

# Tópicos em Aprendizado Estatístico

## Métodos de Reamostragem

Prof. George von Borries

Departamento de Estatística  
Universidade de Brasília

2024

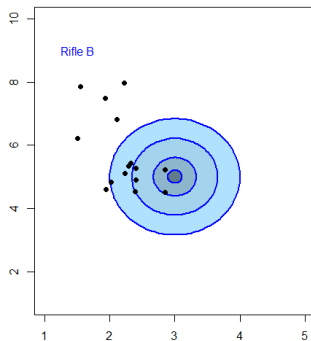
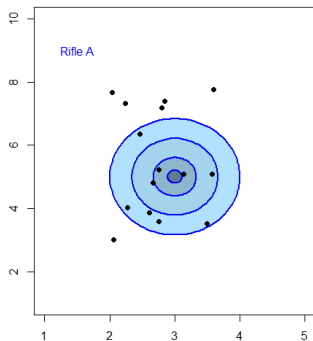


## Qualidade do Modelo



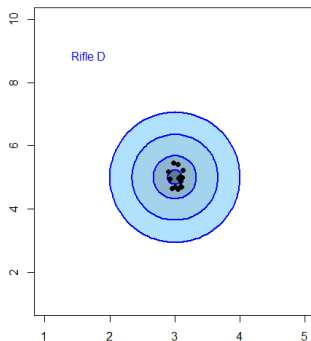
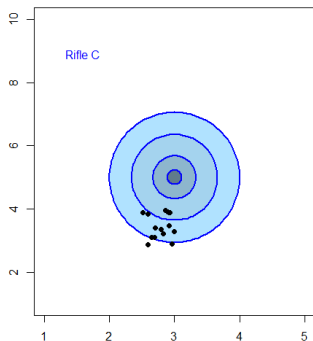
## Nota: Tendenciosidade e precisão.

Quatro rifles (A,B,C,D) foram testados em relação a tendenciosidade e precisão. Especialistas efetuaram 15 tiros com cada rifle.



## Nota: Tendenciosidade e precisão.

Quatro rifles (A,B,C,D) foram testados em relação a tendenciosidade e precisão. Especialistas efetuaram 15 tiros com cada rifle.



- **Estimador não viesado ou não tendencioso:**  $\hat{\theta}$  é um estimador não tendencioso de  $\theta$  se  $E(\hat{\theta}) = \theta$  para todo  $\theta$ , i.e., se a média da distribuição amostral de  $\hat{\theta}$  é igual ao verdadeiro valor da população.
- **Estimador Preciso:** quando os valores do estimador tem pequena variância amostral, i.e., diferentes estimativas estão sempre próximas umas das outras.
- Procuramos sempre um estimador não tendencioso e preciso.



- Suponha que observamos uma resposta  $\mathbf{Y}$  e um conjunto de preditores  $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_p\}$ . Seja a relação de  $\mathbf{Y}$  e  $\mathbf{X}$  obtida por

$$\mathbf{Y} = f(\mathbf{X}) + \epsilon, \quad E[\epsilon] = 0 \text{ e } \text{Var}[\epsilon] = \sigma^2.$$

$f$  é a função verdadeira (fixa desconhecida) e  $\epsilon$  é independente de  $\mathbf{X}$ .

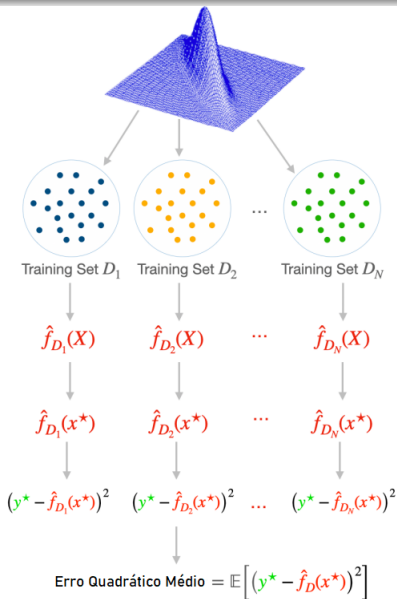
- Uma boa escolha de  $f$  fornece boas previsões<sup>1</sup> de  $\mathbf{Y}$  em novos pontos de  $\mathbf{X} = \mathbf{x}$ .
- Obtemos um dado de treinamento  $\mathcal{D}$  i.i.d. da distribuição geradora e aplicamos um algoritmo de aprendizado  $\mathcal{A}$ .  
A estimativa de  $f$  via  $\mathcal{D}$  será  $\hat{f}_{\mathcal{D}}(\mathbf{X})$ .

- Nossa preocupação é verificar a performance do modelo num conjunto de teste. Para um ponto arbitrário  $\mathbf{X}^*$ ,  $\mathbf{Y}^* = f(\mathbf{X}^*) + \epsilon$ .
- O desempenho do modelo treinado poderá ser avaliado para  $\mathbf{Y}^*$  através do erro quadrático (SE)

$$\text{SE} = [\mathbf{Y}^* - \hat{f}_{\mathcal{D}}(\mathbf{X}^*)]^2$$

- O erro quadrático médio (MSE) é obtido da média do erro quadrático de estimativas obtidas para vários conjuntos de





Fonte: [Akinkunle, 2020.](#)



- Então,  $MSE^2$  será<sup>3</sup>

$$\begin{aligned}
 E \left[ \mathbf{Y} - \hat{f}_D(\mathbf{X}) \right]^2 &= E \left[ (\mathbf{Y} - \hat{f}_D(\mathbf{X})) (\mathbf{Y} - \hat{f}_D(\mathbf{X})) \right] \\
 &= E \left[ \mathbf{Y}^2 - 2\mathbf{Y}\hat{f}_D(\mathbf{X}) + \hat{f}_D^2(\mathbf{X}) \right] \\
 &= E \left[ \mathbf{Y}^2 \right] + E \left[ \hat{f}_D^2(\mathbf{X}) \right] - 2E[\mathbf{Y}] E \left[ \hat{f}_D(\mathbf{X}) \right] \\
 &= \underbrace{\left( E \left[ \hat{f}_D(\mathbf{X}) \right] - f(\mathbf{X}) \right)^2}_{\text{viés do modelo}} + \underbrace{E \left[ \left( \hat{f}_D(\mathbf{X}) - E[\hat{f}_D(\mathbf{X})] \right)^2 \right]}_{\text{variância do modelo}} + \underbrace{\text{Var}(\epsilon)}_{\text{irreduzível}}
 \end{aligned}$$

reduzível

- Em geral, um modelo mais flexível (logo mais complicado) terá maior variância e menor viés.
- Em geral, um modelo mais flexível terá menor interpretabilidade.

<sup>2</sup>Para detalhes do cálculo, ver [Akinkunle, 2020](#)

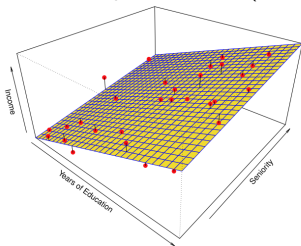
<sup>3</sup>Utilizando  $\mathbf{Y}$  e  $\mathbf{X}$  em vez de  $\mathbf{Y}^*$  e  $\mathbf{X}^*$  para facilitar a notação.0



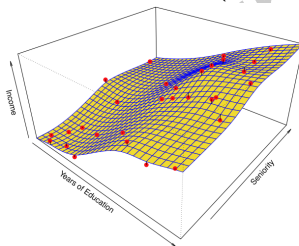


## Exemplo: Relação entre $Y = \text{Renda}$ e $X = \{\text{Formação, Senioridade}\}$

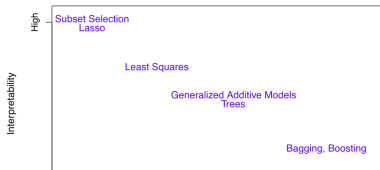
Modelo pouco flexível (linear)



Modelo flexível (não linear)



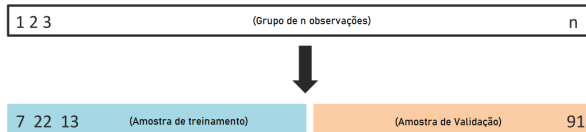
Dados no arquivo [Income2.csv](#) de James et al. (2021)



## Validação Cruzada



- **Validação Cruzada** é um procedimento de partição de dados (treinamento e teste) que permite avaliar:
  - 1 estabilidade das estimativas dos parâmetros;
  - 2 performance do modelo (*Model assessment*);
  - 3 nível de flexibilidade (*Model selection*);
  - 4 precisão de algoritmos de classificação.
- A performance do modelo ajustado para os dados de treinamento pode ser feita via RMSE ou MSE do ajuste utilizando os dados de teste.



Fonte: James et al. (2021)

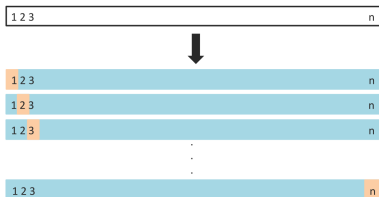


## Procedimento de Validação Cruzada LOOCV<sup>4</sup>

Seja um vetor resposta  $\mathbf{Y}_{(n \times 1)}$  e matriz de covariáveis  $\mathbf{X}_{(p \times n)}$ .

- 1 O modelo de interesse é ajustado  $n$  vezes, eliminando um elemento em cada ajuste.
- 2 A cada passo, eliminamos uma observação de  $\mathbf{Y}$  e  $\mathbf{X}$ . Estimamos o modelo (ex. Regressão) para as  $n - 1$  observações restantes e calculamos  $\text{MSE}_i = (y_i - \hat{y}_i)^2$  para a  $i$ -ésima observação eliminada.
- 3 A estimativa de Validação Cruzada será,

$$\text{CV}_{(n)} = \frac{1}{n} \sum_{i=1}^n \text{MSE}_i$$



<sup>4</sup> Leave-One-Out Cross Validation, também conhecido como Jackknife.



## Procedimento de Validação Cruzada LOOCV

- $CV_{(n)}$  é um procedimento pouco viesado, uma vez que utiliza quase toda amostra para treinamento.
- É uma estimativa mais estável que outros procedimentos de validação cruzada.
- Não existe aleatoriedade na divisão treinamento/validação, logo o resultado será sempre o mesmo para um mesmo conjunto dados.
- É um procedimento geral e pode ser utilizado com qualquer modelo preditivo.

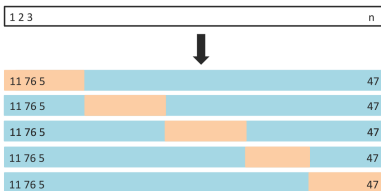


## Procedimento de Validação Cruzada de Ordem $k$ (ou $k$ -fold)

Seja um vetor resposta  $\mathbf{Y}_{(n \times 1)}$  e matriz de covariáveis  $\mathbf{X}_{(p \times n)}$ .

- 1 O modelo de interesse é ajustado  $k$  vezes, em grupos de tamanhos aproximados  $n_k$ , tal que  $\sum_k n_k = n$ .
- 2 A cada passo estimamos o modelo (ex. Regressão) para as  $n - n_k$  observações restantes e calculamos  $\text{MSE}_i = \frac{1}{n_k} \sum_{i=1}^{n_k} (y_i - \hat{y}_i)^2$  para as  $n_k$  observações eliminadas.
- 3 Então, calculamos

$$\text{CV}_{(k)} = \frac{1}{k} \sum_{i=1}^k \text{MSE}_i$$

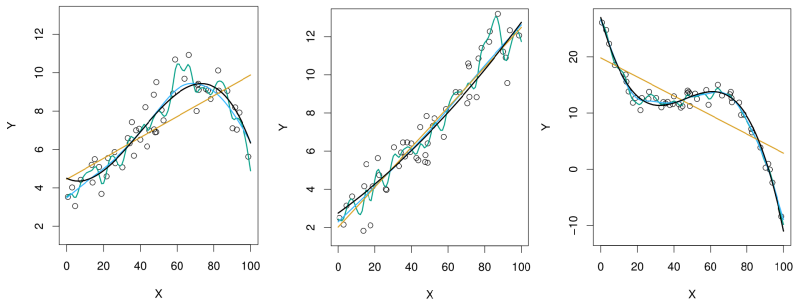


## Procedimento de Validação Cruzada de Ordem $k$ (ou $k$ -fold)

- $CV_{(k)}$  usa menos recurso computacional, mas se torna mais instável a medida que  $n_k$  aumenta (ou  $k$  diminui).
- $CV_{(n)}$  tem menor viés que a estimativa  $CV_{(k)}$ , pois utiliza mais informação.
- Entretanto, os modelos ajustados por LOOCV são mais correlacionados resultando em maior variância para a média destas quantidades (lembre de colinearidade). Desta forma, validação cruzada de ordem  $k$  resulta em menor variância para o erro de teste.



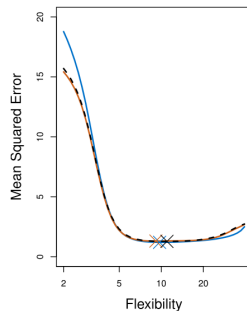
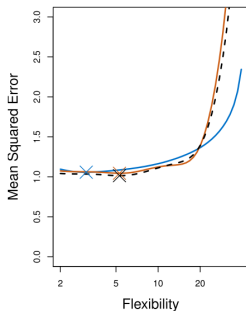
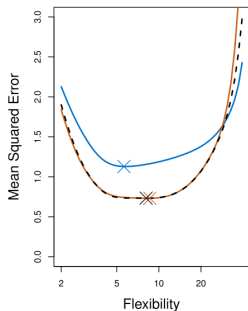
**Exemplo: (James et al., 2021)** Dados gerados e diferentes ajustes. MSE estimado por validação cruzada.



Diferentes polinômios ajustados para três conjuntos de dados gerados com diferentes graus de linearidade (**linear**, suavização intermediária, **suavização forte**). Figuras 2.9, 2.10 e 2.11 de James et al., 2021.







MSE verdadeiro, LOOCV e  $CV_{(10)}$ . X representa o mínimo de cada curva MSE.



O programa `Reamostragem.R` possui os seguintes exemplos de CV:

**Exemplo 1.** Validação cruzada em Regressão por Componentes Principais.

**Exemplo2.** LOOCV para seleção de modelos. Exemplo de Rizzo<sup>5</sup> (2019). Este exemplo utiliza o arquivo `ironslag` do pacote DAAG do R. Este arquivo 53 medidas de teor de ferro utilizando dois métodos, *chemical* e *magnetic*. Rizzo tenta verificar se o método *magnetic* pode ser predito pelos resultados do método *chemical*.

Fonte: Hand, D.J., Daly, F., McConway, K., Lunn, D., and Ostrowski, E. eds (1993) A Handbook of Small Data Sets. London: Chapman & Hall.

Obs: Em situações reais, o objetivo é inverso, uma vez que o método *chemical* é mais complicado que o *magnetic*.

---

<sup>5</sup>Rizzo, M.L. (2019) Statistical Computing with R. Chapman and Hall / CRC, Segunda Edição.



# Bootstrap



O termo Bootstrap foi citado pela primeira vez por Bradley Efron em 1979. Neste texto, Efron faz uma breve descrição do método e procura ressaltar diversas técnicas que sofreram influência do advento da computação.

A denominação *Bootstrap* (Cadarço de Bota) é justificada como sendo um termo eufônico ao termo *Jackknife* (Canivete) de Quenouille (1949) que procurava indicar um método útil em uma grande variedade de situações.

### Definição

Seja  $\mathcal{X} = \{x_1, \dots, x_n\}$  uma amostra aleatória, de tamanho  $n$ , de uma população com função de distribuição  $F$ . Estamos interessados em estimar a característica populacional  $\theta$  definida pelo funcional  $\theta = Q(F)$ . O estimador Bootstrap será  $\hat{\theta} = Q(\hat{F})$ , resultante da aplicação do mesmo funcional sobre a função de distribuição  $\hat{F}$  (substituição de  $F$  por  $\hat{F}$ ) obtida de uma amostra de  $F$ .



A distribuição empírica é uma possível escolha para  $\hat{F}$ . A função de distribuição empírica de  $\mathcal{X}$  é dada por

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x),$$

tal que

$$I(X_i \leq x) = \begin{cases} 1 & \text{se } X_i \leq x \\ 0 & \text{caso contrário,} \end{cases}$$

em que  $I(X_i \leq x)$  é denominada função indicadora.

**Exemplo:** A distribuição empírica da amostra

$\mathcal{X} = \{1, 2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 8, 8, 9, 10\}$  será

$\hat{F}_n = \{0.05, 0.11, 0.21, 0.37, 0.58, 0.68, 0.79, 0.89, 0.95, 1\}$ ,

correspondendo a probabilidade acumulada de cada ponto na amostra.



Efron e Tibshirani<sup>6</sup> (1986) apresentam um esquema simples que permite a compreensão rápida da idéia central em estimação bootstrap, envolvendo a relação entre  $\theta = Q(F)$  e  $\hat{\theta} = Q(\hat{F})$ .

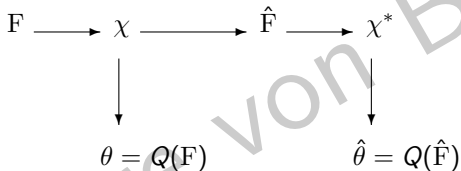


Figura 1: Ilustração Esquemática da Estimação Bootstrap.

<sup>6</sup>Efron, B. e Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1, 1 (1986), 54-77.



**Exemplo:** Suponha que estamos interessados em estimar a média populacional

$$\mu = \int x dF(x) = Q(F) = \theta.$$

O estimador bootstrap de  $\mu$  será obtido através da aplicação do mesmo funcional sobre a função distribuição empírica  $\hat{F}$ ,

$$\bar{X} = \int x d\hat{F}(x) = Q(\hat{F}) = \hat{\theta},$$

que na realidade é a média amostral



**Princípio Bootstrap:** princípio de substituição da população por uma distribuição obtida da amostra e o cálculo da estimativa através de um processo de reamostragem. Hall compara o princípio com uma boneca russa denominada “matryoska”, que é uma sequência de figuras de madeira com pequenas diferenças em cada boneca.



Figura: Boneca denominada Matryoska.

Neste caso, a boneca menor representa a população e  $n_1$  o valor populacional desconhecido. A ideia central é a suposição de que o comportamento em algo que não observamos é semelhante ao comportamento de algo que conseguimos observar nas bonecas maiores.





# Algoritmo de Monte Carlo para Estimação Bootstrap

Este algoritmo é uma aproximação numérica para o processo de substituição de  $F$  por  $\hat{F}$  citado na Definição e que corresponde a substituição de  $F_0$  (população) por  $F_1$  (amostra) e  $F_1$  por  $F_2$  (reamostragem) no Princípio Bootstrap.

Etapas do algoritmo de Monte Carlo para estimação bootstrap:

- 1 Através da seleção aleatória, com reposição, obter um número  $B$  de amostras de  $\mathcal{X}$ , que representaremos por  $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$  e denominaremos de amostras bootstrap (amostras obtidas via  $F_2$ );
- 2 Para cada amostra  $\mathcal{X}_b^*$  calcular a estatística de interesse  $\hat{\theta}_b^* = \hat{\theta}(\mathcal{X}_b^*)$ ,  $b = 1, 2, \dots, B$ , utilizando a mesma fórmula que define  $\hat{\theta} = \theta(F_1)$ ;
- 3 Calcular a estimativa bootstrap ou equação funcional do parâmetro  $\theta_0$ ,

$$\hat{\theta}_B^* = \frac{\sum_{b=1}^B \hat{\theta}_b^*}{B}.$$

Seu desvio-padrão bootstrap será,

$$\hat{\sigma}_B^* = \left( \frac{\sum_{b=1}^B \{ \hat{\theta}_b^* - \hat{\theta}_B^* \}^2}{B - 1} \right)^{1/2}$$



## Bootstrap Paramétrico

Neste caso, assumimos que a distribuição populacional  $F_0$  tem forma funcional conhecida com alguns parâmetros  $\theta_0$  desconhecidos. Desta maneira,  $F_0 = F_{\theta_0}$  representa um elemento da família  $\mathcal{F} = \{F_{\theta}, \theta \in \Omega\}$  de distribuições possíveis. Neste contexto tomamos  $F_1$  como  $F_{\hat{\theta}_1}$ , onde  $\hat{\theta}_1$  representa um estimador de  $\theta_0$ . Neste processo  $\mathcal{X}^*$  será uma amostra (bootstrap) obtida da função de distribuição  $F_{\hat{\theta}_1}$  e o parâmetro calculado para uma amostra de  $F_{\hat{\theta}_1}$ , pelo mesmo método que gerou o cálculo de  $\hat{\theta}_1$ , será representado por  $\hat{\theta}_2^* = \hat{\theta}_2(\mathcal{X}^*)$ . Assim  $F_2$  representa o processo de reamostragem da população definida pela função de distribuição  $F_{\hat{\theta}_1}$ .



## Exemplo: Bootstrap Paramétrico

Considere  $n$  observações de uma população bivariada  $(X, Y)$ . Suponha que desejamos calcular uma estimativa bootstrap do coeficiente de correlação  $\rho$ , entre  $X$  e  $Y$ , i.e,

$$\rho = \frac{\sum_{i=1}^N (X_i - \mu_x)(Y_i - \mu_y)}{\left[ \sum_{i=1}^N (X_i - \mu_x)^2 \right] \left[ \sum_{i=1}^N (Y_i - \mu_y)^2 \right]}$$

Suponha ainda que por alguma razão, acreditamos que a distribuição populacional  $F_0$  é uma função de distribuição normal bivariada com vetor de médias  $\mu$  e matriz de covariância  $\Sigma$  desconhecidos.

Após determinarmos  $F_1 = N_2(\hat{\mu}, \hat{\Sigma})$ , podemos aplicar as etapas 1 a 3 do algoritmo de Monte Carlo, obtendo assim uma estimativa bootstrap da estatística  $\rho$  calculando

$$r = \frac{\sum_{i=1}^n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{\left[ \sum_{i=1}^n (x_i - \hat{\mu}_x)^2 \right] \left[ \sum_{i=1}^n (y_i - \hat{\mu}_y)^2 \right]}$$

sobre  $n$  observações  $(x, y)$  amostradas de uma normal bivariada com parâmetros  $\hat{\mu}$  e  $\hat{\Sigma}$ .



## Bootstrap Não-Paramétrico

Na estimação bootstrap **não-paramétrica** nada se assume sobre  $F_0$  além de que seja uma função de distribuição. A função de distribuição  $F_1$  corresponde à distribuição empírica da amostra  $\mathcal{X}$ , i.e, se  $\mathcal{X}$  é uma amostra aleatória de tamanho  $n$  de  $F_0$ ,  $F_1$  é a distribuição que atribui peso  $1/n$  a cada observação em  $\mathcal{X}$ . De modo análogo,  $F_2$  é a função de distribuição empírica de uma amostra aleatória de uma população com distribuição  $F_1$ , i.e., a distribuição empírica de uma amostra  $\mathcal{X}^*$  retirada aleatoriamente, com reposição, da amostra original  $\mathcal{X}$ . Note que se simbolizarmos a população por  $\mathcal{X}_0$ , então  $\mathcal{X}$  constitui uma amostra aleatória de  $\mathcal{X}_0$  e  $\mathcal{X}^*$  uma amostra aleatória de  $\mathcal{X}$ .



## Intervalos de Confiança Bootstrap

### Intervalo Percentil ( $I_e$ ):

A idéia do intervalo Percentil é muito simples. Seja  $\alpha \in (0, 1)$  e  $\hat{H}(x) = P(\hat{\theta}^* \leq x/\mathcal{X})$  a função de distribuição de probabilidade calculada sobre a distribuição bootstrap de  $\hat{\theta}^*$ . Calculando,

$$\hat{H}^{-1}(\alpha) = \inf\{x : \hat{H}(x) \geq \alpha\}$$

para essa distribuição, o intervalo Percentil (bicaudal) ao nível de  $(1 - 2\alpha) 100\%$  será

$$I_e = (\hat{H}^{-1}(\alpha), \hat{H}^{-1}(1 - \alpha)).$$



## Intervalo Percentil Viés Corrigido ( $I_{vc}$ ):

Seja

$$\hat{y}_\alpha = \hat{H}^{-1}(\alpha) = \hat{H}^{-1}(\Phi(z_\alpha)).$$

A sequência de quantis  $\{\hat{y}_\alpha, 0 < \alpha < 1\}$  estará centralizada empiricamente se  $\hat{y}_{0.5}$  for igual a  $\hat{\theta}$ , i.e., se  $H(\hat{\theta}) = 0.5$ . Quando isto não acontece, uma correção em  $\hat{y}_\alpha$  poderia ser efetuada através da relação,

$$\hat{y}_{vc,\alpha} = \hat{H}^{-1}[\Phi(\hat{m} + (\hat{m} + z_\alpha))] = \hat{H}^{-1}(\Phi(2\hat{m} + z_\alpha))$$

em que

$$\hat{m} = \Phi^{-1}(\hat{H}(\hat{\theta}))$$

é a estimativa da correção de viés e  $\hat{y}_{vc,\alpha}$  o quantil viés-corrigido. O intervalo Percentil Viés Corrigido bilateral com nível de confiança  $(1 - 2\alpha)$  será,

$$I_{vc} = (\hat{y}_{vc,\alpha}; \hat{y}_{vc,1-\alpha}).$$



## Intervalo Percentil Viés Corrigido Acelerado ( $I_{vca}$ ):

Este intervalo é obtido através da introdução da assimetria na correção do intervalo percentil. Assim,

$$\hat{y}_{vca,\alpha} = \hat{H}^{-1}(\Phi[\hat{m} + (\hat{m} + z_{\alpha})(1 - \hat{a}(\hat{m} + z_{\alpha}))^{-1}])$$

sendo  $\hat{m}$  o fator de correção do viés e  $\hat{a}$  o fator de correção de assimetria.

O intervalo de confiança bilateral, viés corrigido acelerado, ao nível  $(1 - 2\alpha)$  será  $I_{vca} = (\hat{y}_{vca,\alpha}; \hat{y}_{vca,1-\alpha})$ .

Uma aproximação para a constante de aceleração é utilizar  $1/6$  do coeficiente de assimetria da distribuição do estimador bootstrap.



## Intervalo Percentil-ta:

Seja  $P \left[ n^{1/2} (\hat{\theta}_1 - \theta_0) / \sigma_0 \leq ta_{(1-\alpha)} | F_0 \right] = 1 - \alpha$ .

Os quantis obtidos irão produzir o intervalo teórico percentil-ta,

$$\mathcal{I}_{ta} = (\hat{\theta}_1 - n^{-1/2} \sigma_0 ta_{(1-\alpha/2)}; \hat{\theta}_1 - n^{-1/2} \sigma_0 ta_{(\alpha/2)})$$

Este intervalo possui caudas iguais e tem nível de confiança  $(1 - 2\alpha)\%$ .

A versão bootstrap do intervalo  $\mathcal{I}_{ta}$  será

$$\mathcal{I}_{\hat{t}a} = (\hat{\theta}_1 - n^{-1/2} \sigma_0 \hat{t}a_{(1-\alpha/2)}; \hat{\theta}_1 - n^{-1/2} \sigma_0 \hat{t}a_{(\alpha/2)})$$





## Intervalo Percentil-tb:

Seja  $P \left[ n^{1/2} (\hat{\theta}_1 - \theta_0) / \hat{\sigma}_1 \leq tb_{(1-\alpha)} | F_0 \right] = 1 - \alpha$ .

Os quantis obtidos irão produzir o intervalo teórico percentil-ta,

$$\mathcal{I}_{tb} = (\hat{\theta}_1 - n^{-1/2} \hat{\sigma}_1 tb_{(1-\alpha/2)}; \hat{\theta}_1 - n^{-1/2} \hat{\sigma}_1 tb_{(\alpha/2)})$$

Este intervalo possui caudas iguais e tem nível de confiança  $(1 - 2\alpha)\%$ .

A versão bootstrap do intervalo  $\mathcal{I}_{ta}$  será

$$\mathcal{I}_{tb} = (\hat{\theta}_1 - n^{-1/2} \hat{\sigma}_1 \hat{t}a_{(1-\alpha/2)}; \hat{\theta}_1 - n^{-1/2} \hat{\sigma}_1 \hat{t}a_{(\alpha/2)})$$



## Bootstrap em Regressão Linear

Considere o modelo univariado de regressão linear múltipla

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

em que  $\mathbf{Y}$  é o vetor  $n \times 1$  das observações da variável dependente,  $\mathbf{X}$  é a matriz  $n \times p$ , de rank  $p$ , das observações de  $p$  variáveis independentes,  $\beta$  é um vetor  $p \times 1$  de parâmetros desconhecidos e  $\epsilon$  é um vetor aleatório  $n \times 1$  de variáveis iid com distribuição comum  $F$  tendo média nula e variância finita  $\sigma^2$  desconhecida.



O estimador de mínimos quadrados ordinários (MQO) de  $\beta$  é o valor que minimiza

$$\begin{aligned}\text{SSE}(\beta) &= (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}^\top \mathbf{Y} - 2 \beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X} \beta\end{aligned}$$

cuja solução  $\hat{\beta}$  satisfaz o sistema de equações normais

$$(\mathbf{X}^\top \mathbf{X}) \hat{\beta} = \mathbf{X}^\top \mathbf{Y}.$$

É claro que

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

e obtida a estimativa  $\hat{\beta}$  de  $\beta$ , o vetor de resíduos estimados será

$$\hat{\epsilon} = \mathbf{Y} - \mathbf{X}\hat{\beta}.$$



Nesta etapa o processo de estimação bootstrap se inicia. Seja  $\hat{F}$  a função de distribuição estimada para a distribuição dos resíduos (por exemplo, a distribuição empírica dos resíduos  $\hat{\epsilon}_i, i = 1, \dots, n$ ). Obtêm-se uma amostra bootstrap  $\mathbf{Y}^*$  calculando-se

$$\mathbf{Y}^* = \mathbf{X}\hat{\beta} + \epsilon^*$$

em que  $\epsilon^{*\top} = (\epsilon_1^*, \dots, \epsilon_n^*)$  é o vetor de amostras iid de  $\hat{F}$ . O novo vetor  $\mathbf{Y}^*$  fornecerá uma estimativa bootstrap  $\hat{\beta}^*$  de  $\beta$ , calculada através da substituição de  $\mathbf{Y}$  por  $\mathbf{Y}^*$ , i.e.,

$$\hat{\beta}^* = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}^*.$$



## Obtenção das Amostras Bootstrap $Y^*$

### Modelo Paramétrico

Assumindo que os elementos do vetor  $\epsilon_{n \times 1}$  pertencem a uma distribuição  $F \sim N(0, \sigma^2)$ , podemos estimar a distribuição residual e obter  $\epsilon^*$  através do seguinte procedimento:

- 1 Estime  $\sigma^2$  através de MSE da estimação de  $\beta$  por MQO, i.e.,

$$\hat{\sigma}^2 = \text{MSE} \Rightarrow \hat{\sigma} = \sqrt{\text{MSE}},$$

(ou utilize o EMV de  $\sigma$ , que difere de  $\hat{\sigma}$  na utilização do fator  $\frac{1}{n}$  ao invés de  $\frac{1}{n-p}$  na fórmula para o cálculo do MSE).

- 2 Use um método convencional para gerar uma amostra  $\mathcal{Z} = \{z_1, \dots, z_n\}$  com  $n$  observações da distribuição  $N(0, 1)$ .
- 3 Obtenha  $\epsilon_i^* = \hat{\sigma} \times z_i \Rightarrow \epsilon^* \sim N(0, \hat{\sigma}^2 \mathbf{I})$ .

Desta forma se tivermos indícios suficientes de que a distribuição  $F$  é normal com média nula e variância  $\sigma^2$ , a distribuição  $\hat{F}$  será consistente para  $F$ .



## Obtenção das Amostras Bootstrap $Y^*$

### Modelo Não-Paramétrico

A abordagem não-paramétrica possui dois procedimentos de estimação para a distribuição residual  $F$ .

**Procedimento 1:**  $\hat{F}$  é a distribuição empírica dos resíduos resultantes da estimação por MQO. Assim  $\epsilon^{*T} = \{\epsilon_1^*, \dots, \epsilon_n^*\}$  será uma amostra aleatória tomada com reposição do vetor  $\hat{\epsilon}$  centralizado (a menos que  $X$  tenha uma coluna de 1's) resultante da estimação por MQO no modelo original.

Este procedimento tem a vantagem de estar livre de algumas suposições sobre a distribuição  $F$ . A desvantagem é não permitir que pontos não pertencendo à amostra original façam parte da amostra bootstrap, fazendo com que estruturas espúrias dos dados originais sejam reproduzidas na amostra bootstrap de  $\hat{\epsilon}$ .



## Obtenção das Amostras Bootstrap $Y^*$

### Modelo Não-Paramétrico

**Procedimento 2:** é uma alternativa intermediária baseada na utilização de um kernel. Nesta solução, a função de distribuição  $F$  deverá ter, adicionalmente, uma densidade e o resíduo bootstrap será uma estimativa suavizada de  $\hat{\epsilon}$ , simbolizada por  $\hat{\epsilon}_h^*$ .

Este procedimento é descrito em detalhes em von Borries<sup>7</sup> (1993) descreve este procedimento e apresenta alguns exemplos utilizando o SAS.

---

<sup>7</sup> von Borries, G. Método Bootstrap: Inferência Estatística e Aplicação em Modelos com Estrutura Não-Linear. Dissertação de Mestrado, UnB, 1993.



## Exemplos: Reamostragem-Bootstrap.R

- 1 Bootstrap Paramétrico e Não-Paramétrico.
- 2 Estimativa da Média Harmônica.
- 3 O pacote boot do R.
- 4 Bootstrap em Regressão Linear

