



XVIII CONGRESSO MUNDIAL DE EPIDEMIOLOGIA
VII CONGRESSO BRASILEIRO DE EPIDEMIOLOGIA

EPIDEMIOLOGIA NA CONSTRUÇÃO DA SAÚDE PARA TODOS:
MÉTODOS PARA UM MUNDO EM TRANSFORMAÇÃO

FIERGS - PORTO ALEGRE - BRASIL - 20 A 24 DE SETEMBRO DE 2008

COMPARANDO RESULTADOS USANDO IMPUTAÇÃO MÚLTIPLA


Luciana Nunes
Mariza Klück
Jandyra Fachel






Setembro - 2008



INTRODUÇÃO

 Dados faltantes são bastante comuns na prática!!!

➤ Problemas:

-  Perda de eficiência nas estimativas
-  Complicação na análise dos dados
-  Viés: respondentes \neq não respondentes

INTRODUÇÃO

👍 **Importante: escolha da abordagem analítica adequada para bancos de dados com observações incompletas.**

Métodos inadequados  **conclusões erradas**

Qual a solução??

😊 Restringir a análise aos sujeitos com dados completos

→ 😞 Viés nas estimativas

😊 Imputar os dados faltantes

O que é imputação???

Técnica usada para “preencher”
os dados faltantes com valores
plausíveis.

Dados faltantes???

Variáveis

Unidades

						
						
						
						
IMPUTAÇÃO						
						
						
						
						

Métodos de imputação

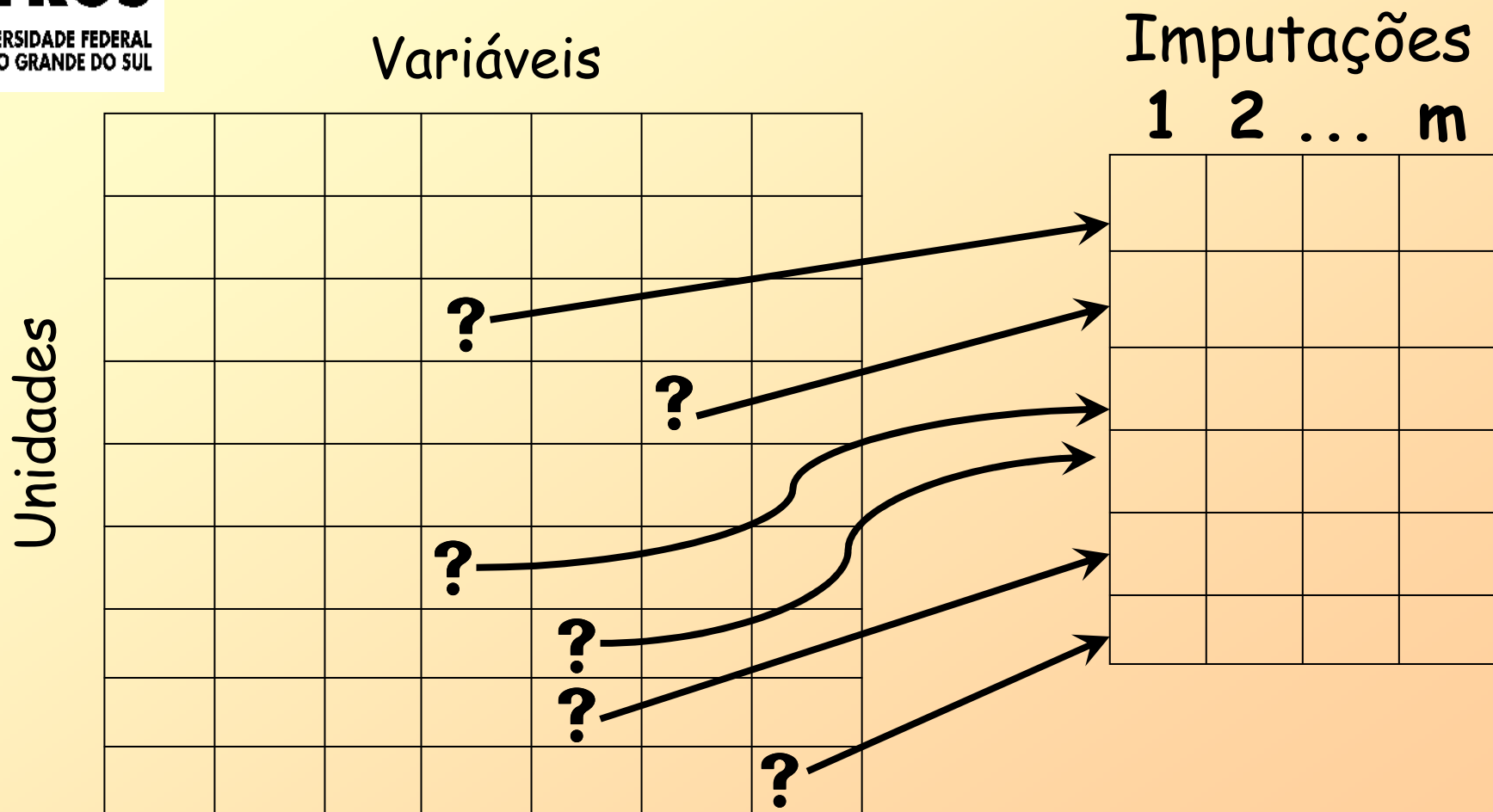
- ❖ Método da substituição (média, mediana)
- ❖ Hot Deck ("doadores")
- ❖ **Métodos de imputação única**
- ❖ Estimativa de máxima-verossimilhança
- ❖ Métodos para dados longitudinais

Imputação única

O dado ausente é preenchido uma única vez e então se utiliza o banco de dados “completo” para as análises.

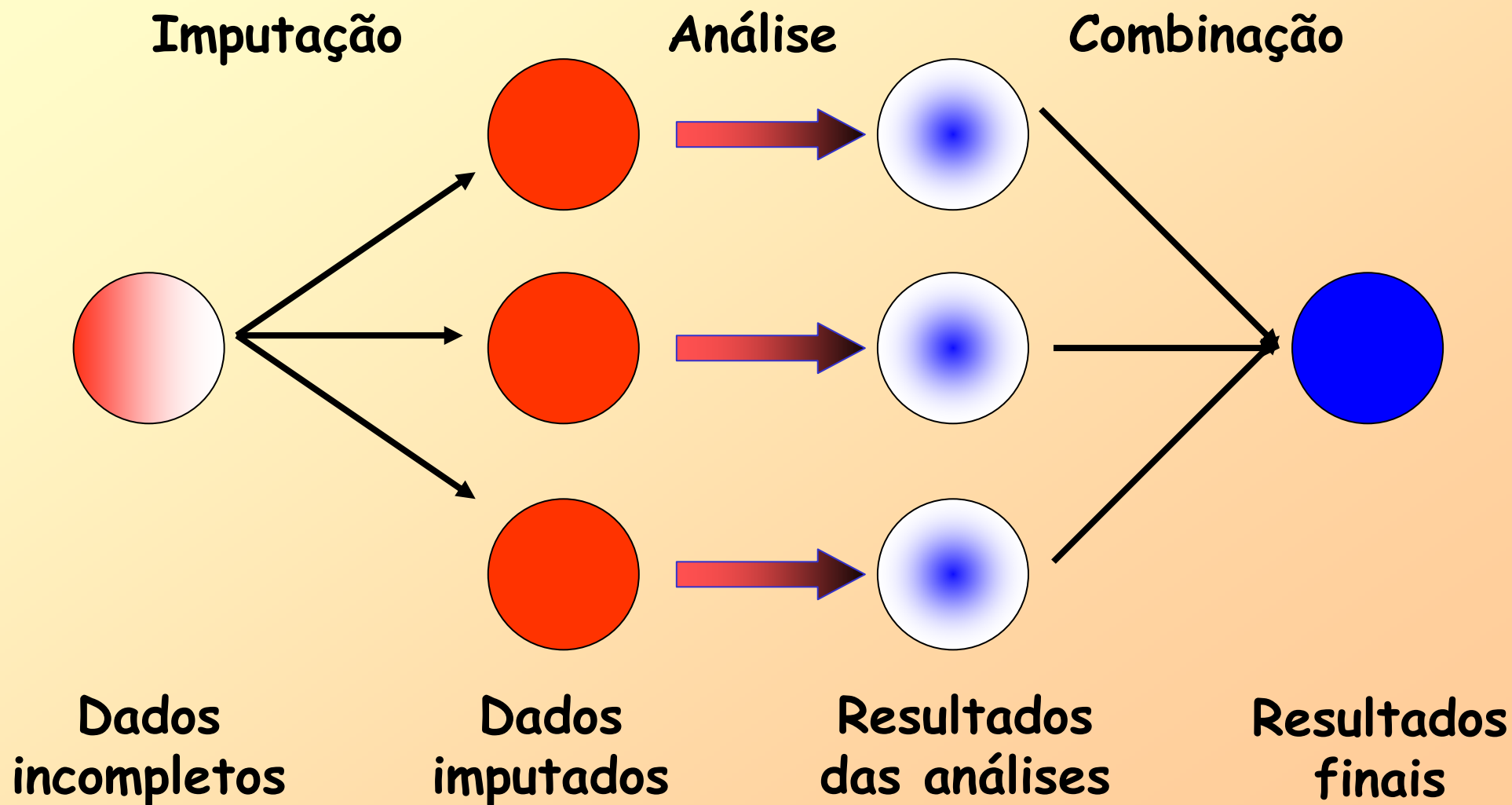
 A incerteza associada à imputação não é levada em conta!!!!

Imputação Múltipla (IM)



Conjunto de dados com m imputações

Idéia geral da IM



Regras de Rubin

A partir de cada análise obtém-se estimativas para o parâmetro de interesse Q_j , $j=1,2,\dots,m$, então a estimativa geral será:

$$\bar{Q} = \frac{1}{m} \sum_{j=1}^m \hat{Q}_j$$

E a variância total levará em conta a variância dentro e entre as imputações.

Objetivos

- ✓ Descrição da metodologia de imputação múltipla, gerando um roteiro para usuários da área da saúde, usando o pacote MICE do R.
- ✓ Comparação de dois métodos de imputação múltipla através de aplicação a dados reais.

Aplicação

❖ **Amostra:** 470 pacientes submetidos a laparotomia. Foram considerados somente os casos completos.

❖ **Fonte de dados:** prontuários de pacientes do HCPA no período de Fev/2000 a Dez/2003.

Bancos incompletos

- **BI-5:** aproximadamente **5%** dos valores da variável albumina foram excluídos aleatoriamente, permanecendo 440 casos completos.
- **BI-20:** aproximadamente **20%** dos valores da variável albumina foram excluídos aleatoriamente, permanecendo 383 casos completos.

Métodos de IM

Os modelos de IM são ajustados sob o paradigma Bayesiano em que, numa regressão linear, a variabilidade entre imputações é gerada através de “m” estimações de β e σ .

Preditive Mean Matching (PMM)

A idéia deste método é a seguinte:

$\hat{y}_{\text{faltante}}$ \longrightarrow Procura mais parecido entre os $\hat{y}_{\text{observados}}$

O valor a ser imputado é o $y_{\text{observado}}$ do
 $\hat{y}_{\text{observados}}$ mais parecido.

Bayesian Linear Regression (*BLR*)

Assim como no método *PMM*,
são estimados β e σ , entretanto os
"m" valores usados para as
imputações são os próprios $\hat{y}_{\text{faltantes}}$

Resultados

Tabela 1 - Estimativas da regressão logística para o banco de dados completo e bancos incompletos (BI-5 e BI-20). Desfecho óbito.

Variáveis independentes	OR [IC95%] e (Erro Padrão) dos modelos logísticos		
	Banco Completo (n=470)	BI-5 (n=440)	BI-20 (n=383)
ASA III	3,4[1,5;7,9] (0,422)	3,0[1,3;6,2] (0,428)	3,3[1,3;8,7] (0,489)
ASA IV/V	20,2[8,8;46,0] (0,421)	16,4[7,1;37,8] (0,426)	22,3[8,6;57,6] (0,484)
Idade ≥ 75	2,9[1,5;5,8] (0,348)	2,7[1,3;5,6] (0,363)	4,0[1,8;8,9] (0,405)
Alb até 2,2g/dl	5,3[2,7;10,5] (0,349)	5,6[2,7;11,7] (0,368)	7,0[3,1;15,9] (0,420)
Alb 2,3 a 3,0g/dl	2,1[1,1;4,1] (0,345)	2,0[1,0;4,2] (0,364)	2,6[1,2;5,9] (0,412)

Resultados

Tabela 4 - Estimativas da regressão logística após imputações múltiplas pelo método *PMM* em diferentes regressões. BI-20 (n=383)

Variáveis independentes	OR [IC95%] e (Erro Padrão) dos modelos logísticos			
	Banco Completo	IM(1)	IM(2)	IM(3)
ASA III	3,4[1,5;7,9] (0,422)	3,3[1,4;7,5] (0,423)	3,3[1,4;7,7] (0,427)	3,3[1,4;7,5] (0,423)
ASA IV/V	20,2[8,8;46,0] (0,421)	20,5[8,9;47,1] (0,425)	20,8[9,0;48,1] (0,427)	20,5[8,9;47,2] (0,425)
Idade ≥ 75	2,9[1,5;5,8] (0,348)	3,1[1,6;6,1] (0,349)	3,1[1,5;6,1] (0,350)	3,1[1,6;6,1] (0,349)
Alb até 2,2g/dl	5,3[2,7;10,5] (0,349)	4,6[2,1;10,1] (0,403)	4,5[1,9;10,9] (0,448)	4,2[2,1;8,8] (0,370)
Alb 2,3 a 3,0g/dl	2,1[1,1;4,1] (0,345)	1,9[0,8;4,2] (0,413)	2,1[0,9;4,6] (0,407)	1,8[0,8;3,8] (0,382)

Discussão

☺ Restringir análise aos casos completos pode levar a conclusões erradas.

☺ Na literatura tem sido recomendado que os dados faltantes sejam imputados.

☺ Imputação evita perda de poder.

Considerações

finais

Imputar ou não imputar?



IMPUTAR



Imputação única ou múltipla?



IMPUTAÇÃO MÚLTIPLA

Imputação múltipla

- **Complexa? Sim**
- **Avanços computacionais**
- **Muita produção científica**
- **Pouco usada na área da Saúde**
- **Menos ainda na literatura brasileira**

Obrigada!!!

