

Como montar meu conjunto de dados?

Eduardo Elias Ribeiro Junior Henrique Aparecido Laureano

Faculdade de Medicina
Universidade Federal de Minas Gerais - UFMG

16 de agosto de 2016

Roteiro

1. O que é um banco de dados?
2. Qual o ponto de partida?
3. Como organizar meu banco de dados?
4. Alguns exemplos

O que é um banco de dados?

Definição

*Um banco de dados é uma coleção organizada de dados que se relaciona de forma a criar algum sentido (informação) e dar mais eficiência durante uma pesquisa ou estudo.*¹

Simplificando:

Banco de dados é uma coleção de dados interligados entre si e organizados para fornecer informações.

¹<https://pt.wikipedia.org>

Dados *vs* Informações

Dados \neq Informações

Dados:

Fatos brutos, em sua forma primária. Muitas vezes os dados podem não fazer sentido sozinhos.

Informações:

Consiste no agrupamento de dados de forma organizada para fazer sentido, gerar conhecimento.

Um banco de dados é uma estrutura de dados organizada que permite a extração de informações.

Qual o ponto de partida?

Definições iniciais

- ▶ Qual o objetivo do estudo?

Definições iniciais

- ▶ Qual o objetivo do estudo?
 - ▶ O que se deseja estudar?

Definições iniciais

- ▶ Qual o objetivo do estudo?
 - ▶ O que se deseja estudar?
 - ▶ Qual a hipótese a ser testada?

Definições iniciais

- ▶ Qual o objetivo do estudo?
 - ▶ O que se deseja estudar?
 - ▶ Qual a hipótese a ser testada?
 - ▶ Qual(is) possível(is) diferença(s) deseja-se verificar?

Definições iniciais

- ▶ Qual o objetivo do estudo?
 - ▶ O que se deseja estudar?
 - ▶ Qual a hipótese a ser testada?
 - ▶ Qual(is) possível(is) diferença(s) deseja-se verificar?
- ▶ Com seus objetivos definidos, quais características dos pacientes precisam ser avaliadas/mensuradas?

Definições iniciais

- ▶ Qual o objetivo do estudo?
 - ▶ O que se deseja estudar?
 - ▶ Qual a hipótese a ser testada?
 - ▶ Qual(is) possível(is) diferença(s) deseja-se verificar?
- ▶ Com seus objetivos definidos, quais características dos pacientes precisam ser avaliadas/mensuradas?
 - ▶ Se for para pecar, peque por excesso! É preferível ter mais informações mensuradas. Assim não se corre o risco de inviabilizar uma possível análise pela ausência do registro de informações.

Como organizar meu banco de dados?

Como organizar meu banco de dados?

- ▶ Nas linhas as observações (unidades experimentais/amostrais: pacientes, indivíduos, planta, etc.);

Como organizar meu banco de dados?

- ▶ Nas linhas as observações (unidades experimentais/amostrais: pacientes, indivíduos, planta, etc.);
- ▶ Nas colunas suas características (informações avaliadas/mensuradas: idade, peso, qtde de fertilizante, etc.);

Como organizar meu banco de dados?

- ▶ Nas linhas as observações (unidades experimentais/amostrais: pacientes, indivíduos, planta, etc.);
- ▶ Nas colunas suas características (informações avaliadas/mensuradas: idade, peso, qtde de fertilizante, etc.);
- ▶ Cada característica mensurada deve ter sua própria coluna na tabela de dados;

Como organizar meu banco de dados?

- ▶ Nas linhas as observações (unidades experimentais/amostrais: pacientes, indivíduos, planta, etc.);
- ▶ Nas colunas suas características (informações avaliadas/mensuradas: idade, peso, qtde de fertilizante, etc.);
- ▶ Cada característica mensurada deve ter sua própria coluna na tabela de dados;
- ▶ Procure atribuir nomes concisos às informações;

Como organizar meu banco de dados?

- ▶ Nas linhas as observações (unidades experimentais/amostrais: pacientes, indivíduos, planta, etc.);
- ▶ Nas colunas suas características (informações avaliadas/mensuradas: idade, peso, qtde de fertilizante, etc.);
- ▶ Cada característica mensurada deve ter sua própria coluna na tabela de dados;
- ▶ Procure atribuir nomes concisos às informações;
- ▶ E se o paciente foi avaliado mais de uma vez em ao menos uma característica?

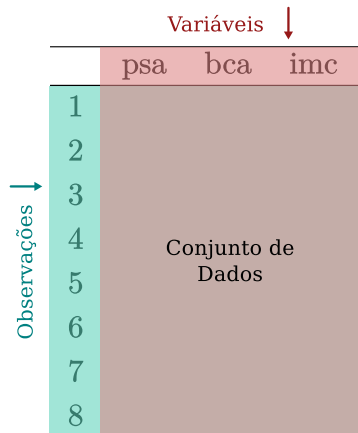
Como organizar meu banco de dados?

- ▶ Nas linhas as observações (unidades experimentais/amostrais: pacientes, indivíduos, planta, etc.);
- ▶ Nas colunas suas características (informações avaliadas/mensuradas: idade, peso, qtde de fertilizante, etc.);
- ▶ Cada característica mensurada deve ter sua própria coluna na tabela de dados;
- ▶ Procure atribuir nomes concisos às informações;
- ▶ E se o paciente foi avaliado mais de uma vez em ao menos uma característica?
 - ▶ Ele deve receber uma nova linha na tabela de dados, de preferência logo na linha abaixo.

Como organizar meu banco de dados?

- ▶ Nas linhas as observações (unidades experimentais/amostrais: pacientes, indivíduos, planta, etc.);
- ▶ Nas colunas suas características (informações avaliadas/mensuradas: idade, peso, qtde de fertilizante, etc.);
- ▶ Cada característica mensurada deve ter sua própria coluna na tabela de dados;
- ▶ Procure atribuir nomes concisos às informações;
- ▶ E se o paciente foi avaliado mais de uma vez em ao menos uma característica?
 - ▶ Ele deve receber uma nova linha na tabela de dados, de preferência logo na linha abaixo.
 - ▶ Nas características que não foram novamente avaliadas repete-se o valor (observação).

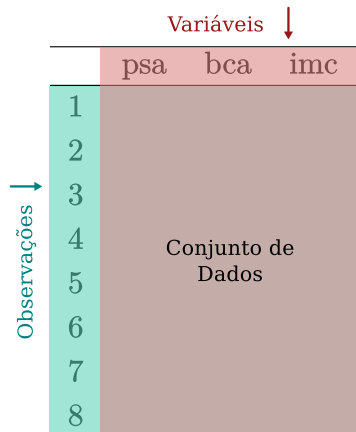
Como organizar meu banco de dados?



The diagram illustrates a data set structure. It features a table with a header row and eight data rows. The header row is labeled 'Variáveis' (Variables) with a downward arrow pointing to the column headers 'psa', 'bca', and 'imc'. The data rows are labeled 'Observações' (Observations) with a downward arrow pointing to the row indices 1 through 8. The main body of the table is a large brown rectangle labeled 'Conjunto de Dados' (Data Set).

Variáveis ↓		psa	bca	imc
Observações ↓	1	Conjunto de Dados		
	2			
	3			
	4			
	5			
	6			
	7			
	8			

Como organizar meu banco de dados?



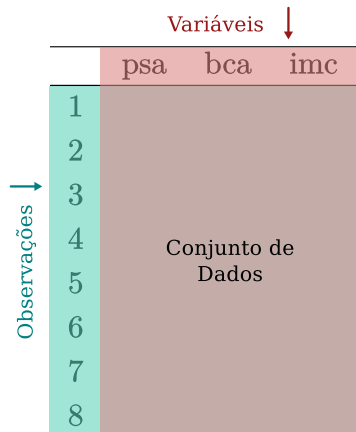
The diagram illustrates a data set structure. It features a table with a header row and eight data rows. The header row is labeled 'Variáveis' (Variables) with a red arrow pointing down to it, and contains three columns: 'psa', 'bca', and 'imc'. The data rows are labeled 'Observações' (Observations) with a teal arrow pointing right to them, and are numbered 1 through 8. The entire data area is labeled 'Conjunto de Dados' (Data Set) in the center. The header row has a light red background, and the data rows have a light teal background.

	psa	bca	imc
1			
2			
3			
4			
5			
6			
7			
8			

Dicas importantes:

- Documente seu conjunto de dados (elabore um dicionário explicando as características/variáveis mensuradas);

Como organizar meu banco de dados?



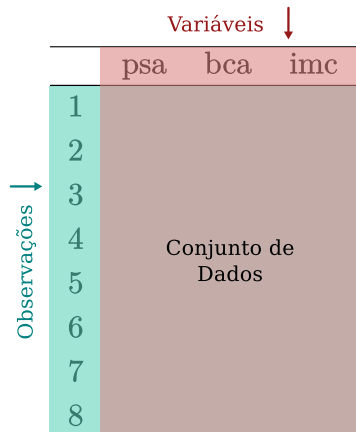
The diagram illustrates a data set structure. At the top, the word 'Variáveis' is written in red, with a red arrow pointing down to the header row of a table. The header row contains three columns: 'psa', 'bca', and 'imc'. To the left of the table, the word 'Observações' is written in teal, with a teal arrow pointing down to the first column of the table. The first column contains numbers 1 through 8. The main body of the table is a large brown rectangle labeled 'Conjunto de Dados' in the center.

	psa	bca	imc
1	Conjunto de Dados		
2			
3			
4			
5			
6			
7			
8			

Dicas importantes:

- Documente seu conjunto de dados (elabore um dicionário explicando as características/variáveis mensuradas);
- Seja cuidadoso ao preencher a tabela de dados.
Siga um padrão para preencher as variáveis categóricas!

Como organizar meu banco de dados?



The diagram illustrates a data table structure. A red arrow labeled 'Variáveis' points to the header row, which contains the variable names 'psa', 'bca', and 'imc'. A teal arrow labeled 'Observações' points to the first column, which contains observation numbers 1 through 8. The main body of the table is a large brown rectangle labeled 'Conjunto de Dados'.

	psa	bca	imc
1			
2			
3			
4			
5			
6			
7			
8			

Dicas importantes:

- Documente seu conjunto de dados (elabore um dicionário explicando as características/variáveis mensuradas);
- Seja cuidadoso ao preencher a tabela de dados.
Siga um padrão para preencher as variáveis categóricas!
- Não resuma informação!

Alguns exemplos

Diabetes em descendentes da tribo indígena Pima

- ▶ id: identificador do paciente
- ▶ npreg: número de gestações
- ▶ glu: concentração de glicose no plasma
- ▶ bp: pressão sanguínea
- ▶ skin: espessura da prega cutânea no tríceps (mm)
- ▶ bmi: índice de massa corporal
- ▶ ped: diabetes pedigree
- ▶ age: idade
- ▶ type: yes ou no para diabetes

id	npreg	glu	bp	skin	bmi	ped	age	type
1	5	86	68	28	30.20	0.36	24	No
2	7	195	70	33	25.10	0.16	55	Yes
3	5	77	82	41	35.80	0.16	35	No
4	0	165	76	43	47.90	0.26	26	No
5	0	107	60	25	26.40	0.13	23	No
6	5	97	76	27	35.60	0.38	52	Yes
7	3	83	58	31	34.30	0.34	25	No

Monitoramento de transplantes (trans) do coração

- ▶ id: identificador do paciente
- ▶ age: idade na hora do trans
- ▶ years: anos após o trans
- ▶ dage: idade do doador
- ▶ sex: sexo (0 = fem, 1 = masc)
- ▶ pdiag: motivo do trans
- ▶ cumrej: soma de episódios de rejeição aguda
- ▶ st: estado na hora da consulta
- ▶ fobs: trans (0 = não, 1 = sim)
- ▶ stmax: estado máximo observado

id	age	years	dage	sex	pdiag	cumrej	st	fobs	stmax
1	52.50	0.00	21	0	IHD	0	1	1	1
1	53.50	1.00	21	0	IHD	2	1	0	1
1	54.50	2.00	21	0	IHD	2	2	0	2
1	55.59	3.09	21	0	IHD	2	2	0	2
1	56.50	4.00	21	0	IHD	3	2	0	2
1	57.49	5.00	21	0	IHD	3	3	0	3
1	58.35	5.85	21	0	IHD	3	4	0	4

10 anos de cirurgia colorretal: complicações e fatores de risco