{\theta}* = (1/B)\sum_{i=1}^{B} \hat{\theta}_i^*$ ser uma média amostral das estimativas Bootstrap. A estimativa Bootstrap do erro padrão de é apenas o desvio padrão da amostra para $\hat{\theta}_i*$, ou

$$s_{\hat{\varsigma}} = \sqrt{\frac{\sum_{i=1}^{B}\left(\hat{\theta}_i^* - \bar{\theta}*\right)^2}{B-1}}$$

Na literatura sobre Bootstrap, o denominador $B-1$ na equação acima é freqüentemente trocado por B. No entanto, para valores grandes geralmente empregados para , há pouca diferença na estimativa produzida para $s_{\hat{\varsigma}}$.

6. APLICAÇÃO

A utilização dos métodos Bootstrap será apresentada via exemplos práticos. Para isso, usou-se o programa R ou Minitab. A base de dados utilizada neste exemplo faz parte do programa S-Plus e está disponível em www.insightful.com/Hesterberg/bootstrap. A Verizon é uma empresa telefônica responsável por uma grande área da região leste dos Estados Unidos. Como tal, cabe a ela fazer o serviço de reparos para os clientes das demais companhias telefônicas dessa região. A Verizon estará sujeita a multas caso os tempos de reparo (tempo para resolver problemas nas linhas telefônicas) para os clientes das empresas concorrentes forem substancialmente maiores que os tempos para os seus próprios clientes. Isso é determinado por meio de testes de hipóteses, negociados junto à Comissão de Serviços Públicos. Começa-se a análise observando a estatística descritiva dos clientes da Verizon.

Estatística Descritiva: Tempo

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Tempo	1664	0	8,412	0,360	14,690	0,000	0,730	3,590	7,080

Figura 1 – Estatística Descritiva dos tempos de reparo dos clientes da Verison.

De acordo com os dados observados, o tempo médio de reparo foi de 8,41 horas com um desvio padrão de 14,69 horas. Estas estatísticas foram extraídas de uma única amostra aleatória. Muito embora a amostra seja de 1664 observações, sendo por isso considerada grande, a amostra não se comportou seguindo uma distribuição Normal de Probabilidade. As figuras a seguir ajudam a entender esta afirmação.

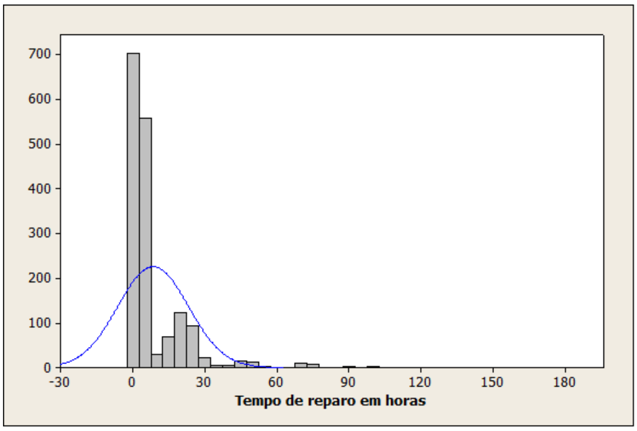

Figura 2 - Gráfico dos tempos de reparo.

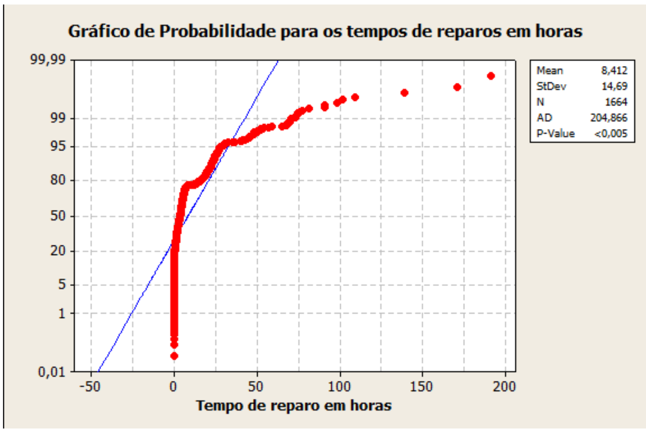

Figura 3 - Gráfico dos tempos de reparo.

De acordo com os gráficos acima, os dados referentes ao tempo de reparo não se comportam seguindo uma distribuição normal. A Figura 3, mostra a estatística de Anderson-Darling igual a 204,866 e o p-valor associado ao teste (<0,005), indicando haver fortes evidências amostrais para rejeitar-se a hipótese de que a distribuição seja proveniente de uma distribuição normal ao nível de 5% de significância. O gráfico a seguir mostra a distribuição de 1000 médias de reamostras para os dados dos tempos de reparo da Verizon, utilizando um histograma e uma curva de densidade.

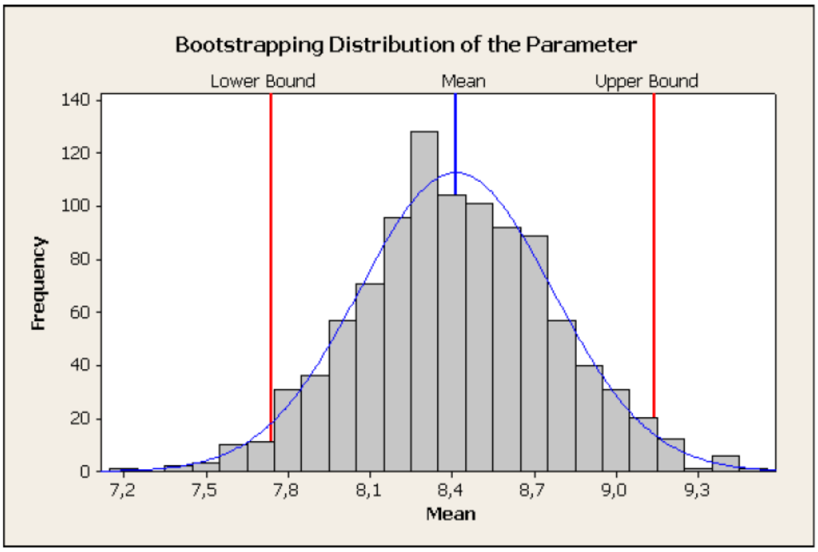

Figura 4 – Simulação Bootstrap de 1000 médias

Vê-se que a distribuição Bootstrap é aproximadamente Normal. O teorema central do limite diz que a distribuição amostral da média $\bar{x}$ é aproximadamente Normal se n for grande. Assim, a forma da distribuição Bootstrap está próxima daquela que se espera que a distribuição amostral tenha. A distribuição Bootstrap está centrada próximo à média da amostra original. Ou seja, como estimador da média da amostra original, a média da distribuição Bootstrap apresenta um viés pequeno. Sabe-se que a distribuição amostral $\bar{x}$ está centrada na média populacional μ, ou seja, que $\bar{x}$ é um estimador não-viciado de μ. Dessa forma, a distribuição das reamostras de novo se comporta (a partir da amostral original) como se espera que a distribuição amostral se comporte (a partir da população). Encontrou-se também o intervalo de confiança para a média e para a mediana para as 1000 média reamostradas.

Para encontrar tal intervalo, utiliza-se uma macro que necessita de três informações: (b, est, alfa). Supondo que o conjunto de valores de interesse se encontra na célula C1 do aplicativo, temos que entrar com as seguintes informações: (b= número de interações).

Bootstrap Confidence Interval

The 95% Bootstrap Confidence Interval (Percentile Method)

Mean	Lower Bound	Upper Bound
8,41480	7,73949	9,13879

Figura 5 – Intervalo de confiança Bootstrap no Minitab

Neste exemplo utilizou-se um total de 1000 interações. A seguir, temos (est). O valor (1) representa que foi solicitado um intervalo de confiança para a média e o valor (2) indica a solicitação de um intervalo de confiança para a mediana. E o último valor de entrada é o nível de significância do teste. Neste exemplo, procurou-se um intervalo ao nível de 95% de confiança. A macro para Minitab for Windows 15 (Confidence Intervals for the Mean or Median using Bootstrap Methods Code), encontra-se disponível na web em: http://www.minitab.com/en-US/support/macros/default.aspx?action=code&id=108. A seguir, encontra-se o intervalo de confiança Bootstrap para a mediana pelo método dos percentis.

Bootstrap Confidence Interval

The 95% Bootstrap Confidence Interval (Percentile Method

Median	Lower Bound	Upper Bound
3,6	3,22	3,82

Figura 6 – Intervalo de confiança para a mediana pelo método dos percentis

O intervalo de confiança para o valor mediano é de 3,22 a 3,82. Foi utilizado o método dos percentis, com um nível de confiança de 95%.

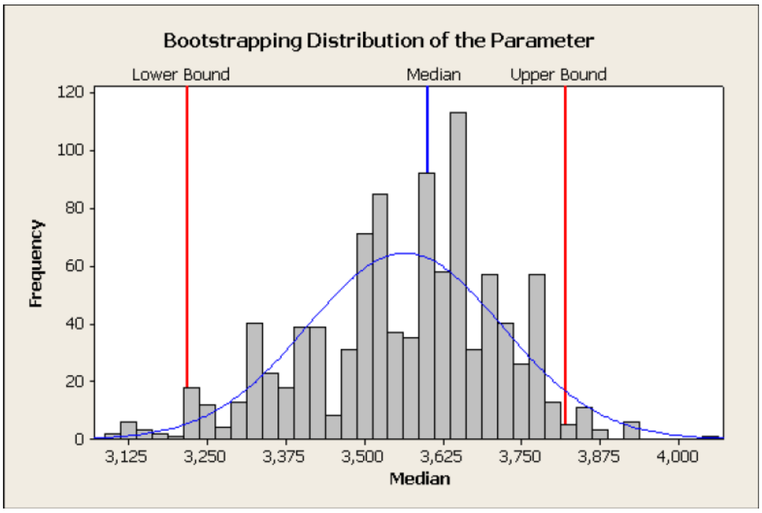

Figura 7 – Simulação Bootstrap

Considere uma amostra qualquer (por exemplo, a amostra MEDIDAS.MTW, disponível em ftp://ftp.est.ufmg.br/pub/fcruz/pacotes/MEDIDAS.MTW. Temos:

original	media_o	reamostra	media_r	media
9,2980	9,3938	11,3871	9,4259	10,8253

Figura 8 – Aplicação das macros nos dados

Suponha que se queira fazer inferência sobre a média da população correspondente, mas como a amostra é muito pequena, decide-se por usar a técnica de Bootstrap (uma técnica de reamostragem), para melhorar a estimativa. Desta forma, deve-se construir-se uma macro que (i) extraia 5 amostras desta amostra (na prática são necessárias umas 200), de igual tamanho, com reposição, (ii) calcule a media de cada uma destas 5 amostras e (iii) disponibilize a diferença entre duas vezes a média da amostra original e média das médias das reamostras (a estimativa melhorada). Suponha que se queira fazer inferência sobre a média da população correspondente, mas como a amostra é muito pequena, decide-se por usar a técnica de Bootstrap (reamostragem), para melhorar a estimativa.

```
MTB > base 1000
MTB > %frederico.mac
Executing from file: frederico.mac
Data Display
Média
9,85763
```

Figura 9 – Aplicação das macros nos dados

No Minitab, obteve-se os seguintes resultados.

original	media_o	reamostra	media_r	media
9,2980	9,90392	11,3871	9,9908	9,85763
9,3938		9,3925	10,0543	
11,3871		11,3871	9,6418	
9,4259		9,2980	10,1064	
10,8253		9,3938	9,9578	

Figura 10 – Saída Computacional

Logo, tem-se a média da amostra original =9,903293 e a média das médias das reamostras 9,85763.

7. CONCLUSÃO

Neste trabalho verificou-se que para se fazer o Bootstrap para uma estatística (por exemplo a média amostral), deve-se retirar centenas de reamostras com reposição a partir da amostra original e calcular a estatística em questão para cada reamostra e inspecionar a distribuição Bootstrap das estatísticas dessas reamostras.

Procurou-se aplicar a metodologia Bootstrap a exemplos práticos e observou-se que a distribuição Bootstrap aproxima-se da distribuição amostral da estatística. Isso é um exemplo do princípio do *plug-in*. Em geral, as distribuições Bootstrap possuem aproximadamente a mesma forma e dispersão da distribuição amostral, porém está centrada na estatística (dos dados originais), ao passo que a distribuição amostral está centrada no parâmetro da população.

Na análise do exemplo Verizon, constatou-se que o Bootstrap não é um substituto para o acréscimo de dados com vistas ao aumento da precisão. Em vez disso, a idéia do Bootstrap é a de se empregar as médias das reamostras para se estimar como a média amostral de uma amostral de tamanho 1664, extraída dessa população, varia em decorrência da amostra aleatória.

A técnica de Bootstrap tenta realizar o que seria desejável realizar na prática, se tal fosse possível: repetir a experiência. As observações são escolhidas de forma aleatória e as estimativas re-calculadas.

REFERÊNCIAS

BOOTH, J.G.; HALL, P.; WOOD, A.T.A. **Balanced importance resampling for the bootstrap.** Annals of Statistics, 21, 286–298, 1993.

DAVISON, A.C.; HINKLEY, D.V. **Bootstrap Methods and Their Application**. Cambridge University Press, 1997.

DAVISON, A.C.; HINKLEY, D.V.; SCHECHTMAN, E. **Efficient Bootstrap Simulation.** Biometrika, 73, 555–566, 1996.

EFRON, B.; TIBSHIRANI, R. "**Bootstrap Methods for Standard Errors, Confidence Intervals and Other measures of Statistician Acuracy**", Statitical Science, Vol. 1, 1986, pp. 54-77.

EFRON, B., **Bootstrap methods: another look at the jackknife Ann.** Stat., Beachood, v.7, p. 1-26, 1979.

HALL, P. The **Bootstrap and edgeworth expansion**. New York: Springer-Verlang, 1992.

HINES, W.W.; MONTGOMERY, D.C.; GOLDSMAN, D.M.; BORROR, C.M. Probabilidade e Estatística na Engenharia. Editora LTC: Rio de Janeiro, p. 227, 2006.

HOEL, P.G. **Introduction to Mathematical Statistics.** John Wiley & Sons. New York, 1971.

MOORE, D. S.; McCABE, G.P.; DUCKWORTH, W.M.; SCLOVE, S.L. **The Practice of Business Statistics: Using data for decisions**. 1a. ed. LTC: Rio de Janeiro, p. 785, 1996.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**, Editora LTC: Rio de Janeiro, p. 155, 1993.

MADDALA, G.S. **Introdução à Econometria.** 3ª. ed., Editora LTC: Rio de Janeiro, p. 318, 2003.

Minitab 15 for Statistical Program. Version 15.

SOUZA, G.S. **Introdução aos Modelos de Regressão Linear e Não Linear.** Editora Embrapa: Brasilia, 1998.

SEBER, G.A.F.; WILD, C.J. **Encontros com o acaso – Um Primeiro Curso de Análise de Dados e Inferência**. Editora LTC: Rio de Janeiro, 2004.

Program R for Statistical. Version 2.11.1. **The R Foundation for Statistical Computing.**