

Avaliação de imóveis com amostra de tamanho reduzido: TÉCNICAS DE REAMOSTRAGEM

Prof. Dr. Lutemberg Florencio



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia

Nivelando conceitos

Distribuição de probabilidade



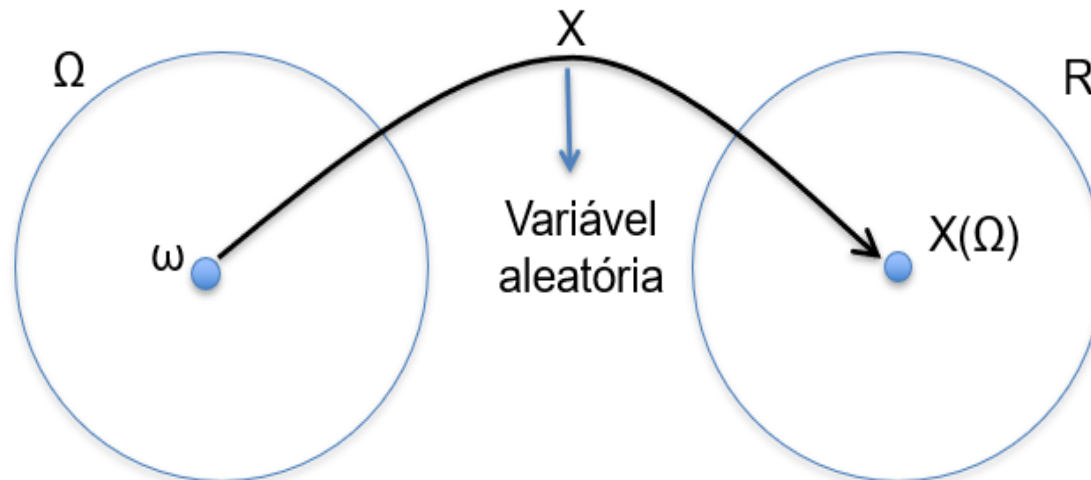
Nivelando conceitos

■ Distribuição de probabilidade

✓ Uma **distribuição de probabilidade** é uma função matemática que descreve a probabilidade de ocorrência de cada possível resultado (valor específico OU intervalo de valores) de uma **variável aleatória**.

□ DEFINIÇÃO FORMAL

✓ Seja E um experimento aleatório e Ω o espaço amostral associado ao experimento. Uma **função** X , que associe a cada resultado $\omega \in \Omega$ um número real $X(\omega)$ é denominado **variável aleatória**.



Nivelando conceitos

■ Distribuição de probabilidade

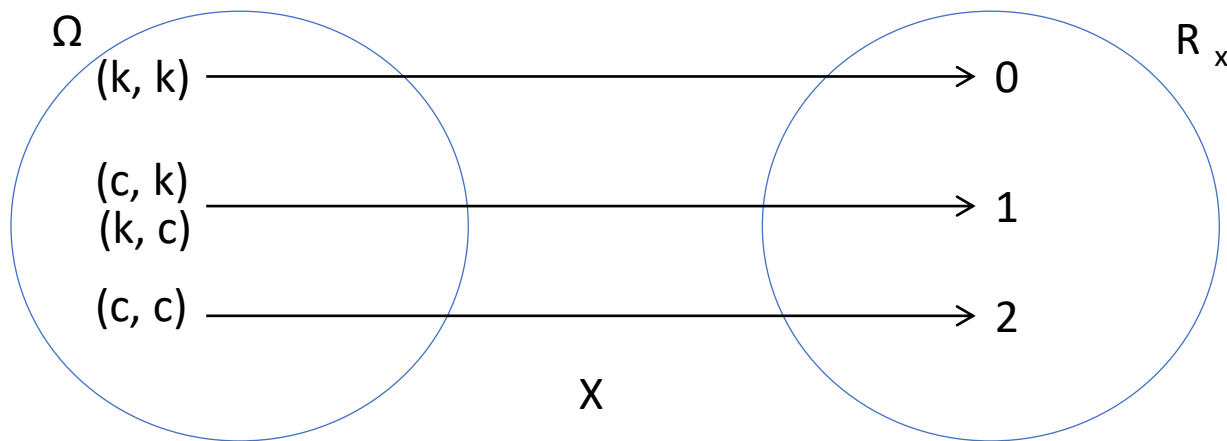
✓ **Exemplo intuitivo:** considere o lançamento de duas moedas:

Então:

E (experimento aleatório): lançamento de duas moedas;

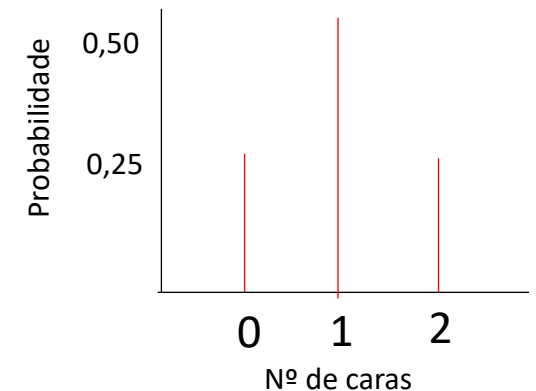
$$\Omega = \{(c, c), (c, k), (k, c), (k, k)\}$$

X (função): nº de caras obtidas nas duas moedas



Distribuição de probabilidade

x	0	1	2
p(x)	0,25	0,5	0,25



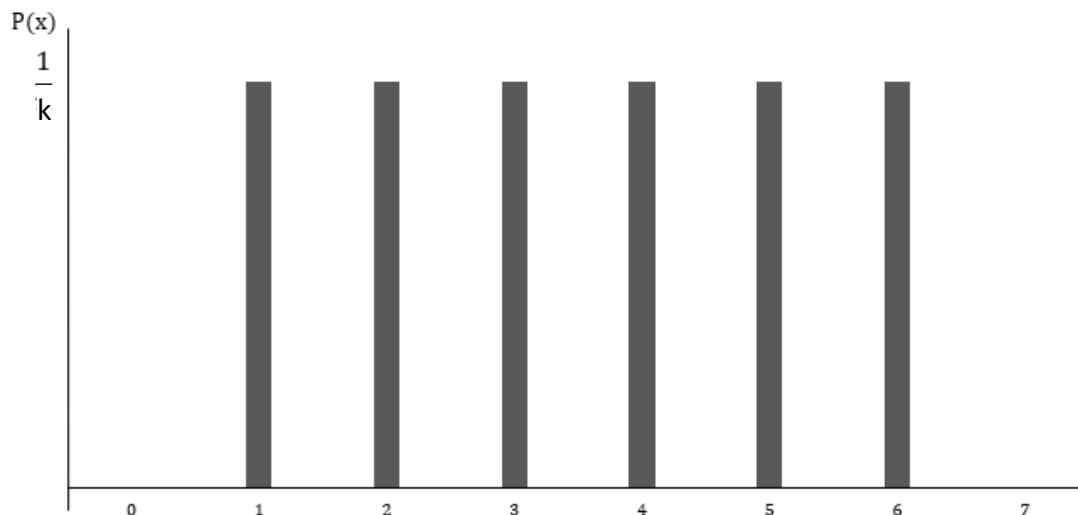
✓ **Variável aleatória:** quantidade que pode assumir diferentes valores em um **experimento aleatório**.

Nivelando conceitos

- Distribuição de probabilidade

✓ As distribuições de probabilidade podem ser representadas de duas formas:

(i) Função de probabilidade (f.p): nesse caso, a variável aleatória X é discreta, e o valores possíveis de X (sua imagem) é finito ou infinito enumerável. Ex.: **distribuição uniforme discreta** ($X \sim U_d(k)$)



I) Fórmula

$P(X=x_i) = 1/6$, para $i=1,2,3,4,5,6$.

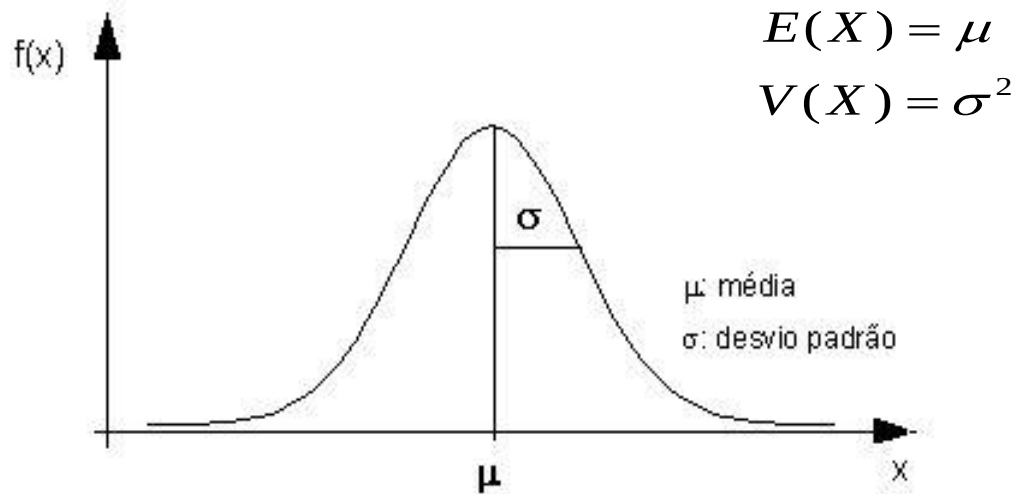
$$E(X) = \frac{1}{k} \sum_{i=1}^k x_i \rightarrow \frac{a+b}{2}$$

$$\text{VAR}(X) = \frac{1}{k} \left[\sum_{i=1}^k x_i^2 - \frac{1}{k} \left(\sum_{i=1}^k x_i \right)^2 \right]$$
$$\rightarrow \frac{(b-a+1)^2 - 1}{12}$$

Nivelando conceitos

- Distribuição de probabilidade

(ii) **Função densidade de probabilidade (f.d.p):** nesse caso, a variável aleatória X é contínua, e o valores possíveis de X (sua imagem) é um intervalo de números reais (não numerável). Ex.: **distribuição normal** $[X \sim N(\mu, \sigma^2)]$



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ para } -\infty < x < +\infty.$$

Nivelando conceitos

- Distribuição de probabilidade
- ✓ Uma **variável aleatória** é caracterizada ou descrita pela sua **distribuição de probabilidade** que, por sua vez, é descrita pelos seus **parâmetros populacionais**. Por exemplo, a média μ e o desvio padrão σ são os parâmetros populacionais da distribuição Normal;
- ✓ **PERGUNTA:** como ter uma primeira percepção acerca da **distribuição de probabilidade** de uma variável aleatória?

HISTOGRAMA

Nivelando conceitos

- Estatística descritiva – Distribuição de frequências

4	6	8	10	11	13
4	7	8	10	12	13
4	7	8	10	12	13
5	7	9	10	12	14
6	7	9	11	12	14
6	8	9	11	13	15

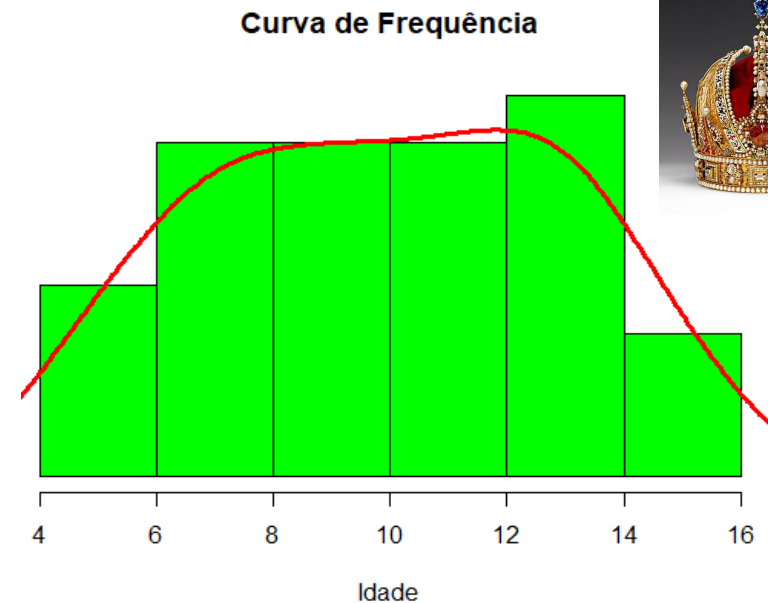
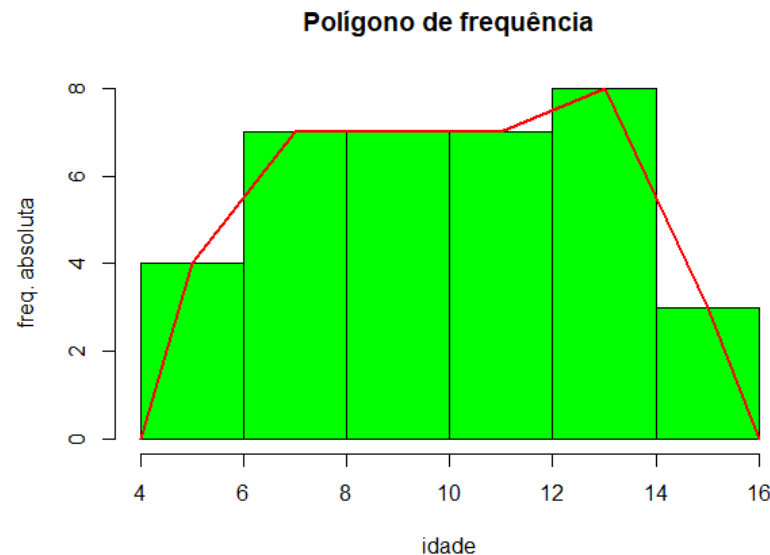
Dados agrupados com intervalos de classes

Idade	Frequência
4 ----6	4
6 ----8	7
8 ----10	7
10 ----12	7
12 ----14	8
14 ----16	3

Nivelando conceitos

■ Estatística descritiva – Gráficos estatísticos

✓ **HISTOGRAMA, POLÍGONO DE FREQUÊNCIA e CURVA DE FREQUÊNCIA:** Podemos concluir* que a linha poligonal (contorno do polígono de frequência) tende a se transformar numa curva – a **curva de frequência** –, mostrando de modo mais evidente, a natureza da **distribuição do conjunto de dados**.



ATENÇÃO: Com base no histograma, têm-se as primeiras percepções sobre:

- Distribuição de probabilidade
- Medidas resumo: posição / dispersão / assimetria / curtose

Motivação

- ✓ Qual o **PROBLEMA (DESAFIO)** de realizar a avaliação de um imóvel com base em uma amostra de tamanho reduzido?



Motivação

- ✓ Da **teoria estatística**, tem-se que a relação entre o **tamanho da amostra (n)** e o **tamanho da população (N)** influencia a precisão (**erro amostral**) das estimativas dos parâmetros populacionais.

Figura 01: Relação entre o tamanho da amostra(n) x tamanho da população(N)

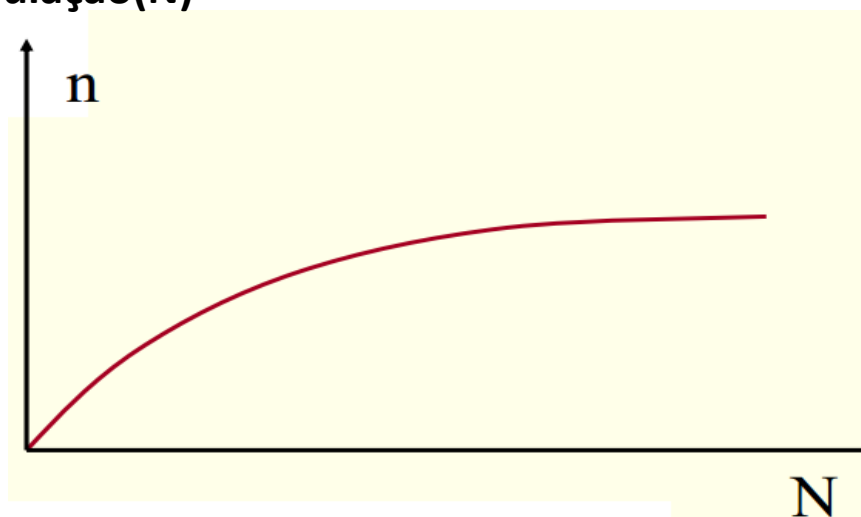
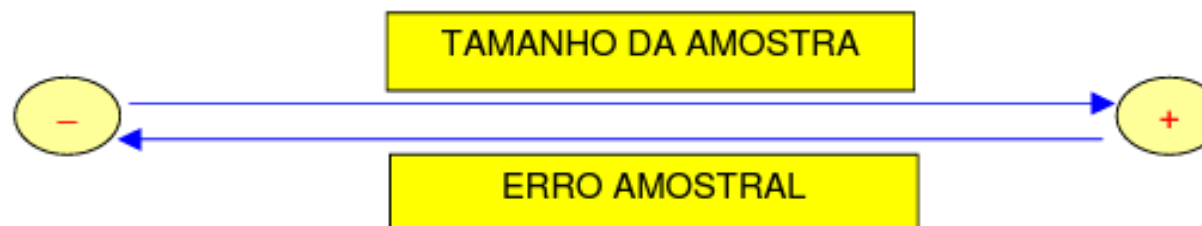


Figura 02: Relação intuitiva entre o tamanho da amostra e o erro amostral



Motivação

- ✓ Dos fundamentos da **teoria econômica**, tem-se que:

O preço do bem pode ser decomposto nos preços de suas k características, representadas por $q = (q_1, q_2, \dots, q_k)$, por meio de uma função do tipo:

$$P(q) = P(q_1, q_2, \dots, q_k).$$

Locacionais, físicas e econômicas



COMO explicar o comportamento do preço (do imóvel) em função de suas características?

Motivação

- **Modelo:** representação simplificada da realidade a partir de uma expressão matemática.

Modelos de
regressão
linear

Modelos de
regressão
espacial

Geoestatística

Métodos
semi-
paramétricos

Inteligência
artificial

Motivação

- Modelo de regressão linear

Var. dependente é expressa por uma **combinação linear nos parâmetros**, sendo dada por:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki} + \hat{e}_i$$



MODELO

Curva de regressão
f (características do
imóvel)

ATENÇÃO:
Estimação dos
parâmetros
X
NBR 14653-2/3

Motivação

- Sobre a estimação dos parâmetros

Nos **modelos de regressão linear**, usualmente os **parâmetros são estimados com base no MQO**. Dados os **pressupostos do modelo clássico de regressão linear**, os estimadores MQO são **BLUE** -- Teorema de **Gauss-Markov** --.

Estimadores MQO: **não-tendenciosidade, eficiência e consistência.**

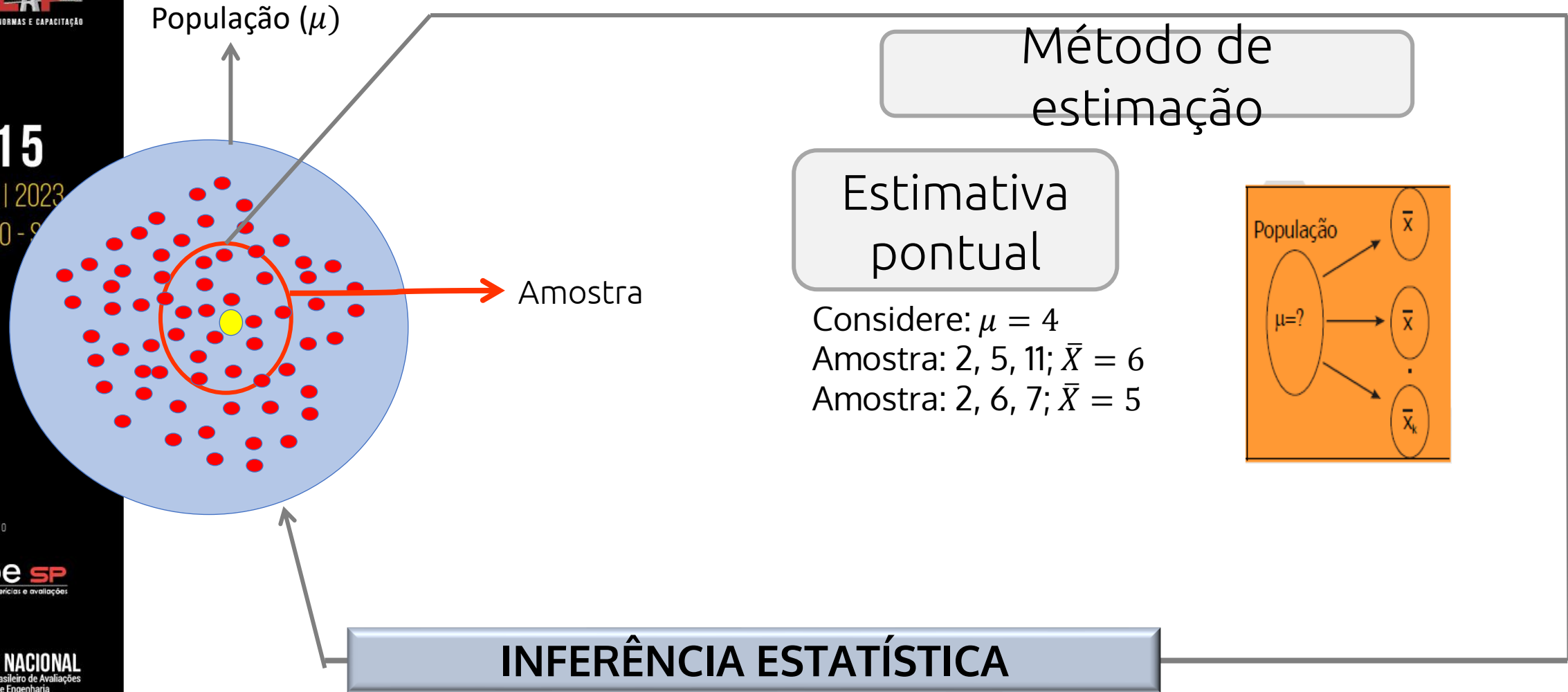
- | | |
|-------------------------------------|---|
| (i) Linearidade | ✓ O Teorema de Gauss-Markov é notável porque não faz nenhuma suposição sobre a distribuição de probabilidade do termo de erro aleatório. |
| (ii) Homocedasticidade | |
| (iii) Não-multicolinearidade | |
| (iv) Não-autocorrelação | |
| (v) Especificação correta do modelo | |

Mas por que isso pode ser um **PROBLEMA** em avaliações com amostras de tamanho reduzido?

ATENÇÃO: Adicionalmente, o estimador MQO fornecerá **estimativas pontuais**. Em que pese as excelentes propriedades do estimador MQO, a estimativa pontual é **“limitada”**... Por que?

Motivação

- Inferência Estatística – Estimação de parâmetros



Em que pese as excelentes propriedades do estimador MQO, a NBR 14653 – provavelmente preocupada com a “limitação” da estimativa pontual e confiabilidade (estatística) do modelo– exige:

Motivação

- Modelo de regressão linear
- ✓ NBR 14653-2/3

Descrição	Grau		
	III	II	I
Amplitude do intervalo de confiança de 80 % em torno da estimativa de tendência central	$\leq 30 \%$	$\leq 40 \%$	$\leq 50 \%$

5	Nível de significância a (somatório do valor das duas caudas) máximo para a rejeição da hipótese nula de cada regressor (teste bicaudal)	10 %	20 %	30 %
6	Nível de significância máximo admitido para a rejeição da hipótese nula do modelo através do teste F de Snedecor	1 %	2 %	5 %

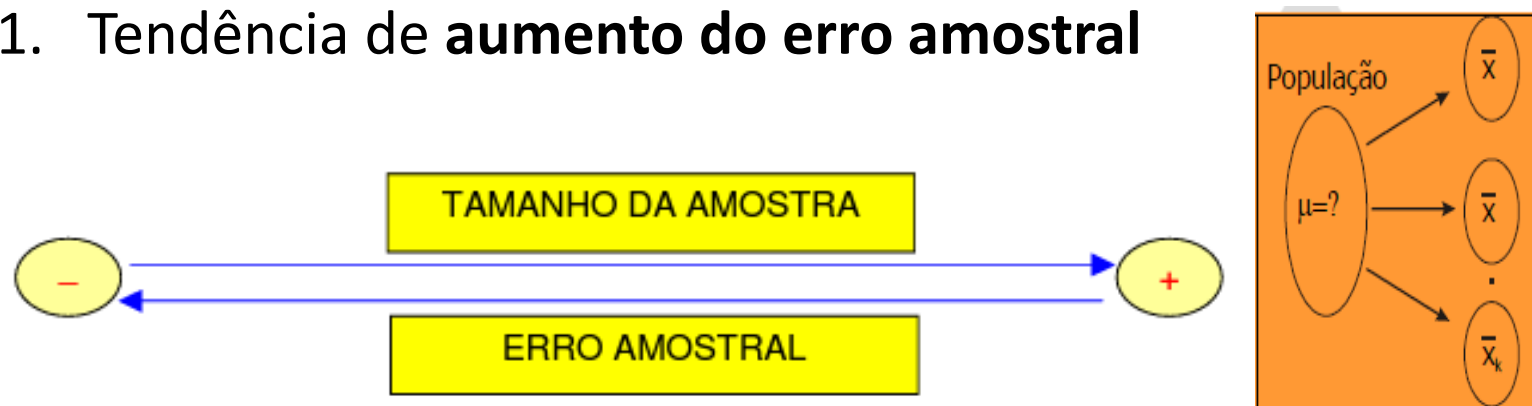
Mas por que isso pode ser um **PROBLEMA** em avaliações com amostras de tamanho reduzido?

Motivação



✓ Qual o **PROBLEMA (DESAFIOS)** de realizar a avaliação de um imóvel com base em uma amostra de tamanho reduzido?

1. Tendência de **aumento do erro amostral**

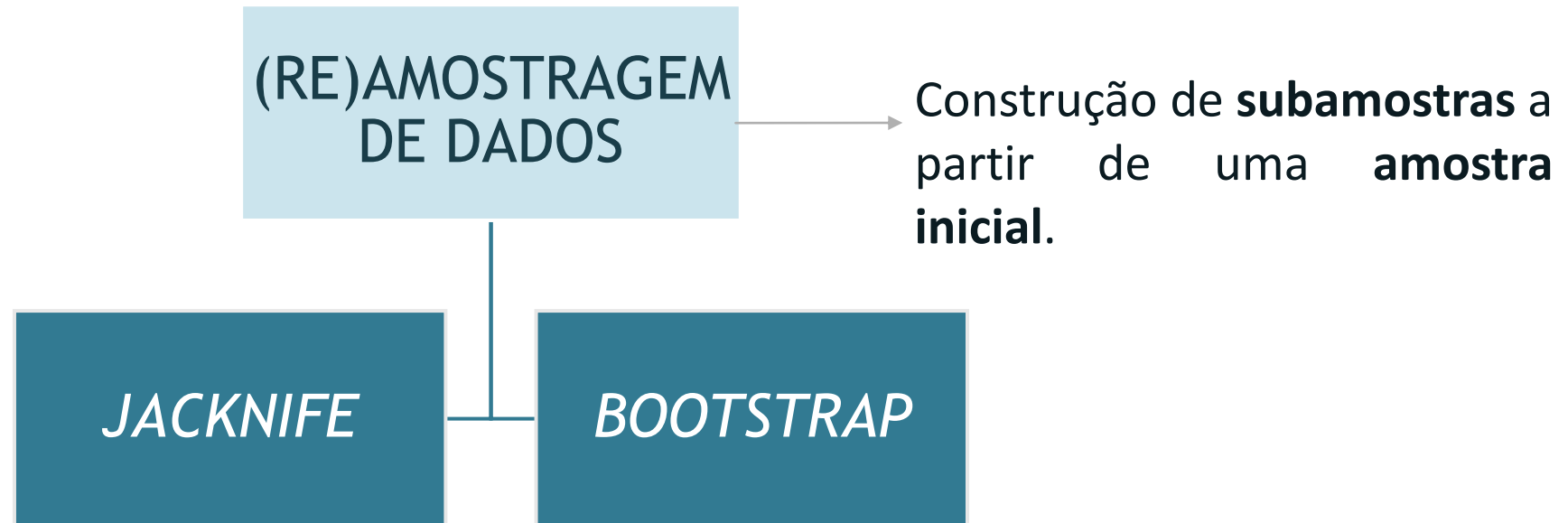


2. Para que seja válido o Teorema de Gauss-Markov – estimador MQO --, as **premissas do MRLC** precisam ser atendidas;

3. Para atender a NBR 14653 – itens 5/6 do grau de fundamentação + grau de precisão – faz-se necessário que os **resíduos tenham distribuição de probabilidade normal**.

Técnicas de reamostragem

✓ Alternativa:



Técnicas de reamostragem

✓ *Jackknife e Bootstrap*

Função de distribuição empírica

A **função de distribuição empírica** atribui probabilidade $1/T$ a cada valor amostral y_t .

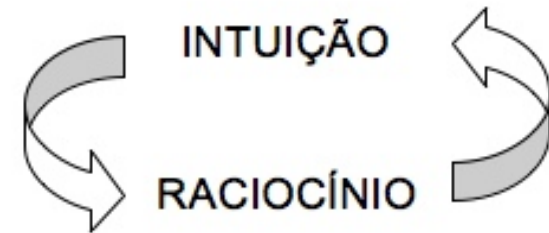
FUNÇÃO DE DISTRIBUIÇÃO EMPÍRICA (FDE):

$$\hat{F}(y) = \frac{\#\{y_t \leq y\}}{T}.$$

Note que os valores da FDE são fixos: $(0, 1/T, 2/T, \dots, T/T)$. Assim, a FDE é equivalente a seus pontos de aumento: $y_{(1)} \leq \dots \leq y_{(T)}$ (valores amostrais ordenados).

Uma vez que a FDE atribui probabilidades iguais para os valores amostrais $y_1, \dots, y_T$, **cada y_t^* é aleatoriamente amostrado desses valores.**

Portanto, a amostra bootstrap é uma amostra aleatória obtida dos dados originais (amostragem com reposição).

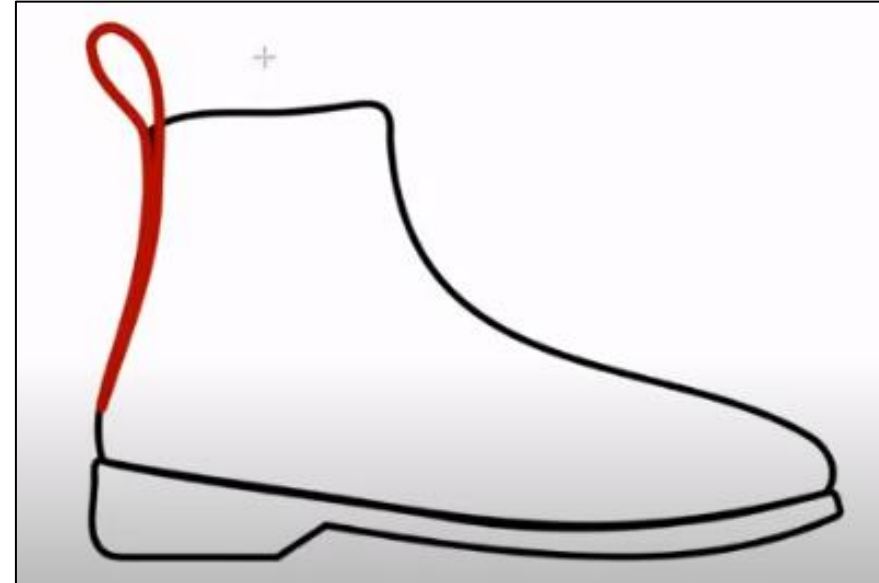


Técnicas de reamostragem

✓ *Jacknife e Bootstrap*



✓ O termo *Jacknife* sugere a ideia de que na ausência de uma ferramenta melhor (mais específica), use o canivete (suíço).

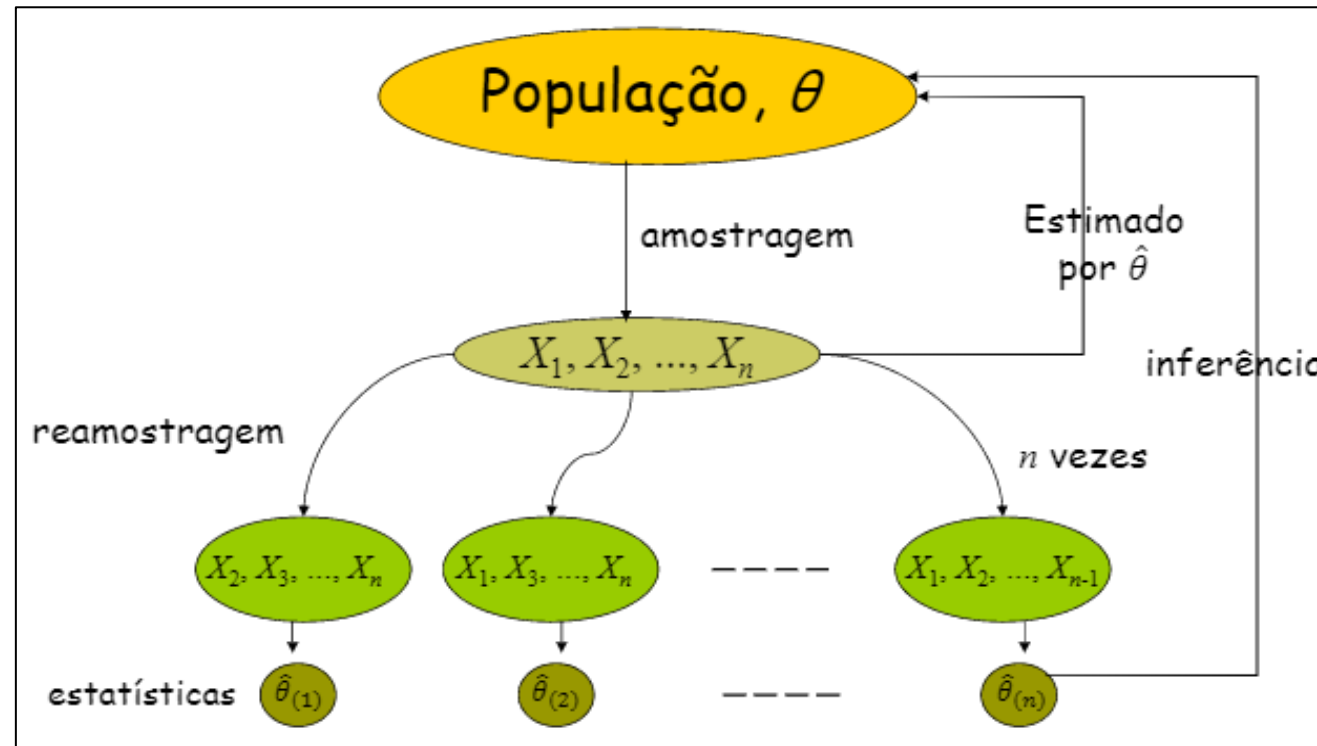


✓ O termo *bootstrap* sugere a ideia de que é possível emergir de um afogamento puxando pelo cadarço do próprio sapato, ou seja, utilizando-se do próprio esforço.

Técnicas de reamostragem

Jackknife

✓ O método *jackknife* é um método inferencial baseado em reamostragem de dados, cuja ideia central é tratar os dados como se fossem a população e amostrar dos dados (de forma **determinística**, cada amostra de tamanho $T-1$) como se estivéssemos amostrando da população e calcular, para cada amostra artificial, a quantidade de interesse.



Fonte: Notas de aula (Prof. Camil Rennó)

Técnicas de reamostragem *Jackknife*

- ✓ O método jackknife foi proposto por Maurice Quenouille (1949) e posteriormente aprimorado por John Tukey (1956).
- ✓ **Limitação:** Podemos ter no máximo T amostras artificiais.

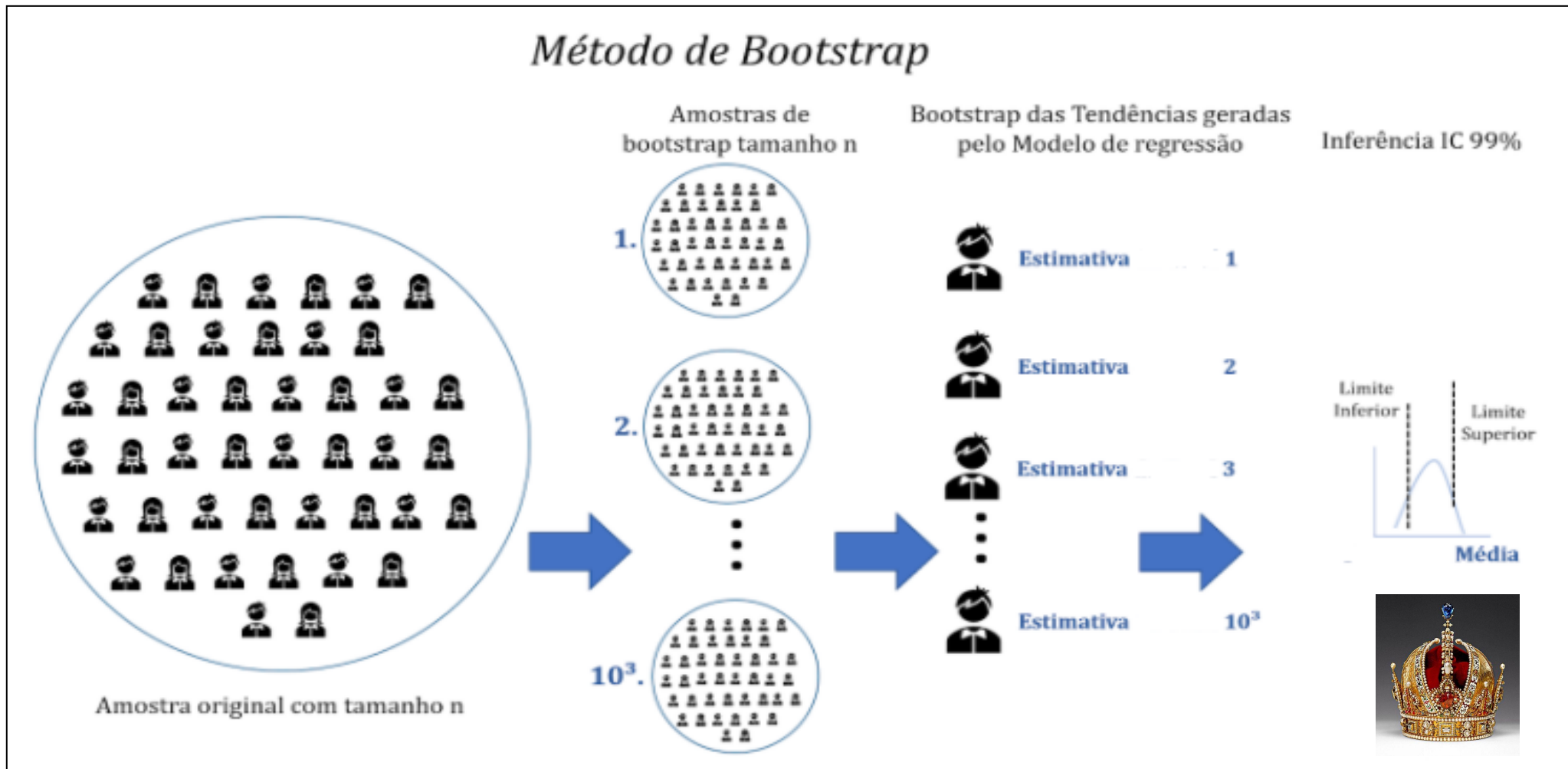
Técnicas de reamostragem

Bootstrap

- ✓ O método *bootstrap* generaliza o método *jackknife* no sentido que substitui o mecanismo determinístico de criação de amostras artificiais por um mecanismo aleatório.
- ✓ *Bootstrap* é um método inferencial baseado em reamostragem de dados, cuja ideia central é tratar os dados como se fossem a população e amostrar dos dados (de forma aleatória, B vezes) como se estivéssemos amostrando da população e calcular, para cada amostra artificial, a quantidade de interesse. Ao final, usar os B valores da quantidade de interesse para estimar sua distribuição de probabilidade.

Técnicas de reamostragem

Bootstrap



Fonte: <https://hupdata.com/adocao-lockdown/>

11^A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

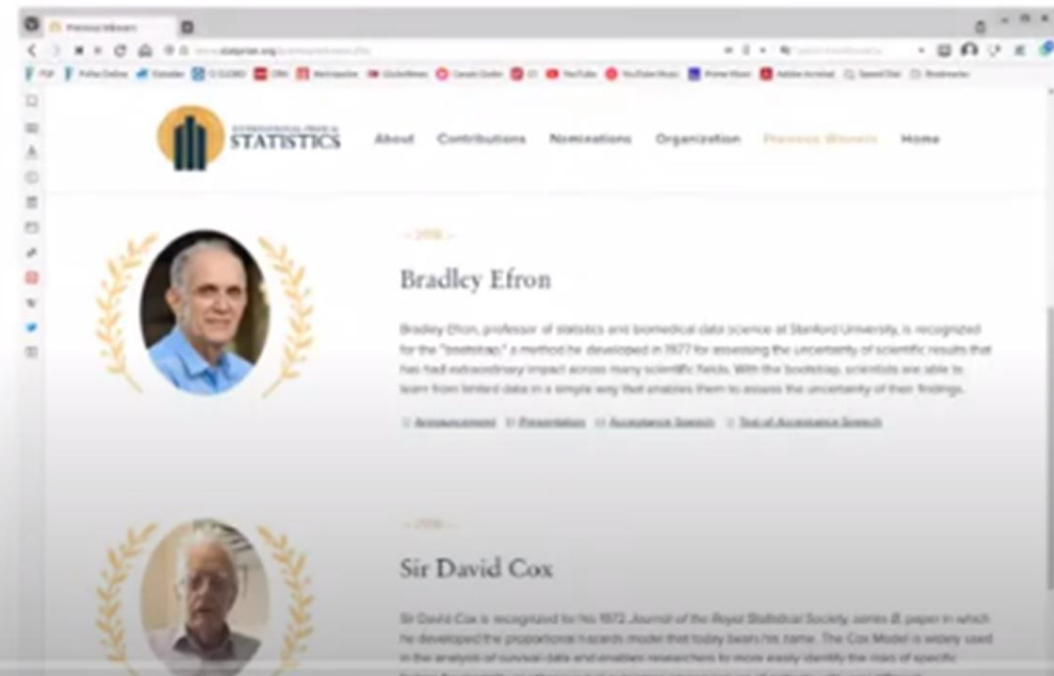
Técnicas de reamostragem *Bootstrap*

[1] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7: 1-25.



Bradley Efron

Vencedores do Prêmio IPS: David Cox (2016), Bradley Efron (2018), Nan Laird (2020); ver <https://www.statprize.org>.



Técnicas de reamostragem *Bootstrap*

- ✓ Como o processo de reamostragem -- com base no método bootstrap -- pode contribuir com as avaliações imobiliárias?

No tratamento dos dados podem ser utilizados, alternativamente e em função da qualidade e da quantidade de dados e informações disponíveis:

- tratamento por fatores: homogeneização por fatores e critérios, fundamentados por estudos conforme 8.2.1.4.2, e posterior análise estatística dos resultados homogeneizados.
- tratamento científico: tratamento de evidências empíricas pelo uso de metodologia científica que leve à indução de modelo validado para o comportamento do mercado.

Técnicas de reamostragem

Bootstrap

✓ Caso prático:

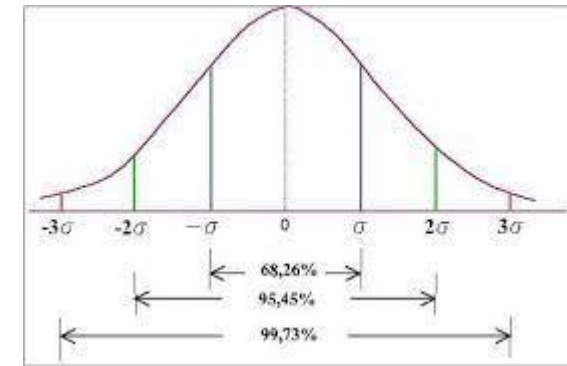
Com base na pesquisa de mercado abaixo estime o valor de um apartamento com área privativa de 100 m² e calcule o intervalo de confiança (com grau de confiança de 80%):

DADO	ÁREA CONST (m ²)	PREÇO TOTAL (R\$ mil)
1	70	522
2	80	572
3	90	785
4	110	764
5	120	1265
6	130	1442

✓ Modelo ajustado pelo MMQ:

$$VI = -581,42 + 14,73 \cdot AP$$

$$VI = 892,14$$



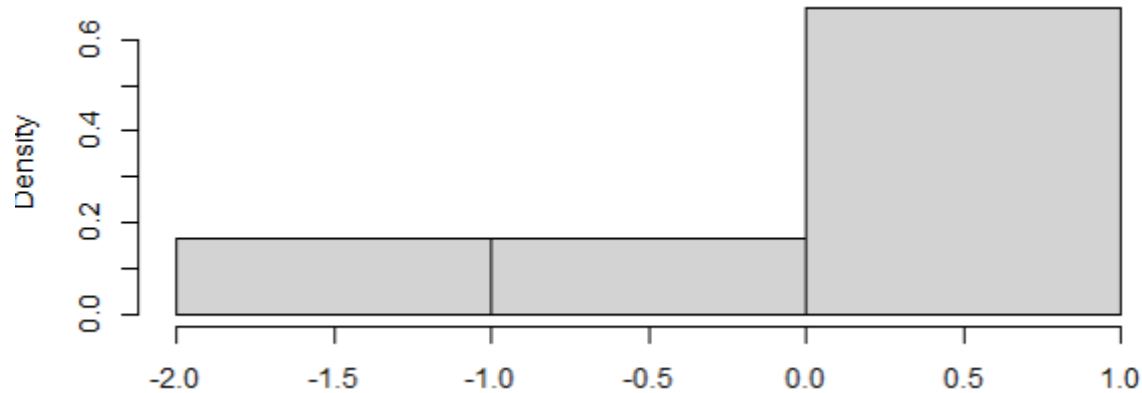
Resíduos

Agora vamos calcular o IC

Técnicas de reamostragem

Bootstrap

✓ Histograma dos resíduos:



✓ Frequência relativa dos resíduos amostrais padronizados nos intervalos

de $[-1; +1]$, $[-1,64; +1,64]$ e $[-1,96; +1,96]$ • 83% • 83% • 100%

Script:

shapiro.test (modelo)

W = 0.76696, p-value = 0.02904



E agora?

E agora?

Técnicas de reamostragem *Bootstrap*

✓ BOOTSTRAP:

NOSSA FERRAMENTA: O ambiente computacional R; ver
<https://www.r-project.org>.

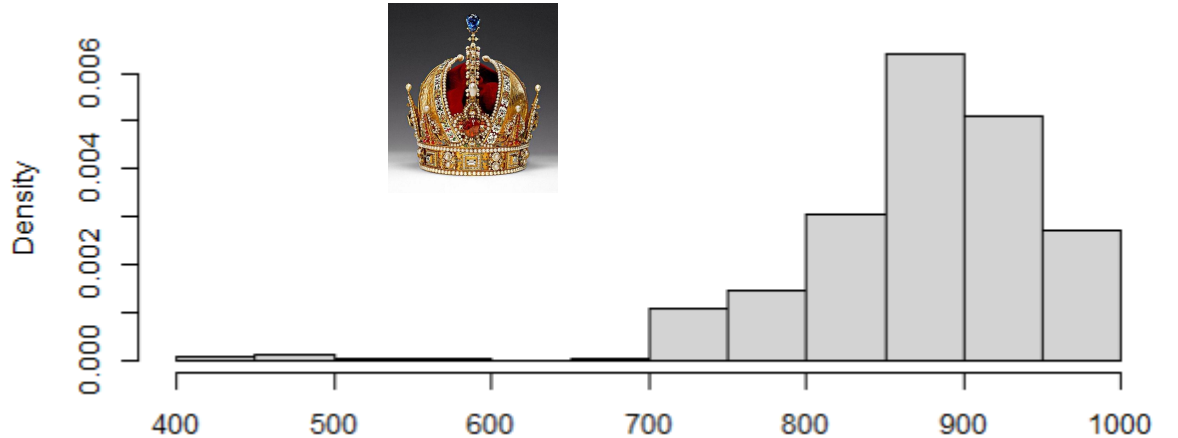
1. Gere 1000 amostras de tamanho 06 (seis) – a partir do conjunto de dados original;
2. Ajuste o modelo de regressão para cada amostra *bootstrap* e registre os coeficientes estimados, bem como a estimativa (média) do valor do avaliando;
3. Use gráficos, como histogramas, para visualizar a distribuição dos coeficientes estimados, bem como da estimativa (média) do valor do avaliando;
4. Calcule os intervalos de confiança para entender a variabilidade dos coeficientes estimados e da estimativa (média) do valor do avaliando.

Técnicas de reamostragem

Bootstrap

✓ BOOTSTRAP: Estimativa (média) do avaliando | Histograma e Interv. Confiança

1	2
6 130 1442	5 120 1265
4 110 764	4 110 764
5 120 1265	6 130 1442
6.1 130 1442	5.1 120 1265
3 90 785	6.1 130 1442
2 80 572	5.2 120 1265
[1,] 850.3937	[990,] 875.1594
[2,] 892.9922	[991,] 881.5353
[3,] 966.7960	[992,] 874.6126
[4,] 852.7985	[993,] 839.6896
[5,] 955.9155	[994,] 957.5167
[6,] 815.5183	[995,] 875.1594
[7,] 845.9412	[996,] 886.1419
[8,] 897.7570	[997,] 727.8841
[9,] 954.3175	[998,] 895.7950
[10,] 937.3536	[999,] 903.8616
	[1000,] 857.8391



$$\bar{X} = 875,03$$

Intervalo de confiança (80%):
[768,68; 955,92]

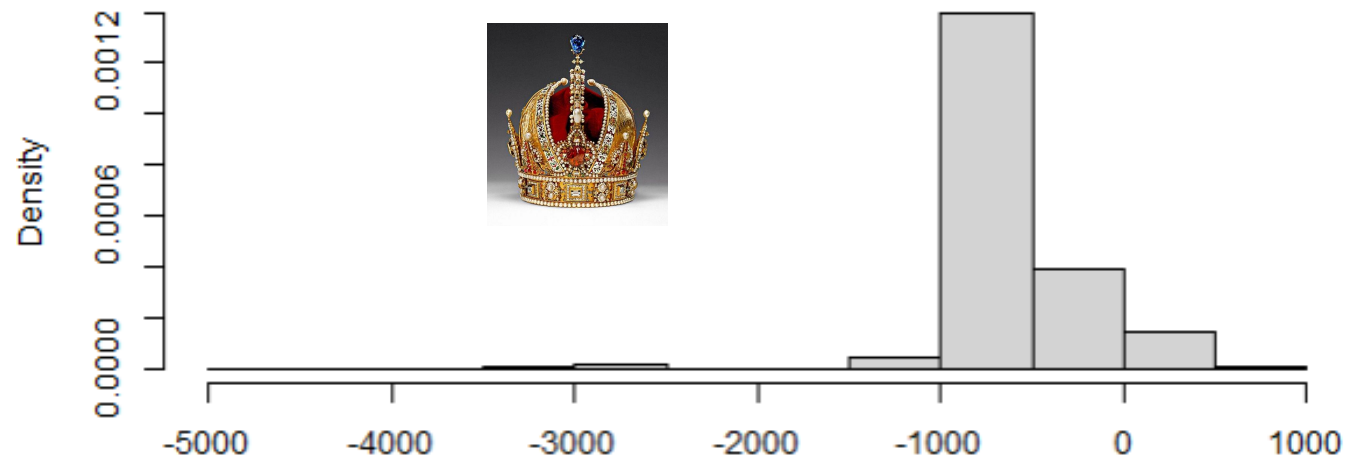
Grau de precisão III

E sobre os coeficientes (parâmetros) do modelo?

Técnicas de reamostragem

Bootstrap

✓ BOOTSTRAP: Intercepto | Histograma e Interv. Confiança



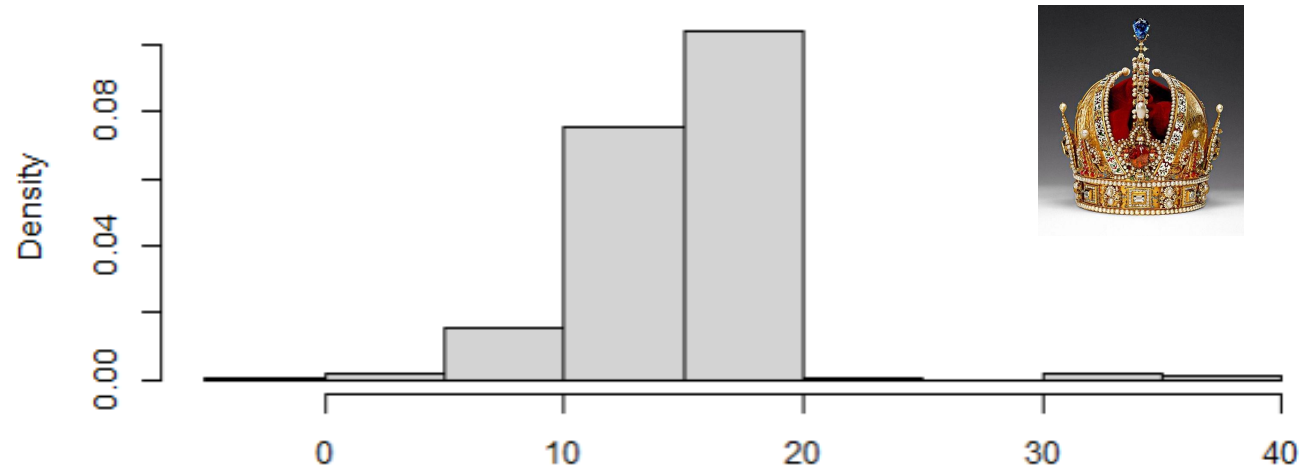
$$\widehat{\beta}_0 = -572,97$$

Intervalo de confiança (80%):
[-830,72; -300,98]

Técnicas de reamostragem

Bootstrap

✓ BOOTSTRAP: Coef angular | Histograma e Interv. Confiança



$$\hat{\beta}_1 = 14,48$$

Intervalo de confiança (80%):
[11,26; 17,37]

Conclusões

✓ O método *bootstrap* é potencialmente viável para uso no contexto das avaliações imobiliárias, sobretudo nos casos em que o número de dados é reduzido e a distribuição de probabilidade é desconhecida, possibilita-se a **estimação pontual e por intervalo de diversos parâmetros de interesse, além da análise da significância dos respectivos parâmetros. Ainda:**

- (i) Cálculo do IC para outros parâmetros, como a moda, mediana ou percentis;
- (ii) Realização de inferências sem a necessidade de verificação dos pressupostos do modelo;
- (iii) Realização de estimativas sob a presença de dados atípicos (legítimos).



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



“Precisamos considerar se as suposições feitas pelo pesquisador são apropriadas aos dados e ao problema. Com demasiada frequência, a pesquisa é baseada em suposições implícitas sobre o problema e os dados -- que provavelmente não estão corretas -- e que produzem estimativas com base nessas premissas”. (GUJARATI, 2004).

Obrigado!

✉ lutembergflorencio@yahoo.com.br

📷 [prof.lutembergflorencio](https://www.instagram.com/prof.lutembergflorencio)

📞 (81) 98886-1406

