Crítica ao uso do Campo de Arbítrio do Avaliador

em situações de escassez de dados de mercado

Luiz Fernando Palin Droubi[a]
Carlos Augusto Zilli[b]
Willian Zonato[c]
Norberto Hochheim[d]

25 de novembro de 2020

GEAP - UFSC

[a] lfpdroubi@gmail.com
[b] carlos.zilli@ifsc.edu.br
[c] will.zonato@gmail.com
[d] norberto.hochheim@ufsc.br

Regressão Linear

Por que arbitrar um valor diferente da média condicional?

Estudos de casos

Conclusões

Regressão Linear

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
    - P. ex.:
        - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
    - P. ex.:
        - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
        - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
        - Se $\varepsilon \approx N(0, \sigma^2)$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ e $\hat{\varepsilon}$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ e $\hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ e $\hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2 \, e \, \hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões
  - P. ex.: Qual o *valor unitário médio* (ou esperado) de um lote, dado que ele possui $A = 360m^2$ e $F = 12m$?

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ e $\hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões
  - P. ex.: Qual o *valor unitário médio* (ou esperado) de um lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \mathbb{E}[\hat{\varepsilon}]$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ e $\hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões
  - P. ex.: Qual o *valor unitário médio* (ou esperado) de um lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \mathbb{E}[\hat{\varepsilon}]$
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\ e\ \hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões
  - P. ex.: Qual o *valor unitário médio* (ou esperado) de um lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \mathbb{E}[\hat{\varepsilon}]$
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12$
    - Incerteza do estimador deve ser verificada com IC

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\, e\, \hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões
  - P. ex.: Qual o *valor unitário médio* (ou esperado) de um lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \mathbb{E}[\hat{\varepsilon}]$
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12$
    - Incerteza do estimador deve ser verificada com IC
  - Mas qual o valor de um novo lote, dado que ele possui $A = 360m^2$ e $F = 12m$?

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
    - P. ex.:
        - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
        - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
        - Se $\varepsilon \approx N(0, \sigma^2)$
        - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
    - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ e $\hat{\varepsilon}$
    - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões
    - P. ex.: Qual o *valor unitário médio* (ou esperado) de um lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
        - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \mathbb{E}[\hat{\varepsilon}]$
        - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12$
        - Incerteza do estimador deve ser verificada com IC
    - Mas qual o valor de um novo lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
        - $\hat{VU} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \hat{\varepsilon}$

Definição

- A regressão linear é uma função que *estima* o valor da *média condicional* de uma população (MATLOFF, 2017, p. 10–11).
  - P. ex.:
    - $VU = \beta_0 + \beta_1 A + \beta_2 F + \varepsilon$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F + \mathbb{E}[\varepsilon]$
    - Se $\varepsilon \approx N(0, \sigma^2)$
    - $\mu(VU) = \mathbb{E}[VU|A, F] = \beta_0 + \beta_1 A + \beta_2 F$
  - São estimados $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2 \, e \, \hat{\varepsilon}$
  - Caso o modelo seja bem especificado, $\hat{\varepsilon}^2 \approx \sigma^2$
- A equação estimada pode ser utilizada para fazer estimativas e previsões
  - P. ex.: Qual o *valor unitário médio* (ou esperado) de um lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \mathbb{E}[\hat{\varepsilon}]$
    - $\mathbb{E}[VU|A = 360, F = 12] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12$
    - Incerteza do estimador deve ser verificada com IC
  - Mas qual o valor de um novo lote, dado que ele possui $A = 360m^2$ e $F = 12m$?
    - $\hat{VU} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 360 + \hat{\beta}_2 \cdot 12 + \hat{\varepsilon}$
    - Incerteza na predição deve ser verificada com IP

Função média condicional

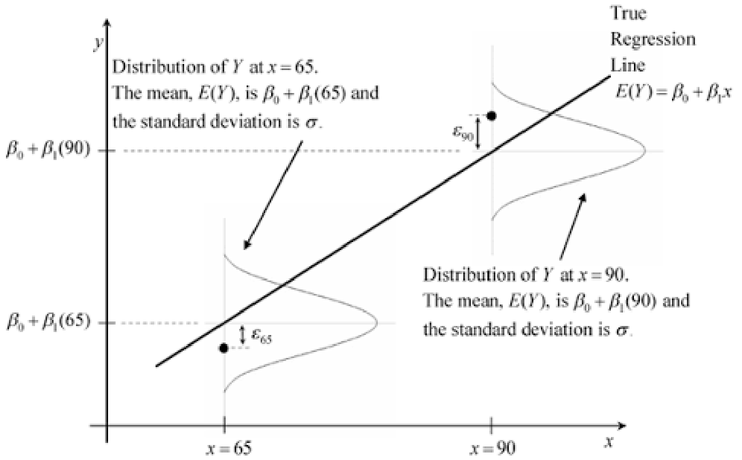

Figura 1: Médias Condicionais.

Estimação vs. previsão de valores

- Sir Francis Galton (1889, p. 62) (*apud* KOENKER, 2000, p. 350) criticou os seus colegas que:

Estimação vs. previsão de valores

- Sir Francis Galton (1889, p. 62) (*apud* KOENKER, 2000, p. 350) criticou os seus colegas que:
- *limitam suas investigações a médias, e não parecem se deleitar com visões mais abrangentes. Suas almas parecem tão enfadonhas ao encanto da variedade quanto ao de um nativo de um de nossos condados ingleses planos, cuja retrospectiva da Suíça foi que, se as montanhas pudessem ser jogadas em seus lagos, dois incômodos seriam removidos de uma só vez.*

Estimação vs. previsão de valores

- Sir Francis Galton (1889, p. 62) (*apud* KOENKER, 2000, p. 350) criticou os seus colegas que:
- *limitam suas investigações a médias, e não parecem se deleitar com visões mais abrangentes. Suas almas parecem tão enfadonhas ao encanto da variedade quanto ao de um nativo de um de nossos condados ingleses planos, cuja retrospectiva da Suíça foi que, se as montanhas pudessem ser jogadas em seus lagos, dois incômodos seriam removidos de uma só vez.*
- Estimação é para parâmetros da população (média, variância, coeficientes, …)

Estimação vs. previsão de valores

- Sir Francis Galton (1889, p. 62) (*apud* KOENKER, 2000, p. 350) criticou os seus colegas que:
- *limitam suas investigações a médias, e não parecem se deleitar com visões mais abrangentes. Suas almas parecem tão enfadonhas ao encanto da variedade quanto ao de um nativo de um de nossos condados ingleses planos, cuja retrospectiva da Suíça foi que, se as montanhas pudessem ser jogadas em seus lagos, dois incômodos seriam removidos de uma só vez.*
- Estimação é para parâmetros da população (média, variância, coeficientes, …)
- Previsão é para novos dados da população (fora da amostra).

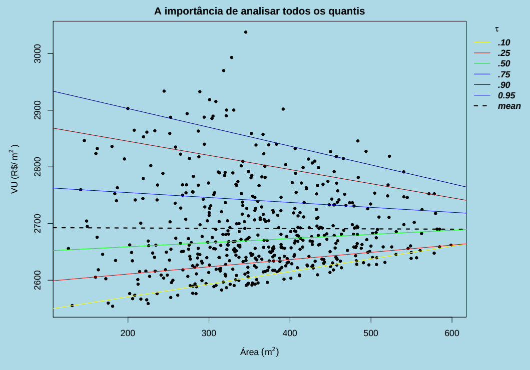

Figura 2: Regressão Quantílica.

Estimação vs. previsão de valores

- NBR 14.653-02 (ABNT, 2011) parece misturar as coisas

Estimação vs. previsão de valores

- NBR 14.653-02 (ABNT, 2011) parece misturar as coisas
  - A.10.1.2 Quando for adotado o valor arbitrado, o intervalo de valores admissíveis deve estar limitado simultaneamente (ver Figura A.2):

Estimação vs. previsão de valores

- NBR 14.653-02 (ABNT, 2011) parece misturar as coisas
    - A.10.1.2 Quando for adotado o valor arbitrado, o intervalo de valores admissíveis deve estar limitado simultaneamente (ver Figura A.2):

Estimação vs. previsão de valores

- NBR 14.653-02 (ABNT, 2011) parece misturar as coisas
  - A.10.1.2 Quando for adotado o valor arbitrado, o intervalo de valores admissíveis deve estar limitado simultaneamente (ver Figura A.2):
  - a) ao intervalo em torno do valor arbitrado com amplitude igual à do *intervalo de predição* ***ou ao intervalo de confiança*** de 80% para a estimativa de tendência central;

Estimação vs. previsão de valores

- NBR 14.653-02 (ABNT, 2011) parece misturar as coisas
  - A.10.1.2 Quando for adotado o valor arbitrado, o intervalo de valores admissíveis deve estar limitado simultaneamente (ver Figura A.2):
  - a) ao intervalo em torno do valor arbitrado com amplitude igual à do *intervalo de predição* ***ou ao intervalo de confiança*** de 80% para a estimativa de tendência central;

Estimação vs. previsão de valores

- NBR 14.653-02 (ABNT, 2011) parece misturar as coisas
  - A.10.1.2 Quando for adotado o valor arbitrado, o intervalo de valores admissíveis deve estar limitado simultaneamente (ver Figura A.2):
  - a) ao intervalo em torno do valor arbitrado com amplitude igual à do *intervalo de predição* ***ou ao intervalo de confiança*** de 80% para a estimativa de tendência central;
  - b) ao campo de arbítrio em torno da estimativa de tendência central.

Por que arbitrar um valor diferente da média condicional?

Campo de Arbítrio

- **campo de arbítrio**
  intervalo de variação no entorno do estimador pontual adotado na avaliação, dentro do qual pode-se arbitrar o valor do bem, **desde que justificado pela existência de características próprias não contempladas no modelo**. (ABNT, 2019, 3, grifo nosso)

Campo de Arbítrio

- **campo de arbítrio**
  intervalo de variação no entorno do estimador pontual adotado na avaliação, dentro do qual pode-se arbitrar o valor do bem, **desde que justificado pela existência de características próprias não contempladas no modelo**. (ABNT, 2019, 3, grifo nosso)

- **8.2.1.5.2** O campo de arbítrio pode ser utilizado **quando variáveis relevantes para a avaliação do imóvel não tiverem sido contempladas no modelo**, por escassez de dados de mercado, por inexistência de fatores de homogeneização aplicáveis ou porque essas variáveis não se apresentaram estatisticamente significantes em modelos de regressão, desde que a amplitude de até mais ou menos 15 % seja suficiente para absorver as influências não consideradas e que os ajustes sejam justificados.
  (ABNT, 2011, 17, grifo nosso)

Campo de Arbítrio

- **campo de arbítrio**
  intervalo de variação no entorno do estimador pontual adotado na avaliação, dentro do qual pode-se arbitrar o valor do bem, **desde que justificado pela existência de características próprias não contempladas no modelo**. (ABNT, 2019, 3, grifo nosso)

- **8.2.1.5.2** O campo de arbítrio pode ser utilizado **quando variáveis relevantes para a avaliação do imóvel não tiverem sido contempladas no modelo**, por escassez de dados de mercado, por inexistência de fatores de homogeneização aplicáveis ou porque essas variáveis não se apresentaram estatisticamente significantes em modelos de regressão, desde que a amplitude de até mais ou menos 15 % seja suficiente para absorver as influências não consideradas e que os ajustes sejam justificados.
  (ABNT, 2011, 17, grifo nosso)

- ***Características próprias não contempladas no modelo*** vs. ***variáveis relevantes não contempladas no modelo.***

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo
- A NBR 14.653-2 prevê a utilização do CA

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo
- A NBR 14.653-2 prevê a utilização do CA
- Entendemos que o CA pode ser um bom fator limitante

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo
- A NBR 14.653-2 prevê a utilização do CA
- Entendemos que o CA pode ser um bom fator limitante
- Mas não dá qualquer parâmetro para a previsão de valores

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo
- A NBR 14.653-2 prevê a utilização do CA
- Entendemos que o CA pode ser um bom fator limitante
- Mas não dá qualquer parâmetro para a previsão de valores
  - P. Ex.: Lote em situação de esquina, sem variável situação

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo
- A NBR 14.653-2 prevê a utilização do CA
- Entendemos que o CA pode ser um bom fator limitante
- Mas não dá qualquer parâmetro para a previsão de valores
  - P. Ex.: Lote em situação de esquina, sem variável situação
  - Para isto, pode-se utilizar o CA. Mas em qual magnitude?

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo
- A NBR 14.653-2 prevê a utilização do CA
- Entendemos que o CA pode ser um bom fator limitante
- Mas não dá qualquer parâmetro para a previsão de valores
  - P. Ex.: Lote em situação de esquina, sem variável situação
  - Para isto, pode-se utilizar o CA. Mas em qual magnitude?
  - Quanto a posição na quadra valoriza o lote em relação à situação de meio-de-quadra?

Problema da omissão de variável relevante

- A NBR 14.653-2 não apresenta uma definição rigorosa de *variável relevante*.
- Seleção de variáveis é um tema complexo para o qual os estatísticos ainda não tem solução fechada.
- Mas é consenso que, para evitar *overfitting*, nem todas as variáveis podem ser incluídas no modelo final
  - $p < \sqrt{n}$ (John Tukey *apud* Matloff (2017))
- Mesmos variáveis com significância estatística podem ter que ser retiradas do modelo
- A NBR 14.653-2 prevê a utilização do CA
- Entendemos que o CA pode ser um bom fator limitante
- Mas não dá qualquer parâmetro para a previsão de valores
  - P. Ex.: Lote em situação de esquina, sem variável situação
  - Para isto, pode-se utilizar o CA. Mas em qual magnitude?
  - Quanto a posição na quadra valoriza o lote em relação à situação de meio-de-quadra?
  - Intervalos de Predição podem ser instrutivos

```
## Call:
## bestfit(formula = valor ~ area_total + quartos + suites + garagens +
##     dist_b_mar + padrao, data = dados, subset = -c(31, 39), transf = c("rec",
##     "rsqrt", "log", "sqrt"))
##
## Best (Chosen) Transformations:
##      id valor area_total  quartos   suites garagens dist_b_mar padrao adj_R2
## 633 514   log   identity identity identity identity        log    rec   0.95
##
## Best (Chosen) fit LM summary:
##                       Estimate Std. Error t value Pr(>|t|) signif
## (Intercept)           13.56386   0.230495   58.85   <2e-16    ***
## identity(area_total)   0.00147   0.000288    5.11    8e-06    ***
## identity(quartos)      0.16357   0.035356    4.63    4e-05    ***
## identity(suites)       0.06096   0.033686    1.81     0.08    ***
## identity(garagens)     0.20873   0.033416    6.25    2e-07    ***
## log(dist_b_mar)       -0.14072   0.027197   -5.17    6e-06    ***
## rec(padrao)           -0.56284   0.105001   -5.36    3e-06    ***
## ---
##  Signif. codes:  0 '***' 0.10 '**' 0.20 '*' 0.30 ' ' 1
## NBR-14.653-2 check:
## Number of market data used:
## [1] "n = 48 >= 42 --> Grau III"
## Max significance level allowed for each predictor:
## [1] "t máximo = 7.77 %  < 10% --> Grau III"
## Max significance level allowed for F-test:
## [1] "p-valor F = 3.25e-24 % < 1% --> Grau III"
```

Seleção de Variáveis

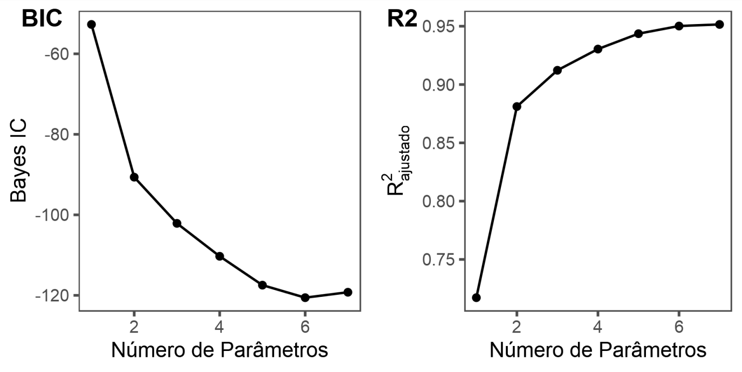

Figura 3: Seleção de Variáveis baseada em critérios de ajuste.

Seleção de Variáveis

- Modelo com variável suítes, IC: 961.660,64, 924.768,13, 1.000.024,94; Amplitude: 7,83%

- Modelo com variável suítes, IC: 961.660,64, 924.768,13, 1.000.024,94; Amplitude: 7,83%
- Modelo sem variável suítes, IC: 988.870,76, 955.381,88, 1.023.533,52; Amplitude: 6,89%

Seleção de Variáveis

- Modelo com variável suítes, IC: 961.660,64, 924.768,13, 1.000.024,94; Amplitude: 7,83%
- Modelo sem variável suítes, IC: 988.870,76, 955.381,88, 1.023.533,52; Amplitude: 6,89%
- Modelo com variável suítes, IP: 961.660,64, 802.017,63, 1.153.080,88; Amplitude: 36,51%

Seleção de Variáveis

- Modelo com variável suítes, IC: 961.660,64, 924.768,13, 1.000.024,94; Amplitude: 7,83%
- Modelo sem variável suítes, IC: 988.870,76, 955.381,88, 1.023.533,52; Amplitude: 6,89%
- Modelo com variável suítes, IP: 961.660,64, 802.017,63, 1.153.080,88; Amplitude: 36,51%
- Modelo sem variável suítes, IP: 988.870,76, 821.719,16, 1.190.023,82; Amplitude: 37,24%

Viés e Variância

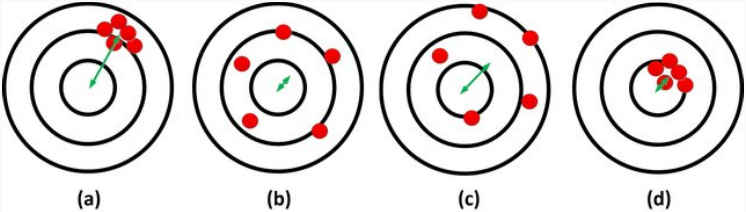

Figura 4: Viés e Variância. Fonte: Ghojogh e Crowley (2019).

Tradeoff

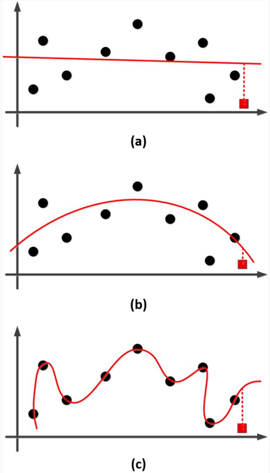

Figura 5: Overfitting. Fonte: Ghojogh e Crowley (2019).

Capacidade do Modelo

- Não se pode incluir todas as variáveis

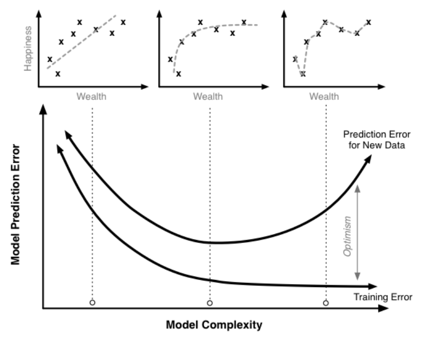

Figura 6: Capacidade do modelo. Fonte: autor desconhecido.

Capacidade do Modelo

- Não se pode incluir todas as variáveis

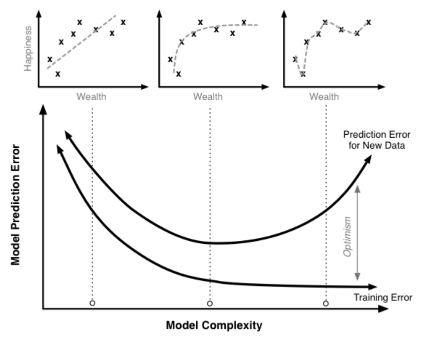

Figura 6: Capacidade do modelo. Fonte: autor desconhecido.

- Moral da estória: *If I had more time, I would have written a shorter letter* – Blaise Pascal *apud* Matloff (2017).

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (Ped)

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (Ped)
    - Outro possui melhor topografia ($Topo$)

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (Ped)
    - Outro possui melhor topografia ($Topo$)
    - Outro possui melhor Posição na Quadra (PQ)

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (Ped)
    - Outro possui melhor topografia ($Topo$)
    - Outro possui melhor Posição na Quadra (PQ)
    - Outro tem melhor posição solar (PS)

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (Ped)
    - Outro possui melhor topografia ($Topo$)
    - Outro possui melhor Posição na Quadra (PQ)
    - Outro tem melhor posição solar (PS)
    - Devido a estas pequenas diferenças, $\varepsilon \neq 0$

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (*Ped*)
    - Outro possui melhor topografia (*Topo*)
    - Outro possui melhor Posição na Quadra (*PQ*)
    - Outro tem melhor posição solar (*PS*)
    - Devido a estas pequenas diferenças, $\varepsilon \neq 0$
    - Porém, alguns lotes tem melhor *Ped*, *Topo*, *PQ* **e** *PS*.

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (*Ped*)
    - Outro possui melhor topografia (*Topo*)
    - Outro possui melhor Posição na Quadra (*PQ*)
    - Outro tem melhor posição solar (*PS*)
    - Devido a estas pequenas diferenças, $\varepsilon \neq 0$
    - Porém, alguns lotes tem melhor *Ped*, *Topo*, *PQ* **e** *PS*.
    - Enquanto outros tem pior *Ped*, *Topo*, *PQ* **e** *PS*.

Exemplo

- Na impossibilidade de se modelar todas as variáveis, cabe ao profissional de Engenharia de Avaliações identificar quando um imóvel tem valor acima ou abaixo da média do mercado
  - P. ex.: Existem diversos lotes com $A = 360m^2$ e $F = 12m$, porém nem todos são iguais
    - Um possui melhor pedologia (*Ped*)
    - Outro possui melhor topografia (*Topo*)
    - Outro possui melhor Posição na Quadra (*PQ*)
    - Outro tem melhor posição solar (*PS*)
    - Devido a estas pequenas diferenças, $\varepsilon \neq 0$
    - Porém, alguns lotes tem melhor *Ped*, *Topo*, *PQ* **e** *PS*.
    - Enquanto outros tem pior *Ped*, *Topo*, *PQ* **e** *PS*.
    - É justo avaliá-los com o valor médio?

Estudos de casos

Omissão de variável relevante

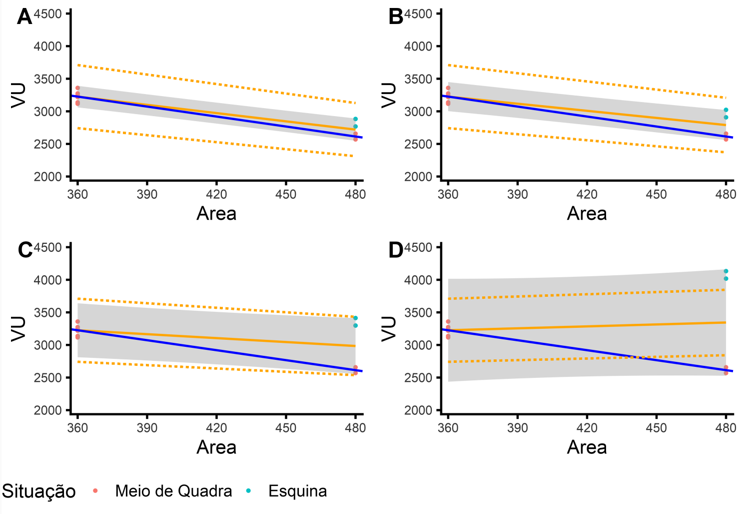

Figura 7: Omissão de variável relevante: quatro exemplos ilustrativos. Fonte: os autores.

- Modelo sem micronumerosidade

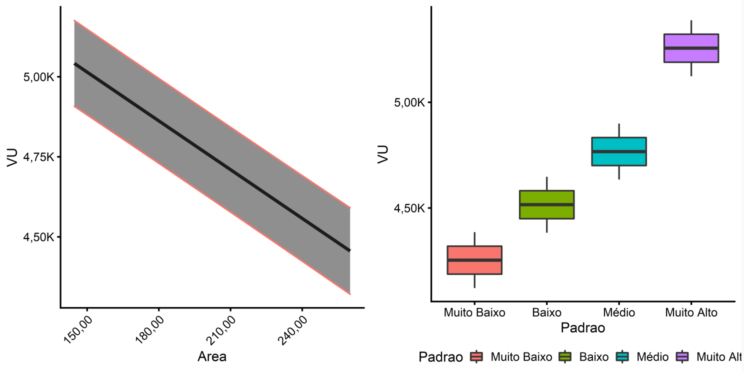

Figura 8: Micronumerosidade: perda de informação no modelo. Fonte: os autores.

Micronumerosidade

- Modelo sem micronumerosidade

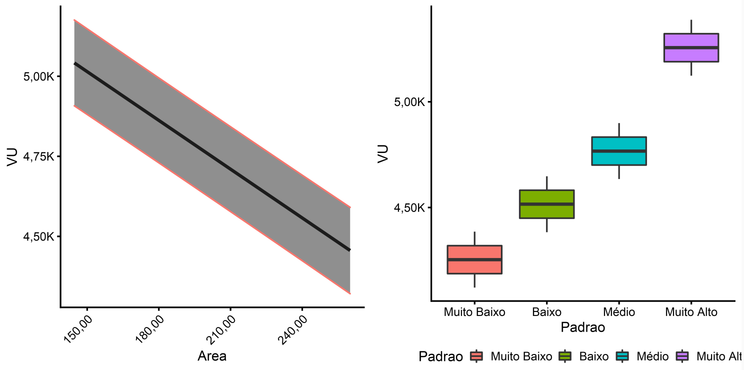

Figura 8: Micronumerosidade: perda de informação no modelo. Fonte: os autores.

- Quanto vale imóvel de alto padrão?

Micronumerosidade (2)

- Modelo com poucos dados de padrão alto (<3)

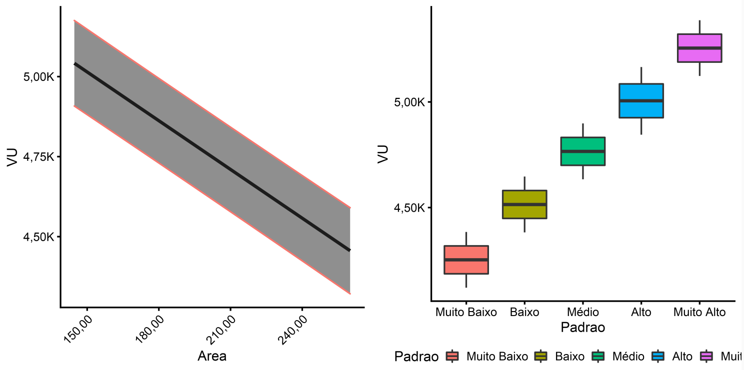

Figura 9: Pequeno número de dados: Estimação pobre. Fonte: os autores.

Micronumerosidade (2)

- Modelo com poucos dados de padrão alto (<3)

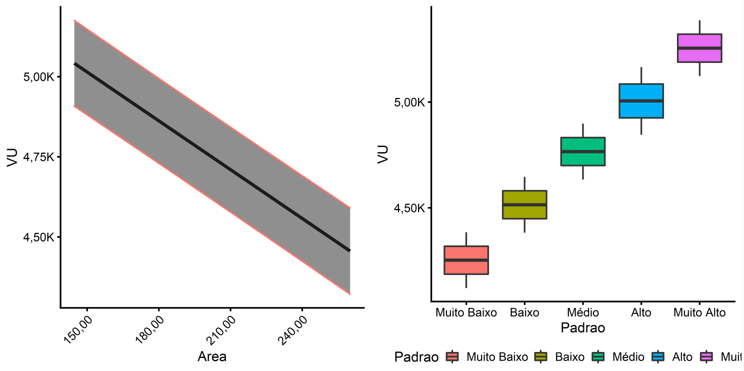

Figura 9: Pequeno número de dados: Estimação pobre. Fonte: os autores.

- $Var(VU|Padrao = alto) > Var(VU|Padrao \neq alto)$

Avaliação Intervalar

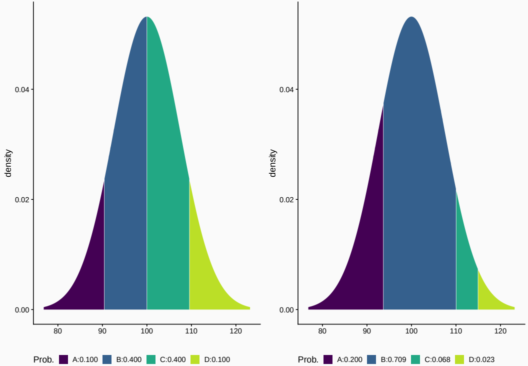

Figura 10: Avaliação Intervalar. Intervalos incoerentes.

- Faz sentido arbitrar um valor maior do que a média e reportar um intervalo admissível com valores abaixo da média?

Avaliação Intervalar (2)

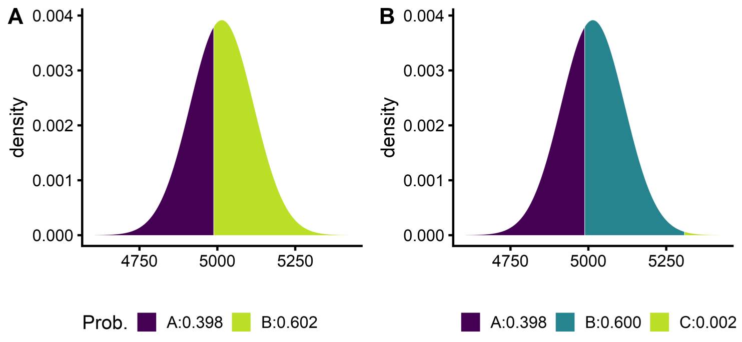

Figura 11: Estimação intervalar: uma abordagem probabilística. Fonte: os autores.

- Arbitragem de valor acima da média:

Avaliação Intervalar (2)

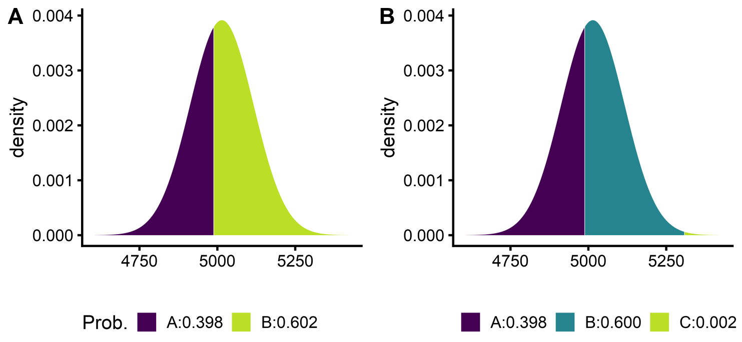

Figura 11: Estimação intervalar: uma abordagem probabilística. Fonte: os autores.

- Arbitragem de valor acima da média:
  - valores admissíveis: $[IC_{inf}; CA_{sup}]$ ou $[IC_{inf}; IP_{sup}]$

Avaliação Intervalar (2)

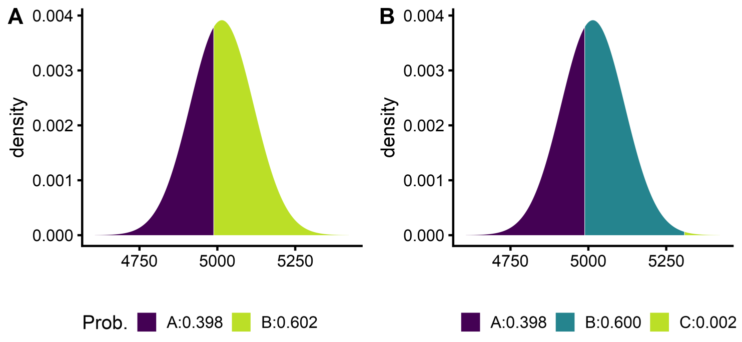

Figura 11: Estimação intervalar: uma abordagem probabilística. Fonte: os autores.

- Arbitragem de valor acima da média:
  - valores admissíveis: $[IC_{inf}; CA_{sup}]$ ou $[IC_{inf}; IP_{sup}]$
  - valores admissíveis: $[IC_{inf}; min(CA_{sup}, IP_{sup})]$

Avaliação Intervalar (2)

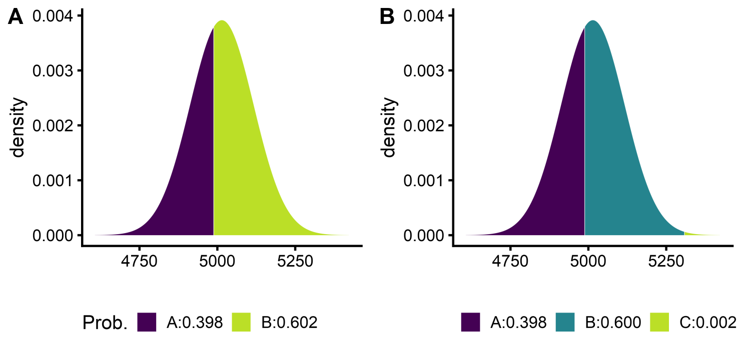

Figura 11: Estimação intervalar: uma abordagem probabilística. Fonte: os autores.

- Arbitragem de valor acima da média:
  - valores admissíveis: $[IC_{inf}; CA_{sup}]$ ou $[IC_{inf}; IP_{sup}]$
  - valores admissíveis: $[IC_{inf}; min(CA_{sup}, IP_{sup})]$
  - valores admissíveis: $Y_{arb} \pm min(Y_{arb} - IC_{inf}; min(IP_{sup}, CA_{sup}) - Y_{arb})$ (simétrico)

Conclusões

Conclusões

- O campo de arbítrio (CA) é um artifício que possibilita ao avaliador a escolha de um valor para o bem-avaliando diferente do valor (médio) ajustado com o modelo de regressão linear.

Conclusões

- O campo de arbítrio (CA) é um artifício que possibilita ao avaliador a escolha de um valor para o bem-avaliando diferente do valor (médio) ajustado com o modelo de regressão linear.

- A NBR 14653-2 (2011) permite que o CA seja utilizado apenas quando da omissão de variável relevante. Esta exigência não consta na parte 1 (2019)

Conclusões

- O campo de arbítrio (CA) é um artifício que possibilita ao avaliador a escolha de um valor para o bem-avaliando diferente do valor (médio) ajustado com o modelo de regressão linear.
- A NBR 14653-2 (2011) permite que o CA seja utilizado apenas quando da omissão de variável relevante. Esta exigência não consta na parte 1 (2019)
- O cálculo do IP é um bom parâmetro para estimar o *quanto* o valor do bem pode/deve ser majorado ou minorado.

Conclusões

- O campo de arbítrio (CA) é um artifício que possibilita ao avaliador a escolha de um valor para o bem-avaliando diferente do valor (médio) ajustado com o modelo de regressão linear.
- A NBR 14653-2 (2011) permite que o CA seja utilizado apenas quando da omissão de variável relevante. Esta exigência não consta na parte 1 (2019)
- O cálculo do IP é um bom parâmetro para estimar o *quanto* o valor do bem pode/deve ser majorado ou minorado.
- O IC não é um bom parâmetro para a arbitragem de valores.

Conclusões

- O campo de arbítrio (CA) é um artifício que possibilita ao avaliador a escolha de um valor para o bem-avaliando diferente do valor (médio) ajustado com o modelo de regressão linear.
- A NBR 14653-2 (2011) permite que o CA seja utilizado apenas quando da omissão de variável relevante. Esta exigência não consta na parte 1 (2019)
- O cálculo do IP é um bom parâmetro para estimar o *quanto* o valor do bem pode/deve ser majorado ou minorado.
- O IC não é um bom parâmetro para a arbitragem de valores.
- O CA deve ser utilizado, portanto, apenas como um fator limitante.

Conclusões

- O campo de arbítrio (CA) é um artifício que possibilita ao avaliador a escolha de um valor para o bem-avaliando diferente do valor (médio) ajustado com o modelo de regressão linear.
- A NBR 14653-2 (2011) permite que o CA seja utilizado apenas quando da omissão de variável relevante. Esta exigência não consta na parte 1 (2019)
- O cálculo do IP é um bom parâmetro para estimar o *quanto* o valor do bem pode/deve ser majorado ou minorado.
- O IC não é um bom parâmetro para a arbitragem de valores.
- O CA deve ser utilizado, portanto, apenas como um fator limitante.
- A micronumerosidade deveria ser revista. Deixar que os dados falem.

ABNT. **NBR 14653-1**: Avaliação de Bens – Parte 1: Procedimentos Gerais. Rio de Janeiro, 27 jun. 2019. p. 19.

________. **NBR 14653-2**: Avaliação de Bens – Parte 2: Imóveis Urbanos. Rio de Janeiro, fev. 2011. p. 4.

GALTON, F. **Natural Inheritence**. London: MacMillan, 1889.

GHOJOGH, Benyamin; CROWLEY, Mark. **The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial**. [S.l.: s.n.], 2019. arXiv: 1905.12787 [stat.ML].

KOENKER, Roger W. Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics. **Journal of Econometrics**, v. 95, n. 2, p. 347–374, 2000. ISSN 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(99)00043-3. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0304407699000433>.

MATLOFF, Norman. **Statistical Regression and Classification**: From Linear Models to Machine Learning. Boca Raton, Florida: Chapman & Hall, 2017.