

Resultados do projeto “Estatística descritiva de um banco de dados de pacientes pediátricos com tumores do sistema nervoso central”

Um caderno aberto de pesquisa

Francisco Hélder Cavalcante Félix, Centro Pediátrico do Câncer - Hospital Infantil Albert Sabin

Abstract: Um banco de dados de pacientes pediátricos com tumores do sistema nervoso central diagnosticados em um grande hospital estadual foi criado pelos autores. Os resultados e análises foram colocados num repositório do serviço GitHub. O código das análises foi checado com o serviço de integração contínua em nuvem Travis Ci e o resultado final publicado neste arquivo. Este arquivo traz tabelas, gráficos e texto mostrando os resultados do projeto. Foi elaborado em Rmarkdown, utilizando a linguagem de marcação simplificada Markdown com “pedaços” de código da linguagem estatística R entremeados. O arquivo foi avaliado pelo pacote rmarkdown e compilado para o formato pdf neste texto. Devido à integração contínua, pode ser atualizado em tempo real, enquanto dados novos são acrescentados.

Keywords: tumores do sistema nervoso central, cancerologia pediátrica, estatística descritiva, rmarkdown, integração contínua, ciência aberta

February 13, 2018

Introdução:

A ciência aberta baseia-se principalmente na capacidade de divulgar (compartilhar) eletronicamente as informações coletadas (dados brutos) e produzidas (análises e seus resultados) de um projeto de pesquisa através da internet. Dessa forma, 2 consequências advêm imediatamente: 1 - Transparência da informação e do processo científicos, inclusive para públicos não técnicos. 2 - Capacidade irrestrita de comentários, tanto por especialistas (análogo à *revisão por pares*), quanto por não especialistas (que poderíamos chamar de *revisão cidadã*).

O observador arguto já pode levantar a questão de que o controle de comentários numa plataforma é dos controladores daquele serviço, ou seja, é possível criar um canal de comunicação tipo “ciência aberta”, porém unidirecional, sem recepção de comentários (ou pior, é possível censurá-los). Independentemente disso, os comentários a uma publicação livremente disponível na internet podem ser publicados em qualquer canal sem relação com o canal original e referenciado ao primeiro. Ou seja, não há como verdadeiramente censurar comentários a uma publicação livre na internet.

Outra característica importante da ciência aberta é a capacidade de *reuso* de informações, o que pode ser entendido como a principal utilidade social da ciência aberta. Esta característica simples tem o potencial de otimizar a produção científica a nível global. Bastaria isso para justificar a implementação em larga escala da ciência aberta. Outros benefícios podem ser descritos de forma ilimitada.

Esta é a implementação de ciência aberta que criei, baseando-me largamente em projetos já existentes. Trata-se de um caderno de pesquisa aberto, armazenado num repositório remoto para o programa Git (existem vários), gerado através de um serviço de integração contínua (CI) em nuvem (vários idem) e com a estrutura de um pacote da linguagem estatística R, usada para as análises. *Não se trata de um pacote verdadeiro*, apesar de ter um diretório de código R e um arquivo de definições DESCRIPTION. O objetivo desse mimetismo é facilitar as análises numa plataforma de CI. Um pacote de R é um programa com funções utilizáveis. Não é isso que este(s) caderno(s) é(são). Assim, propositadamente deixo de fora partes imprescindíveis de um pacote, como o NAMESPACE e os manuais.

Na seção a seguir, são mostrados resultados de análises estatísticas concernentes a este caderno aberto de pesquisa em particular. Todos os dados pertinentes a seres humanos são adequadamente desidentificados.

Análises:

```
require(pander)
require(survival)
snc<-read.csv('../data/snc.csv')
attach(snc)
```

```
barplot(summary(as.factor(snc$sex)),names.arg=c("masculino","feminino"),
xlab="Sexo",width=0.5,xlim=c(0,1.7),space=0.5,col=0)
```

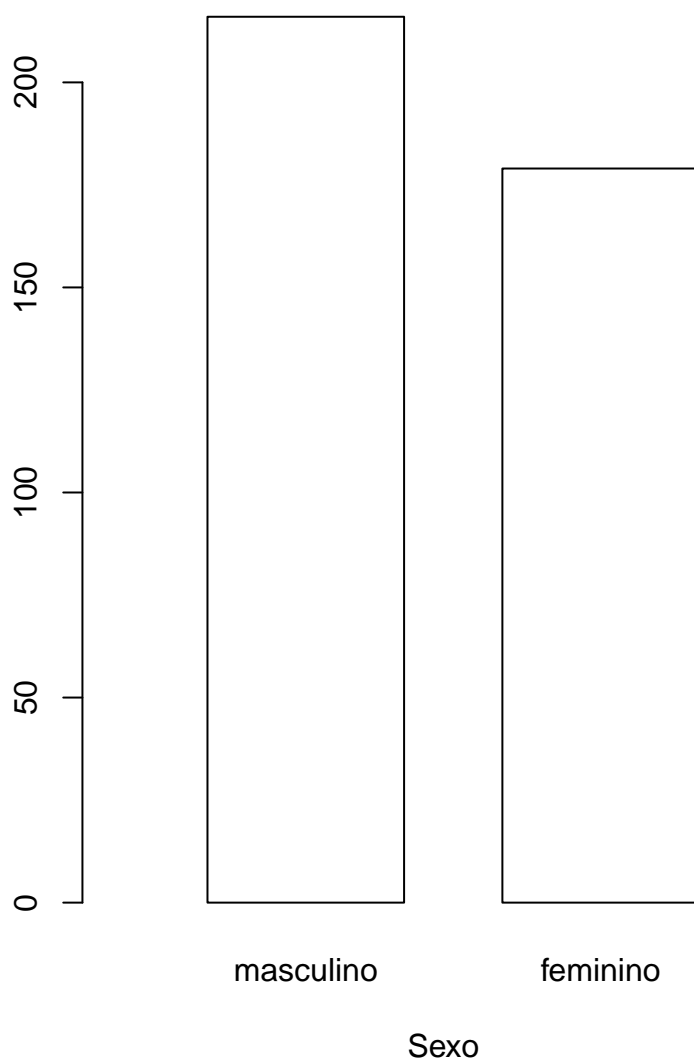


Figura 1: número de pacientes segundo o sexo.

```
boxplot(snc$age/365.25,xlab="Idade (anos)",boxwex=0.6,staplewex=0.4,
frame.plot=F)
```

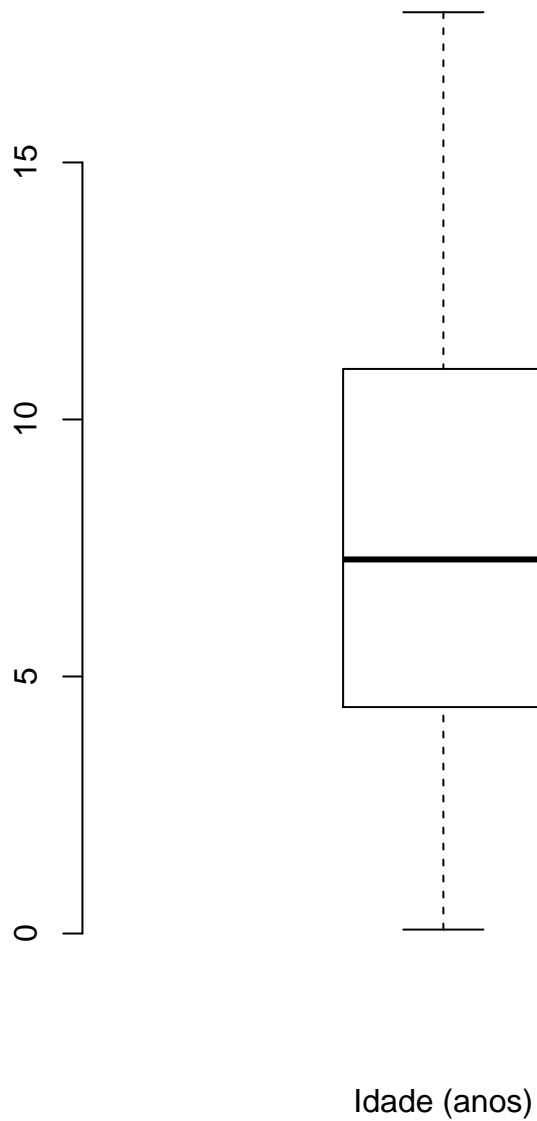


Figura 2: idade dos pacientes ao diagnóstico.

```
require(DescTools)
panderOptions('table.split.table', Inf)
set.caption("Procedência dos pacientes")
pander(Freq(origin), style = 'rmarkdown')
```

Table 1: Procedência dos pacientes

level	freq	perc	cumfreq	cumperc
Abaiara	1	0.002532	1	0.002532
Acarape	1	0.002532	2	0.005063
Acarau	1	0.002532	3	0.007595
Acopiara	2	0.005063	5	0.01266
Aiuaba	1	0.002532	6	0.01519
Alcantaras	1	0.002532	7	0.01772
Alto Santo	1	0.002532	8	0.02025
Amontada	2	0.005063	10	0.02532
Apuiaries	2	0.005063	12	0.03038
Aquiraz	3	0.007595	15	0.03797
Aracati	2	0.005063	17	0.04304
Aracoiaba	3	0.007595	20	0.05063
Ararende	1	0.002532	21	0.05316
Aurora	1	0.002532	22	0.0557
Barbalha	3	0.007595	25	0.06329
Barreira	3	0.007595	28	0.07089
Barro	2	0.005063	30	0.07595
Baturite	2	0.005063	32	0.08101
Beberibe	2	0.005063	34	0.08608
Bela Cruz	2	0.005063	36	0.09114
Boa Viagem	1	0.002532	37	0.09367
Boa Vista	1	0.002532	38	0.0962
Camocim	2	0.005063	40	0.1013
Caninde	7	0.01772	47	0.119
Capistrano	3	0.007595	50	0.1266
Carire	2	0.005063	52	0.1316
Cascavel	1	0.002532	53	0.1342
Catunda	1	0.002532	54	0.1367
Caucaia	21	0.05316	75	0.1899
Cedro	2	0.005063	77	0.1949
Chaval	1	0.002532	78	0.1975
Crateus	1	0.002532	79	0.2
Crato	3	0.007595	82	0.2076
Cruz	1	0.002532	83	0.2101
Eusebio	4	0.01013	87	0.2203
Farias Brito	1	0.002532	88	0.2228
Forquilha	2	0.005063	90	0.2278
Fortaleza	131	0.3316	221	0.5595
Frecheirinha	1	0.002532	222	0.562
Graca	1	0.002532	223	0.5646
Granja	1	0.002532	224	0.5671
Granjeiro	1	0.002532	225	0.5696
Guaiuba	3	0.007595	228	0.5772
Guaraciaba Do Norte	2	0.005063	230	0.5823
Hidrolandia	1	0.002532	231	0.5848
Horizonte	4	0.01013	235	0.5949
Ibaretama	1	0.002532	236	0.5975
Ibicuitinga	2	0.005063	238	0.6025
Icapui	1	0.002532	239	0.6051
Ico	2	0.005063	241	0.6101

level	freq	perc	cumfreq	cumperc
Iguatu	5	0.01266	246	0.6228
Independencia	1	0.002532	247	0.6253
Ipu	2	0.005063	249	0.6304
Iracema	2	0.005063	251	0.6354
Iraucuba	1	0.002532	252	0.638
Itaigaba	1	0.002532	253	0.6405
Itapaje	6	0.01519	259	0.6557
Itapipoca	11	0.02785	270	0.6835
Itarema	4	0.01013	274	0.6937
Itatira	2	0.005063	276	0.6987
Jaguaribe	2	0.005063	278	0.7038
Jaguaruana	1	0.002532	279	0.7063
Juazeiro do Norte	1	0.002532	280	0.7089
Juazeiro Do Norte	4	0.01013	284	0.719
Jucas	3	0.007595	287	0.7266
Lavras Da Mangabeira	3	0.007595	290	0.7342
Limoeiro	3	0.007595	293	0.7418
Madalena	2	0.005063	295	0.7468
Maracanau	7	0.01772	302	0.7646
Maranguape	4	0.01013	306	0.7747
Martinopole	1	0.002532	307	0.7772
Massape	2	0.005063	309	0.7823
Meruoca	1	0.002532	310	0.7848
Miraima	1	0.002532	311	0.7873
Missao Velha	1	0.002532	312	0.7899
Mombaca	1	0.002532	313	0.7924
Morada Nova	1	0.002532	314	0.7949
Moraujo	2	0.005063	316	0.8
Morava Nova	2	0.005063	318	0.8051
Mossoro	1	0.002532	319	0.8076
Nova Olinda	1	0.002532	320	0.8101
Novo Oriente	1	0.002532	321	0.8127
Ocara	2	0.005063	323	0.8177
Oros	2	0.005063	325	0.8228
Pacajus	4	0.01013	329	0.8329
Pacatuba	6	0.01519	335	0.8481
Pacoti	1	0.002532	336	0.8506
Pacuja	1	0.002532	337	0.8532
Palmacia	1	0.002532	338	0.8557
Paraipaba	2	0.005063	340	0.8608
Parambu	3	0.007595	343	0.8684
Pedra Branca	2	0.005063	345	0.8734
Penaforte	1	0.002532	346	0.8759
Pentecoste	2	0.005063	348	0.881
Pindoretama	3	0.007595	351	0.8886
Pires Ferreira	1	0.002532	352	0.8911
Quiterianopolis	2	0.005063	354	0.8962
Quixada	6	0.01519	360	0.9114
Quixeramobim	2	0.005063	362	0.9165
Reriutaba	2	0.005063	364	0.9215
Russas	2	0.005063	366	0.9266
Santa Quiteria	1	0.002532	367	0.9291

level	freq	perc	cumfreq	cumperc
Santana Do Acarau	2	0.005063	369	0.9342
Sao Benedito	4	0.01013	373	0.9443
Sao Luis Do Curu	1	0.002532	374	0.9468
Sobral	6	0.01519	380	0.962
Solonopole	1	0.002532	381	0.9646
Tabuleiro Do Norte	1	0.002532	382	0.9671
Tarauaca	1	0.002532	383	0.9696
Taua	1	0.002532	384	0.9722
Tiangua	4	0.01013	388	0.9823
Tairi	1	0.002532	389	0.9848
Tururu	1	0.002532	390	0.9873
Ubajara	3	0.007595	393	0.9949
Umirim	1	0.002532	394	0.9975
Uruburetama	1	0.002532	395	1

Dados e código para replicação estão disponíveis no repositório do GitHub do projeto