Como montar meu conjunto de dados?

Eduardo Elias Ribeiro Junior Henrique Aparecido Laureano

Faculdade de Medicina
Universide Federal de Minas Gerais - UFMG

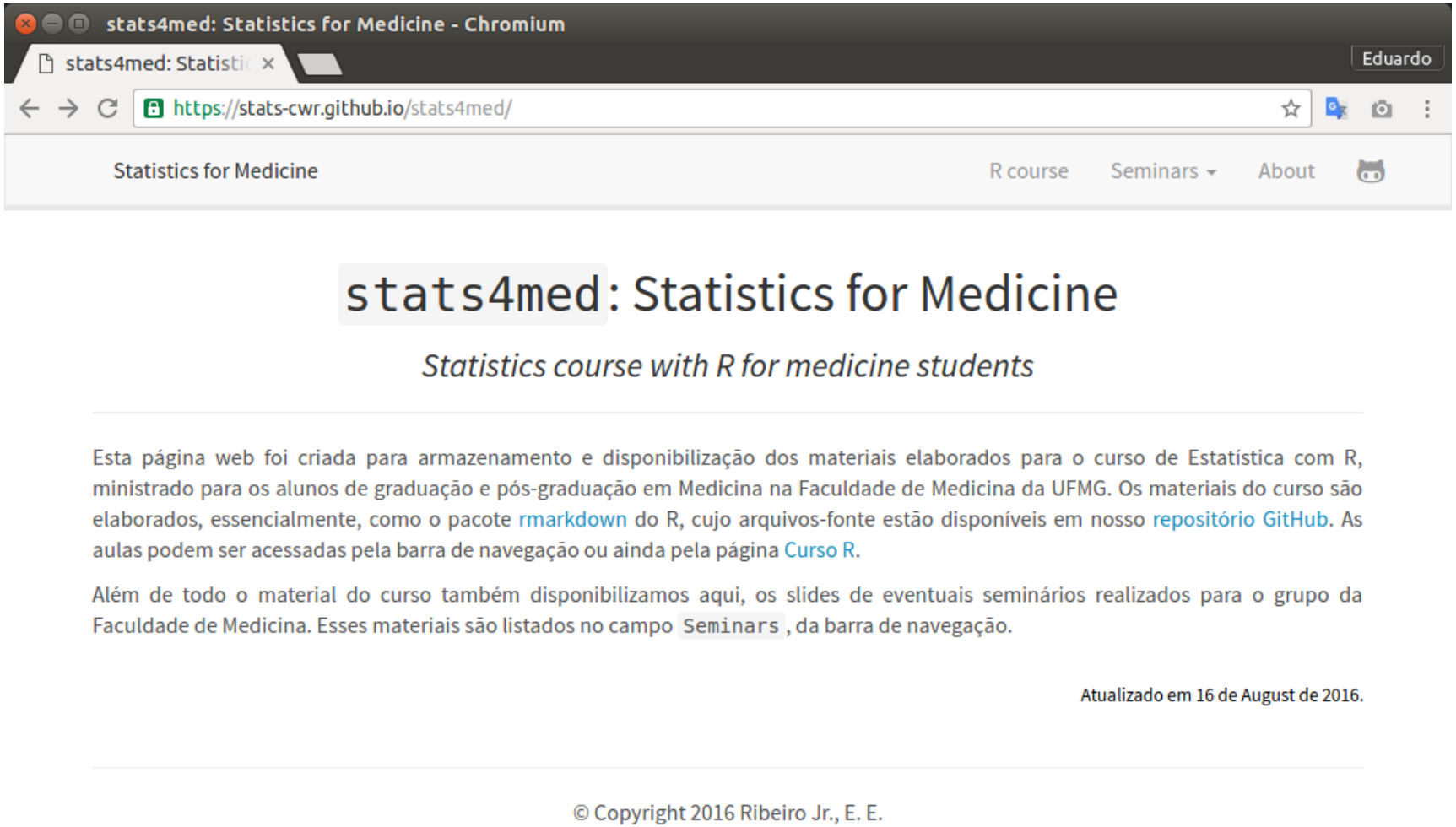
16 de agosto de 2016

Roteiro

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

Disponibilização

`https://stats-cwr.github.io/stats4med`



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 conjunto de dados?

Como organizar meu conjunto de dados?

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

Como organizar meu conjunto 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 conjunto 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 conjunto 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 conjunto 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?

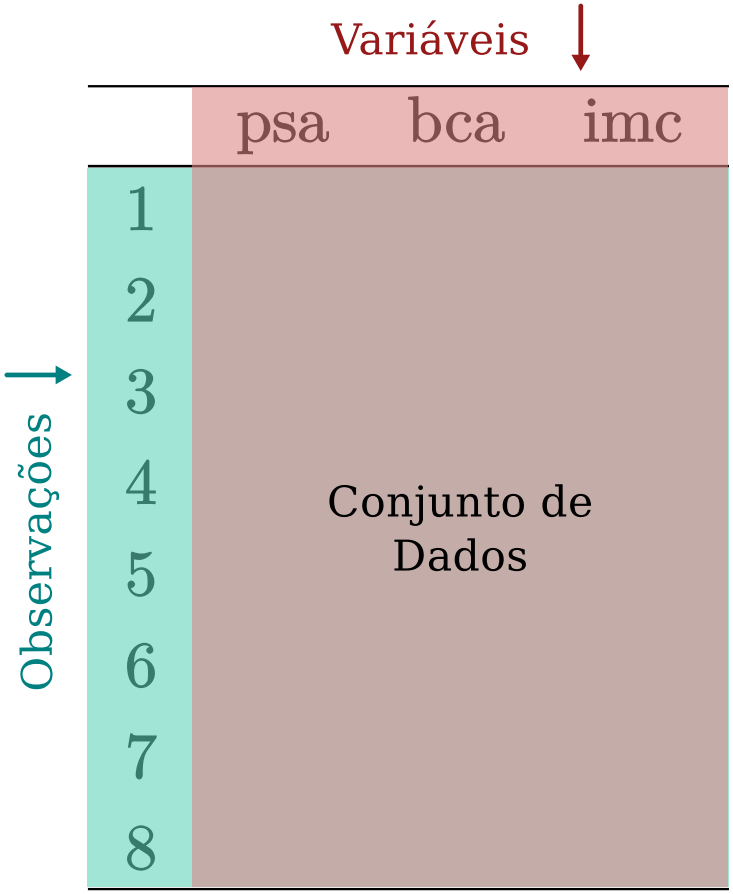
Como organizar meu conjunto de dados?

- 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.

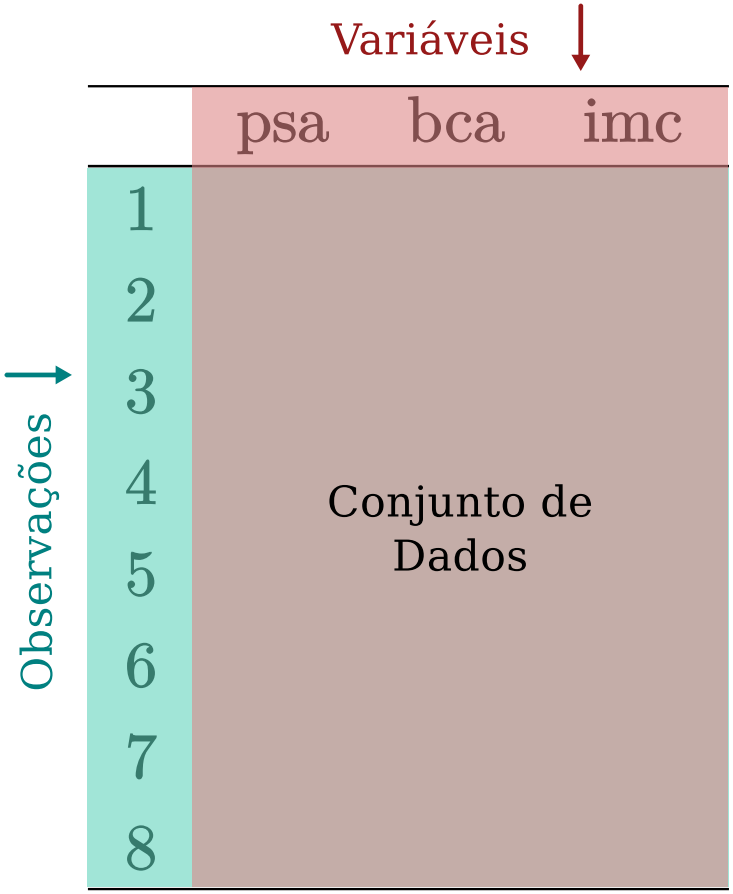
Como organizar meu conjunto de dados?

- 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?



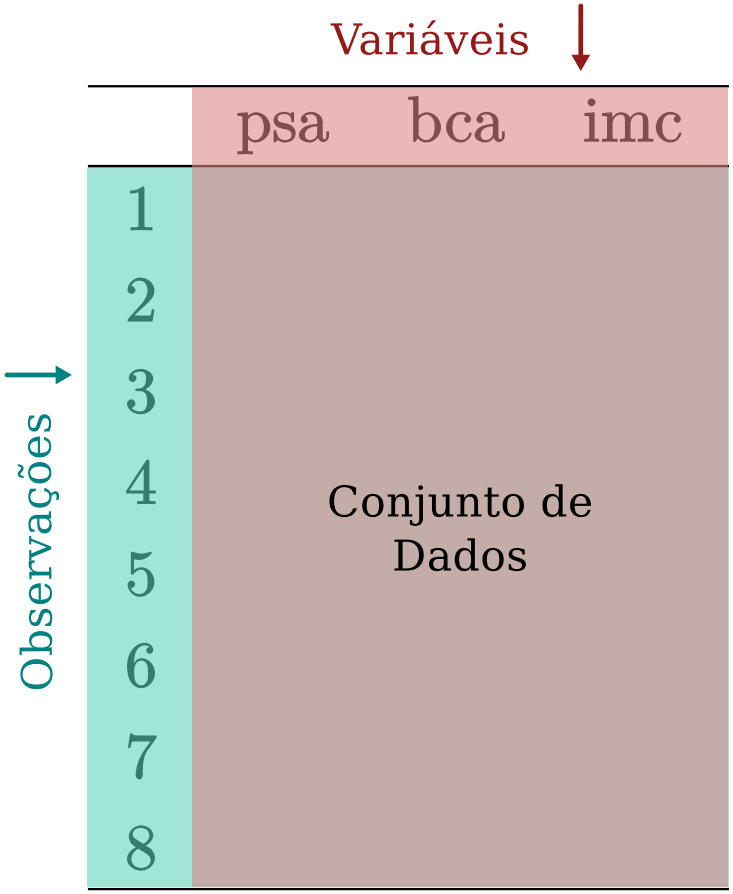
Como organizar meu banco de dados?



Dicas importantes:

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

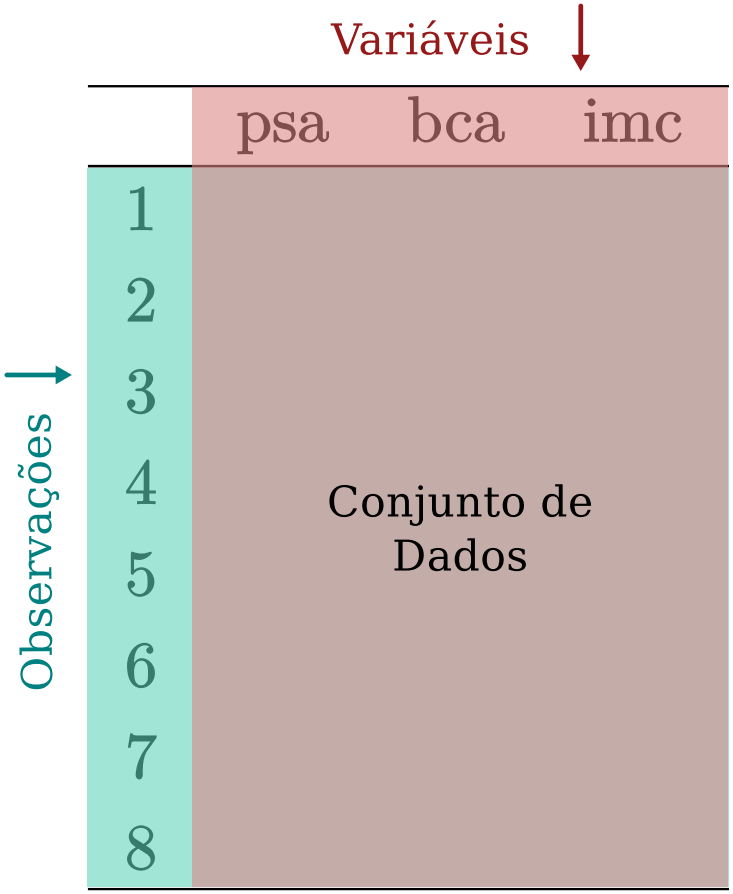
Como organizar meu banco de dados?



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 as variáveis categóricas!

Como organizar meu banco de dados?



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 as variáveis categóricas!
- Não resuma os conjunto de dados! Isto será feito posteriormente, na análise.

Alguns exemplos

Diabetes em descendentes da tribo indígina 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

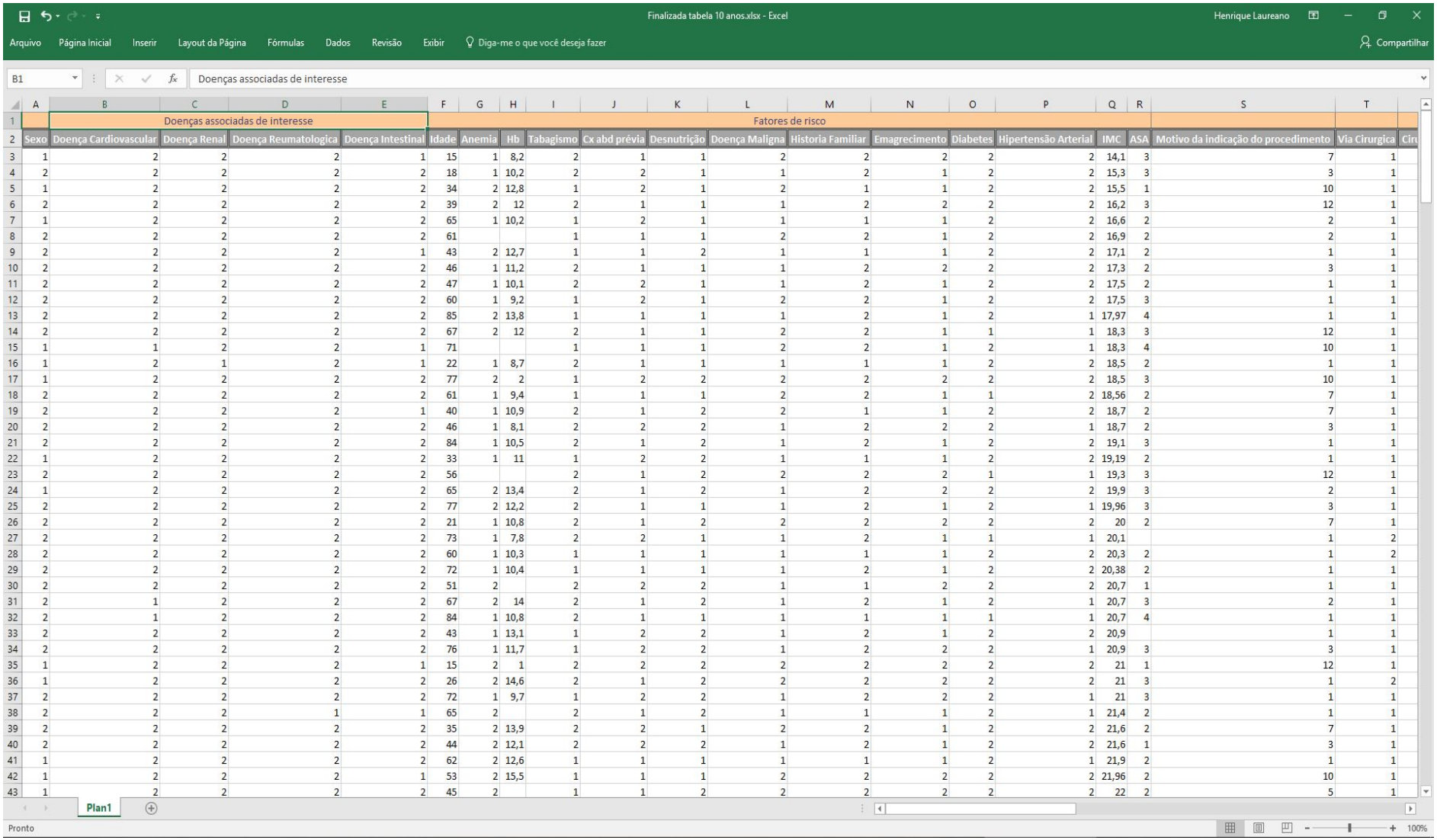
id	npreg	glu	bp	skin	bmi	ped	age	type
1	5	86	68	28	30.20	0.36	24	No
2	7	195	70	33	25.10	0.16	55	Yes
3	5	77	82	41	35.80	0.16	35	No
4	0	165	76	43	47.90	0.26	26	No
5	0	107	60	25	26.40	0.13	23	No
6	5	97	76	27	35.60	0.38	52	Yes
7	3	83	58	31	34.30	0.34	25	No

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



Finalizada tabela 10 anos.xlsx - Excel — Henrique Laureano

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibir | Diga-me o que você deseja fazer | Compartilhar

B1: Doenças associadas de interesse

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		Doenças associadas de interesse				Fatores de risco														
2	Sexo	Doença Cardiovascular	Doença Renal	Doença Reumatologica	Doença Intestinal	Idade	Anemia	Hb	Tabagismo	Cx abd prévia	Desnutrição	Doença Maligna	Historia Familiar	Emagrecimento	Diabetes	Hipertensão Arterial	IMC	ASA	Motivo da indicação do procedimento	Via Cirurgica
3	1	2	2	2	1	15	1	8,2	2	1	1	2	2	2	2	2	14,1	3	7	1
4	2	2	2	2	2	18	1	10,2	2	2	1	1	2	1	2	2	15,3	3	3	1
5	1	2	2	2	2	34	2	12,8	1	2	1	2	1	1	2	2	15,5	1	10	1
6	2	2	2	2	2	39	2	12	2	1	1	1	2	2	2	2	16,2	3	12	1
7	1	2	2	2	2	65	1	10,2	1	2	1	1	1	1	2	2	16,6	2	2	1
8	2	2	2	2	2	61			1	1	1	2	2	1	2	2	16,9	2	2	1
9	2	2	2	2	1	43	2	12,7	1	1	2	1	1	1	2	2	17,1	2	1	1
10	2	2	2	2	2	46	1	11,2	2	1	1	1	2	2	2	2	17,3	2	3	1
11	2	2	2	2	2	47	1	10,1	2	2	1	1	2	1	2	2	17,5	2	1	1
12	2	2	2	2	2	60	1	9,2	1	2	1	2	2	1	2	2	17,5	3	1	1
13	2	2	2	2	2	85	2	13,8	1	1	1	1	2	1	2	1	17,97	4	1	1
14	2	2	2	2	2	67	2	12	2	1	1	2	2	1	1	1	18,3	3	12	1
15	1	1	2	2	1	71			1	1	1	2	2	1	2	1	18,3	4	10	1
16	1	2	1	2	1	22	1	8,7	2	1	1	1	1	1	2	2	18,5	2	1	1
17	1	2	2	2	2	77	2	2	1	2	2	2	2	2	2	2	18,5	3	10	1
18	2	2	2	2	2	61	1	9,4	1	1	1	2	2	1	1	2	18,56	2	7	1
19	2	2	2	2	1	40	1	10,9	2	1	2	2	1	1	2	2	18,7	2	7	1
20	2	2	2	2	2	46	1	8,1	2	2	2	1	2	2	2	1	18,7	2	3	1
21	2	2	2	2	2	84	1	10,5	2	1	2	1	2	1	2	2	19,1	3	1	1
22	1	2	2	2	2	33	1	11	1	2	2	1	1	1	2	2	19,19	2	1	1
23	2	2	2	2	2	56			2	1	2	2	2	2	1	1	19,3	3	12	1
24	1	2	2	2	2	65	2	13,4	2	1	2	1	2	2	2	2	19,9	3	2	1
25	2	2	2	2	2	77	2	12,2	2	1	1	1	2	1	2	1	19,96	3	3	1
26	2	2	2	2	2	21	1	10,8	2	1	2	2	2	2	2	2	20	2	7	1
27	2	2	2	2	2	73	1	7,8	2	2	1	1	2	1	1	1	20,1		1	2
28	2	2	2	2	2	60	1	10,3	1	1	1	1	1	1	2	2	20,3	2	1	2
29	2	2	2	2	2	72	1	10,4	1	1	1	1	2	1	2	2	20,38	2	1	1
30	2	2	2	2	2	51	2		2	2	2	1	1	2	2	2	20,7	1	1	1
31	2	1	2	2	2	67	2	14	2	1	2	1	2	1	2	1	20,7	3	2	1
32	2	1	2	2	2	84	1	10,8	2	1	1	1	1	1	1	1	20,7	4	1	1
33	2	2	2	2	2	43	1	13,1	1	2	2	1	2	1	2	2	20,9		1	1
34	2	2	2	2	2	76	1	11,7	1	2	2	1	2	2	2	1	20,9	3	3	1
35	1	2	2	2	1	15	2	1	2	2	2	2	2	2	2	2	21	1	12	1
36	1	2	2	2	2	26	2	14,6	2	1	2	2	2	2	2	2	21	3	1	2
37	2	2	2	2	2	72	1	9,7	1	2	2	1	2	2	2	1	21	3	1	1
38	2	2	2	1	1	65	2		2	1	2	1	1	1	2	1	21,4	2	1	1
39	2	2	2	2	2	35	2	13,9	2	2	1	2	2	1	2	2	21,6	2	7	1
40	2	2	2	2	2	44	2	12,1	2	2	2	1	2	1	2	2	21,6	1	3	1
41	1	2	2	2	2	62	2	12,6	1	1	1	1	1	1	2	1	21,9	2	1	1
42	1	2	2	2	1	53	2	15,5	1	1	1	2	2	2	2	2	21,96	2	10	1
43	1	2	2	2	2	45	2		1	1	2	2	2	2	2	2	22	2	5	1

Plan1

Pronto — 100%