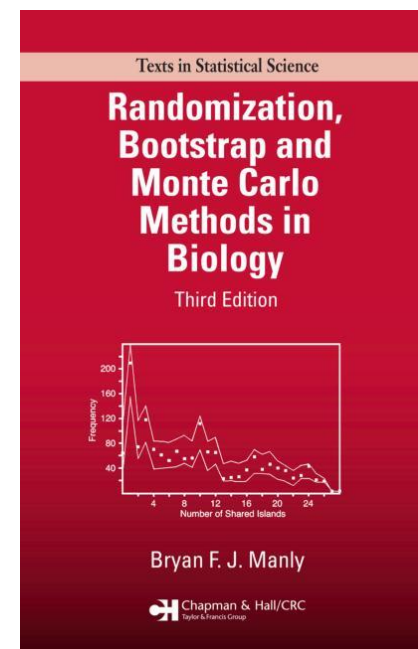


# 3

## Métodos *Bootstrap*



- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3<sup>a</sup> ed., Chapman & Hall/CR Press, Boca Raton, 2007.

---

### ESQUEMA DO CAPÍTULO

#### 3.1 REAMOSTRAGEM COM REPOSIÇÃO

#### 3.2 LIMITES DE CONFIANÇA BOOTSTRAP PADRÃO

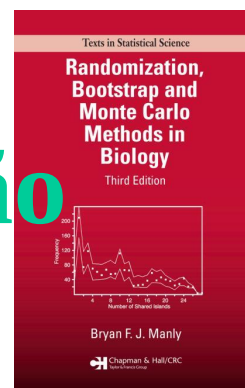
#### 3.3 LIMITES DE CONFIANÇA PERCENTÍLICOS SIMPLES

#### 3.4 OBSERVAÇÕES FINAIS

## 3.1 Reamostragem com Reposição

- Essencialmente, a ideia do *bootstrap* é utilizar, **na ausência de qualquer outra informação**, a distribuição de uma amostra aleatória de tamanho  $n$  de uma população de interesse, como guia da distribuição desta população;
- Em outras palavras, uma população infinita que consiste nos  $n$  valores da amostra obtida, **cada um com igual probabilidade  $1/n$  de ocorrência**, é utilizada para modelar a população real desconhecida.

# 3.1 Reamostragem com Reposição (cont.)



- O termo *bootstrap* foi introduzido por Efron (1979), para denotar uma técnica que é supostamente análoga a alguém saindo de um atoleiro puxando a si mesmo pelos cabelos ([Raspe, Rudolf Erich. As Aventuras do Barão de Munchausen](#));
- Este termo segue o mesmo espírito de Tukey (1958), que denotou *jackknife* o esquema de reamostragem que ele criou anteriormente.

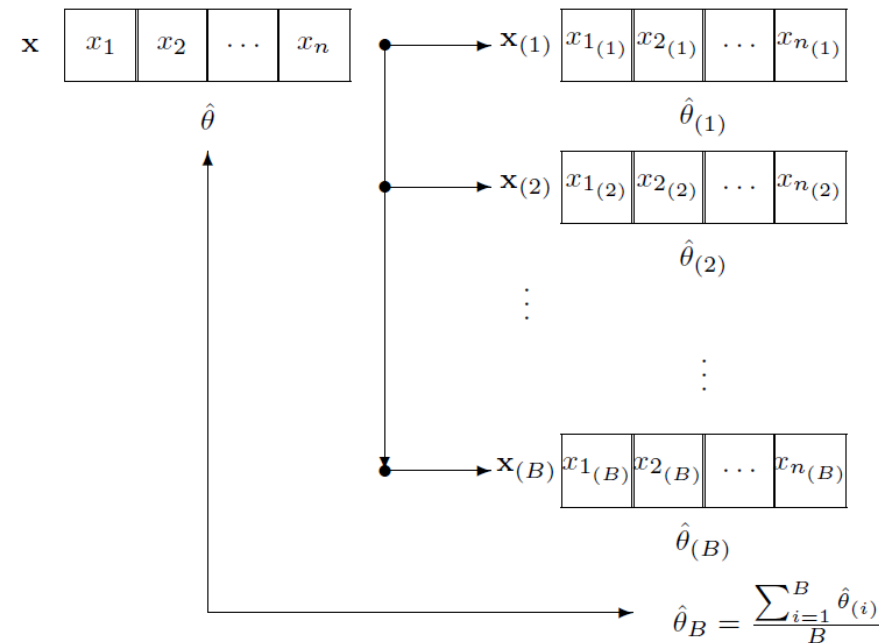


# 3.1 Reamostragem com Reposição (cont.)

A **Figura** ilustra o *bootstrap*: com

- determinação da estimativa  $\hat{\theta}$ , a partir da amostra original,
- $B$  reamostras *bootstrap* (com reposição),  $\mathbf{x}_{(i)}$ ,
- a estimativa  $\hat{\theta}_{(i)}$ , a partir da  $i$ -ésima reamostra;
- e a **estimativa bootstrap**

$$\hat{\theta}_B = \sum_{i=1}^B \hat{\theta}_{(i)} / B$$

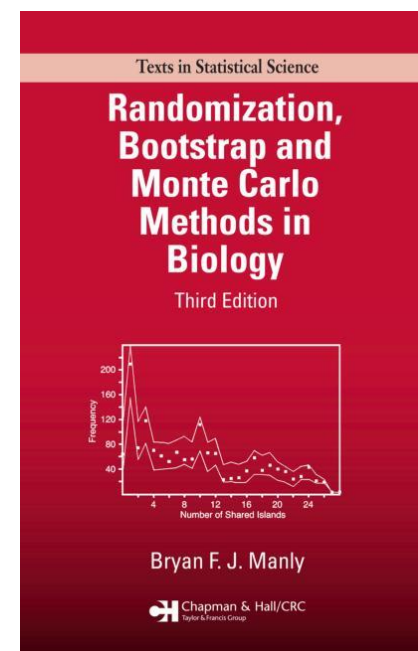


**Figura:** O *bootstrap*

# 3

## Métodos *Bootstrap*

- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3ª ed., Chapman & Hall/CR Press, Boca Raton, 2007.



---

### ESQUEMA DO CAPÍTULO

3.1 REAMOSTRAGEM COM REPOSIÇÃO

3.2 LIMITES DE CONFIANÇA BOOTSTRAP  
PADRÃO

3.3 LIMITES DE CONFIANÇA PERCENTÍLICOS  
SIMPLES

3.4 OBSERVAÇÕES FINAIS

## 3.2 Limites de Confiança *Bootstrap* Padrão

- É o método mais **simples** para obtenção de limites de confiança via *bootstrap*;
- O princípio é que se o estimador de interesse  $\hat{\theta}$  é normalmente distribuído, com media  $\theta$  e desvio padrão  $\sigma$ , então há uma probabilidade  $1-\alpha$  que a afirmação abaixo seja **verdadeira**:

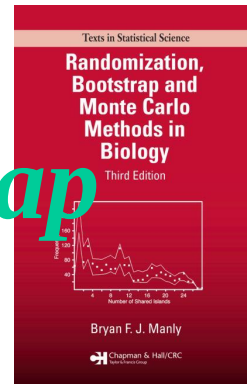
$$\theta - z_{\alpha/2} \sigma < \hat{\theta} < \theta + z_{\alpha/2} \sigma,$$

em que  $z_{\alpha/2}$  é o percentil  $(1-\alpha/2)$  da distribuição normal padrão;

- **Equivalentemente**,

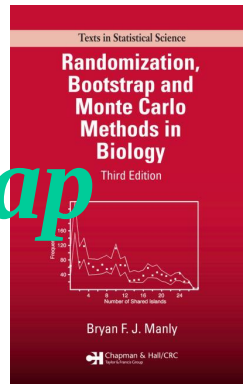
$$\hat{\theta} - z_{\alpha/2} \sigma < \theta < \hat{\theta} + z_{\alpha/2} \sigma$$

## 3.2 Limites de Confiança *Bootstrap* Padrão (cont.)



- O limite de confiança *bootstrap* padrão utiliza reamostragens com reposição, para estimação do desvio padrão do estimador, denominado **desvio padrão *bootstrap***;
- Genericamente, um intervalo de confiança *bootstrap* padrão de  $(1-\alpha)100\%$  é **dado** por:  
estimativa  $\pm z_{\alpha/2} \times (\text{desvio padrão } *bootstrap*)$
- Por **exemplo**, utilizando-se  $z_{\alpha/2} = 1,96$ , tem-se um intervalo de confiança *bootstrap* de 95%.

## 3.2 Limites de Confiança *Bootstrap* Padrão (cont.)



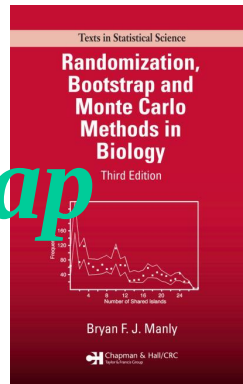
- Os requisitos para que estes limites de confiança funcionem são:
  1. O estimador  $\hat{\theta}$  tem uma distribuição **aproximadamente normal**;
  2. O estimador  $\hat{\theta}$  é **não viciado**, ou seja, seu valor médio é  $\theta$ , para repetidas amostras da população de interesse;
  3. A reamostragem *bootstrap* fornece uma **boa aproximação** para o desvio padrão  $\sigma$ .
- O requisito 2 pode ser evitado, pela **correção** do vício durante o *bootstrap*:

$$\text{vício} = \hat{\theta}_B - \hat{\theta};$$

$$\hat{\theta}_{\text{corrigido}} = \hat{\theta} - \text{vício} = 2\hat{\theta} - \hat{\theta}_B$$



## 3.2 Limites de Confiança *Bootstrap* Padrão (cont.)



### Exemplo 3.2.1:

- Para a estimação do desvio padrão populacional, suponha que seja extraída uma amostra aleatória de tamanho 20 de uma certa população:

3,56; 0,69; 0,10; 1,84; 3,93; 1,25; 0,18; 1,13; 0,27; 0,50; 0,67; 0,01; 0,61; 0,82; 1,70; 0,39; 0,11; 1,20; 1,21; 0,72

- Suponha que você decida utilizar o seguinte estimador para o desvio padrão (que é sabidamente viciado):

$$\hat{\sigma} = \sqrt{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 / 20 \right]}$$

- Calcular os limites de confiança *bootstrap* padrão de 95%.

# 3.2 Limites de Confiança *Bootstrap* Padrão (cont.)

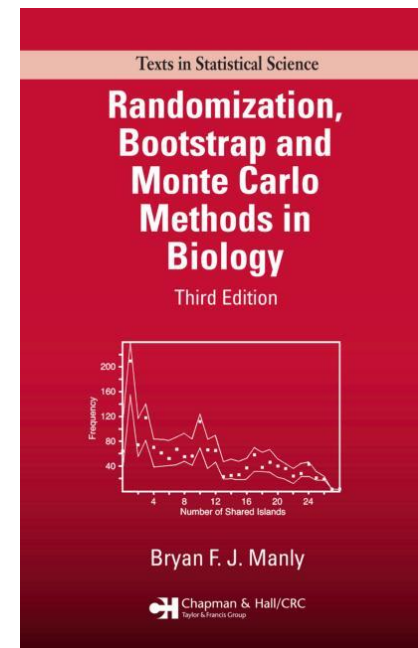
## Exemplo 3.2.1 (cont.):

|                    | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13                                                              | 14   | 15   | 6                          | 17   | 18   | 19   | 20    | desvio padrão |
|--------------------|------|------|------|------|------|------|------|------|------|------|------|------|-----------------------------------------------------------------|------|------|----------------------------|------|------|------|-------|---------------|
| amostra original   | 3,56 | 0,69 | 0,10 | 1,84 | 3,93 | 1,25 | 0,18 | 1,13 | 0,27 | 0,50 | 0,67 | 0,01 | 0,61                                                            | 0,82 | 1,70 | 0,39                       | 0,11 | 1,20 | 1,21 | 0,72  | 1,033         |
| amostras bootstrap |      |      |      |      |      |      |      |      |      |      |      |      |                                                                 |      |      |                            |      |      |      |       |               |
| 1                  | 0,10 | 1,20 | 0,10 | 1,21 | 1,13 | 0,18 | 0,69 | 0,72 | 1,20 | 0,10 | 0,72 | 1,20 | 0,39                                                            | 0,01 | 0,01 | 0,11                       | 1,84 | 1,70 | 1,13 | 1,84  | 0,633         |
| 2                  | 0,39 | 0,39 | 0,82 | 1,13 | 3,93 | 1,25 | 0,18 | 0,11 | 0,18 | 0,67 | 1,25 | 0,72 | 0,10                                                            | 3,93 | 0,01 | 0,67                       | 1,84 | 1,13 | 1,70 | 1,84  | 1,120         |
| 3                  | 0,50 | 1,20 | 1,84 | 0,69 | 3,56 | 0,18 | 0,11 | 3,56 | 0,39 | 1,25 | 0,27 | 1,20 | 3,93                                                            | 1,25 | 0,72 | 0,50                       | 0,01 | 3,56 | 0,50 | 0,61  | 1,300         |
| 4                  | 0,11 | 1,70 | 0,67 | 0,27 | 0,01 | 0,67 | 0,10 | 0,82 | 0,69 | 3,93 | 1,84 | 0,50 | 1,21                                                            | 0,39 | 1,20 | 0,61                       | 1,21 | 3,56 | 1,21 | 0,01  | 1,070         |
| 5                  | 0,39 | 0,69 | 0,82 | 1,13 | 0,01 | 1,25 | 0,18 | 1,25 | 0,82 | 3,56 | 0,72 | 1,13 | 0,69                                                            | 0,39 | 0,69 | 3,93                       | 1,84 | 1,21 | 3,93 | 0,82  | 1,170         |
| ⋮                  |      |      |      |      |      |      |      |      |      |      |      |      |                                                                 |      |      |                            |      |      |      |       |               |
| 996                | 3,56 | 1,13 | 0,82 | 0,01 | 0,01 | 1,84 | 0,27 | 1,70 | 0,27 | 3,93 | 0,67 | 1,13 | 0,69                                                            | 1,13 | 0,10 | 3,56                       | 1,84 | 0,27 | 0,82 | 0,69  | 1,200         |
| 997                | 1,84 | 0,69 | 0,67 | 3,93 | 0,18 | 1,70 | 0,67 | 0,72 | 0,11 | 0,27 | 0,50 | 0,69 | 1,84                                                            | 1,21 | 3,93 | 0,10                       | 0,82 | 0,50 | 0,72 | 3,56  | 1,220         |
| 998                | 0,82 | 1,84 | 0,10 | 0,67 | 3,56 | 0,67 | 1,70 | 0,72 | 1,21 | 1,20 | 3,93 | 1,25 | 0,82                                                            | 0,11 | 0,11 | 0,67                       | 0,61 | 0,18 | 0,39 | 3,93  | 1,220         |
| 999                | 1,20 | 0,01 | 3,93 | 1,21 | 0,50 | 0,18 | 0,01 | 0,72 | 0,61 | 0,01 | 3,56 | 0,69 | 0,72                                                            | 1,25 | 0,01 | 0,27                       | 1,84 | 1,70 | 0,18 | 0,39  | 1,110         |
| 1000               | 0,50 | 1,70 | 1,70 | 0,82 | 0,61 | 1,20 | 0,61 | 0,72 | 3,93 | 3,56 | 0,69 | 3,56 | 0,82                                                            | 0,50 | 0,18 | 1,84                       | 0,10 | 3,93 | 0,67 | 0,27  | 1,300         |
|                    |      |      |      |      |      |      |      |      |      |      |      |      |                                                                 |      |      |                            |      |      |      |       |               |
|                    |      |      |      |      |      |      |      |      |      |      |      |      | Estimativa bootstrap (média das estimativas)                    |      |      |                            |      |      |      |       | 0,978         |
|                    |      |      |      |      |      |      |      |      |      |      |      |      | Desvio padrão bootstrap (desvio padrão das estimativas)         |      |      |                            |      |      |      |       | 0,241         |
|                    |      |      |      |      |      |      |      |      |      |      |      |      | Estimativa do vício (estimativa bootstrap - estimativa amostra) |      |      |                            |      |      |      |       | -0,055        |
|                    |      |      |      |      |      |      |      |      |      |      |      |      | Estimativa corrigida (estimativa amostra - estimativa vício)    |      |      |                            |      |      |      |       | 1,087         |
| ●                  |      |      |      |      |      |      |      |      |      |      |      |      |                                                                 |      |      | Limite inferior do IC(95%) |      |      |      | 0,614 |               |
|                    |      |      |      |      |      |      |      |      |      |      |      |      |                                                                 |      |      | Limite superior do IC(95%) |      |      |      | 1,561 |               |

# 3

## Métodos *Bootstrap*

- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3ª ed., Chapman & Hall/CR Press, Boca Raton, 2007.



---

### ESQUEMA DO CAPÍTULO

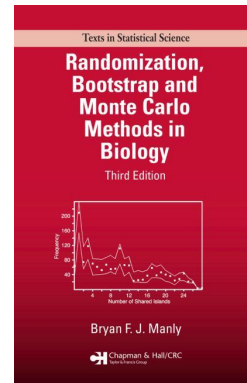
3.1 REAMOSTRAGEM COM REPOSIÇÃO

3.2 LIMITES DE CONFIANÇA BOOTSTRAP  
PADRÃO

3.3 LIMITES DE CONFIANÇA PERCENTÍLICOS  
SIMPLES

3.4 OBSERVAÇÕES FINAIS

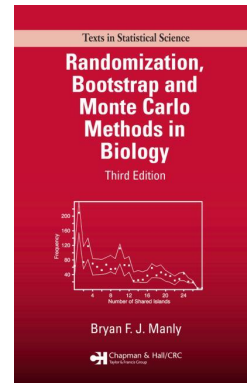
## 3.3 Limites de Confiança Percentílicos Simples



- **Várias** das abordagens para construção de intervalos de confiança são baseadas em tentativas de aproximar os percentis da distribuição de um estimador pelos percentis gerados por *bootstrap*;
- Nessa linha, Efron (1979) desenvolveu um método **simples** para cálculo dos limites de confiança;
- Assim, os limites de confiança de  $(1-\alpha)100\%$  para um determinado parâmetro são meramente aqueles  $(1-\alpha)100\%$  valores **centrais**, das estimativas obtidas via *bootstrap* na amostra original.

## 3.3 Limites de Confiança

### Percentílicos Simples (cont.)



- Suponha que o problema seja encontrar o **intervalo de confiança** de  $(1-\alpha)100\%$  para um parâmetro  $\theta$ , para o qual é conhecida uma estimativa  $\hat{\theta}$ , baseada em uma amostra aleatória  $x_1, x_2, \dots, x_n$ , de uma população de interesse;
- Suponha que exista uma função crescente monótona, tal que os valores transformados  $f(\hat{\theta})$  sejam **normalmente** distribuídos com média em  $f(\theta)$  e desvio padrão igual a 1;
- Nesse caso, a afirmação a seguir é **verdadeira**, com probabilidade  $1-\alpha$ :

$$f(\theta) - z_{\alpha/2} < f(\hat{\theta}) < f(\theta) + z_{\alpha/2}$$

## 3.3 Limites de Confiança Percentílicos Simples (cont.)

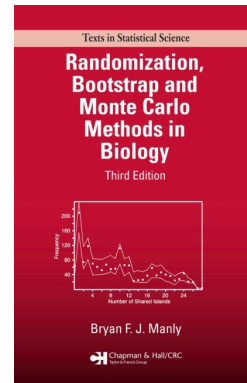
- Rearranjando os termos, segue que

$$f(\hat{\theta}) - z_{\alpha/2} < f(\theta) < f(\hat{\theta}) + z_{\alpha/2}$$

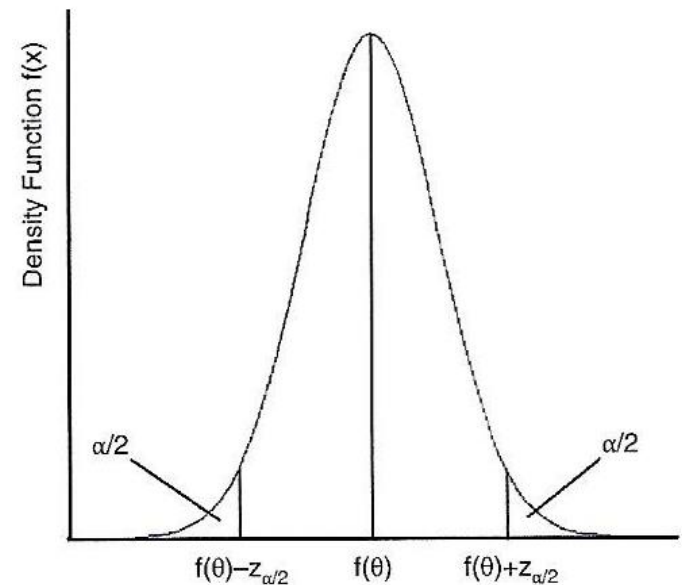
que é o **intervalo de confiança** de  $(1-\alpha)100\%$  para  $f(\theta)$ ;

- Se a função de transformação  $f$  é conhecida, então os limites do intervalo podem ser transformados via função inversa  $f^{-1}$  de forma a se obter um intervalo de confiança para  $\theta$  com o **mesmo** nível de confiança  $(1-\alpha)100\%$ ;
- O **problema** é que a função  $f$  nem sempre é conhecida.

## 3.3 Limites de Confiança Percentílicos Simples (cont.)



- Entretanto, a distribuição do estimador transformado  $f(\hat{\theta}_B)$  pode ser gerada via reamostragem *bootstrap* dos dados originais, com a acurácia **desejada**;
- Tal distribuição deve ser similar àquela apresentada na **Figura**.



**Figura:** A distribuição normal do estimador transformado, para o intervalo de confiança percentílico

## 3.3 Limites de Confiança

### Percentílicos Simples (cont.)

- A premissa básica do método é que a distribuição das reamostras *bootstrap* minetiza a distribuição de interesse;
- Assim, a distribuição do erro em  $\hat{\theta}$ ,  

$$\varepsilon = \hat{\theta} - \theta,$$
deveria ser aproximada pela distribuição do erro obtido pela reamostragem *bootstrap*,

- Dessa forma, a distribuição *bootstrap* de  $\varepsilon_B$  pode ser gerada para encontrar os dois erros  $\varepsilon_L$  e  $\varepsilon_H$ , tais que:

$$\text{Prob}\left(\varepsilon_L < \hat{\theta}_B - \hat{\theta} < \varepsilon_H\right) = 1 - \alpha$$



## 3.3 Limites de Confiança Percentílicos Simples (cont.)

- Definindo-se

$$\varepsilon_L = \hat{\theta}_L - \hat{\theta} \text{ e } \varepsilon_H = \hat{\theta}_H - \hat{\theta}$$

em que  $\hat{\theta}_L$  é o valor que é excedido com probabilidade  $1 - \alpha/2$  na distribuição *bootstrap* e

em que  $\hat{\theta}_H$  é o valor que é excedido com probabilidade  $\alpha/2$  na distribuição *bootstrap*;

- têm-se os seguintes **limites de confiança percentílicos do tipo #1**:

$$\text{Prob}\left(\hat{\theta}_L < \theta < \hat{\theta}_H\right) = 1 - \alpha$$

## 3.3 Limites de Confiança Percentílicos Simples (cont.)

- Por outro lado, assumindo-se que a distribuição *bootstrap* dos erros é uma **boa** aproximação para a distribuição real dos erros, pode ser afirmado que

$$\text{Prob}\left(\varepsilon_L < \hat{\theta} - \theta < \varepsilon_H\right) = 1 - \alpha$$

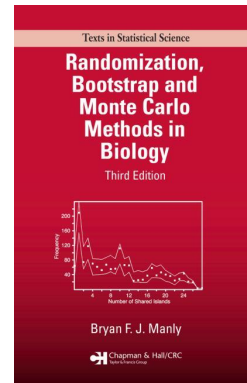
tal que

$$\text{Prob}\left(\hat{\theta} - \varepsilon_H < \theta < \hat{\theta} - \varepsilon_L\right) = 1 - \alpha$$

- têm-se então os **limites de confiança percentílicos do tipo #2**,

$$\text{Prob}\left(2\hat{\theta} - \hat{\theta}_H < \theta < 2\hat{\theta} - \hat{\theta}_L\right) = 1 - \alpha$$

## 3.3 Limites de Confiança Percentílicos Simples (cont.)



### Exemplo 3.3.1 (cont.):

- Dadas as 20 observações de uma amostra aleatória a seguir, **determinar um intervalo de confiança percentílico de 95%** para o desvio padrão,  
3,56; 0,69; 0,10; 1,84; 3,93; 1,25; 0,18; 1,13; 0,27;  
0,50; 0,67; 0,01; 0,61; 0,82; 1,70; 0,39; 0,11; 1,20;  
1,21; 0,72
- a partir do seguinte estimador **viciado** para o desvio padrão:

$$\hat{\sigma} = \sqrt{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 / 20 \right]} = 1,03$$

## 3.3 Limites de Confiança Percentílicos Simples (cont.)

### Exemplo 3.3.1

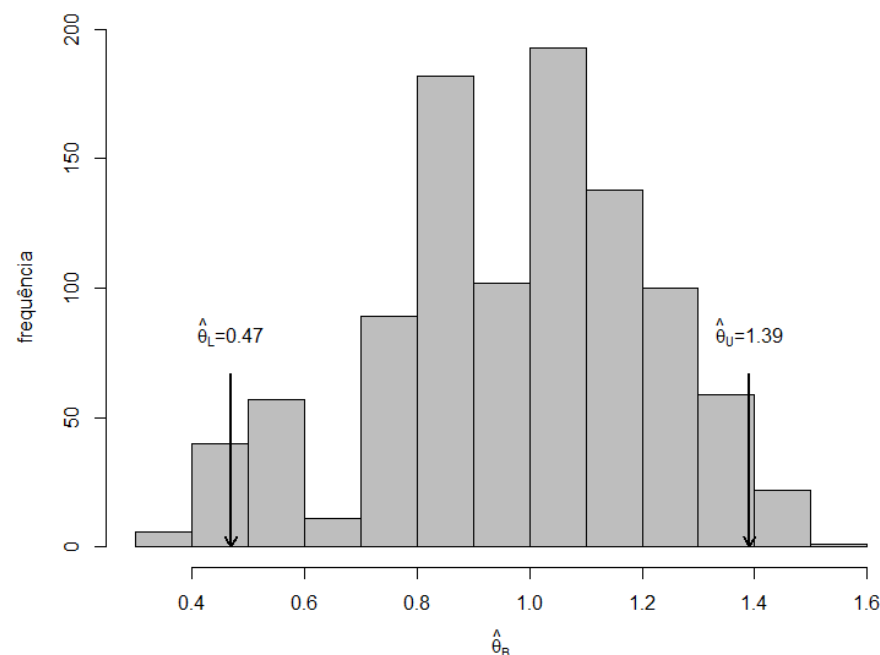
(cont.):

- Os limites de confiança percentílicos do tipo #1 são tais que a seguinte equação é satisfeita:

$$\text{Prob}(\hat{\theta}_L < \theta < \hat{\theta}_H) = 1 - \alpha$$

- o que fornece o seguinte resultado, conforme ilustrado na Figura:

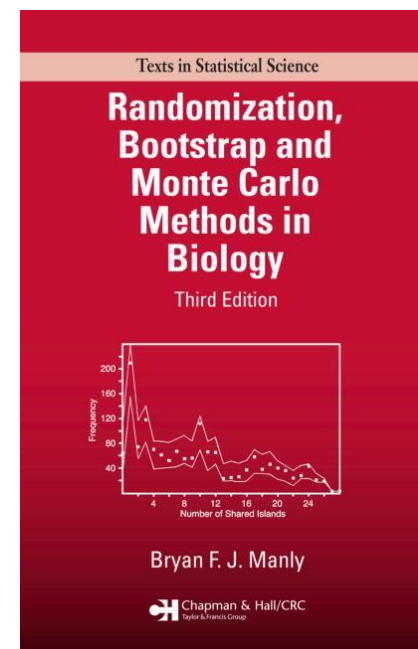
$$\hat{\theta}_L = 0,47 \text{ e } \hat{\theta}_H = 1,39$$



**Figura:** Histograma de estimativas para  $\sigma$ , obtidas via 1.000 amostras *bootstrap* e os percentis  $\alpha/2$  e  $1 - \alpha/2$

# 3

## Métodos *Bootstrap*



- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3<sup>a</sup> ed., Chapman & Hall/CR Press, Boca Raton, 2007.

---

### ESQUEMA DO CAPÍTULO

3.1 REAMOSTRAGEM COM REPOSIÇÃO

3.2 LIMITES DE CONFIANÇA BOOTSTRAP  
PADRÃO

3.3 LIMITES DE CONFIANÇA PERCENTÍLICOS  
SIMPLES

3.4 OBSERVAÇÕES FINAIS

## 3.4 Observações Finais

- Neste capítulo, são apresentados alguns métodos **básicos** para construção de intervalos de confiança;
- Muita **pesquisa** tem sido desenvolvida recentemente na área de bootstrap, que inclui o desenvolvimento de novos métodos para:
  - o cálculo dos limites de confiança para diversos parâmetros populacionais, bem como
  - testes de significância.