

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)
```

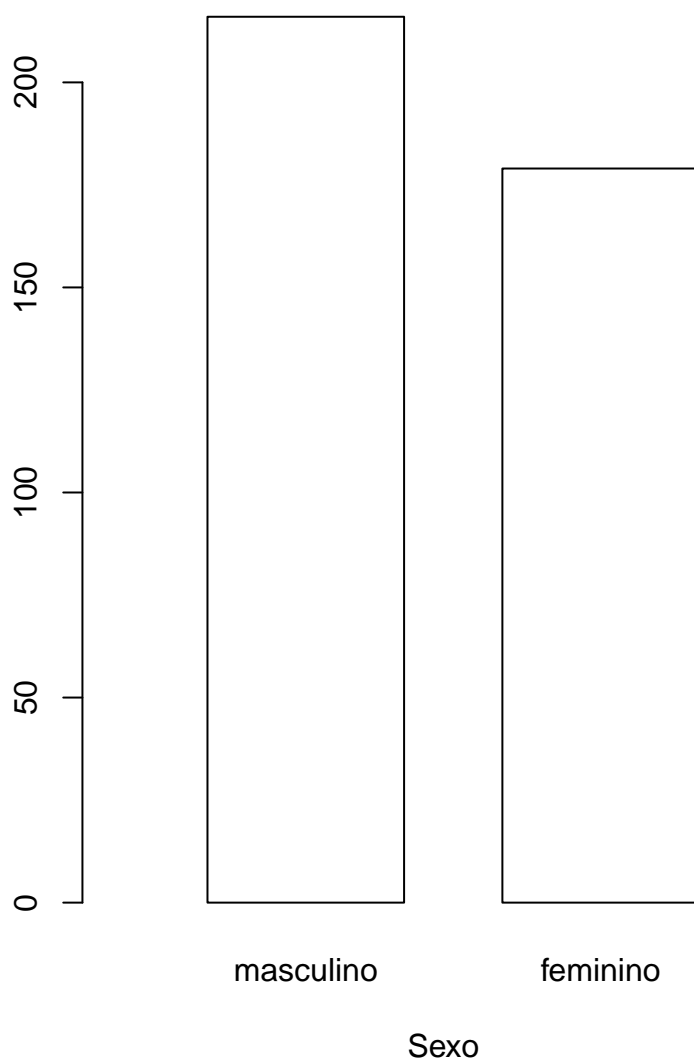


Figura 1: número de pacientes segundo o sexo.

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

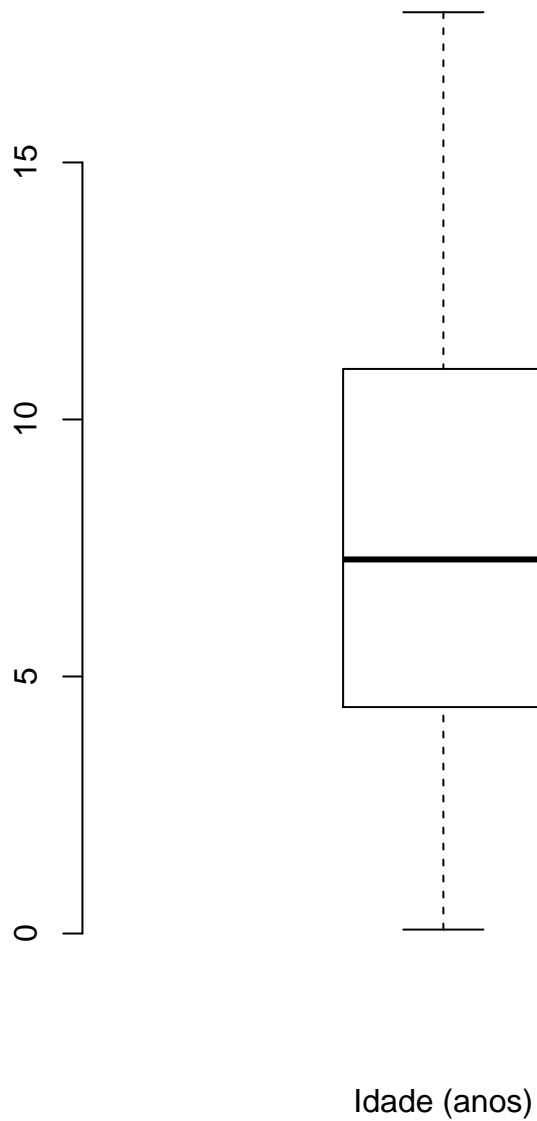


Figura 2: idade dos pacientes ao diagnóstico.

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

Table 1: Procedência dos pacientes

level	freq	perc	cumfreq	cumperc
Fortaleza	131	0.3316	131	0.3316
Caucaia	21	0.05316	152	0.3848
Itapipoca	11	0.02785	163	0.4127
Caninde	7	0.01772	170	0.4304
Maracanau	7	0.01772	177	0.4481
Itapaje	6	0.01519	183	0.4633
Pacatuba	6	0.01519	189	0.4785
Quixada	6	0.01519	195	0.4937
Sobral	6	0.01519	201	0.5089
Iguatu	5	0.01266	206	0.5215
Eusebio	4	0.01013	210	0.5316
Horizonte	4	0.01013	214	0.5418
Itarema	4	0.01013	218	0.5519
Juazeiro Do Norte	4	0.01013	222	0.562
Maranguape	4	0.01013	226	0.5722
Pacajus	4	0.01013	230	0.5823
Sao Benedito	4	0.01013	234	0.5924
Tiangua	4	0.01013	238	0.6025
Aquiraz	3	0.007595	241	0.6101
Aracoiaba	3	0.007595	244	0.6177
Barbalha	3	0.007595	247	0.6253
Barreira	3	0.007595	250	0.6329
Capistrano	3	0.007595	253	0.6405
Crato	3	0.007595	256	0.6481
Guaiuba	3	0.007595	259	0.6557
Jucas	3	0.007595	262	0.6633
Lavras Da Mangabeira	3	0.007595	265	0.6709
Limoeiro	3	0.007595	268	0.6785
Parambu	3	0.007595	271	0.6861
Pindoretama	3	0.007595	274	0.6937
Ubajara	3	0.007595	277	0.7013
Acopiara	2	0.005063	279	0.7063
Amontada	2	0.005063	281	0.7114
Apuiaries	2	0.005063	283	0.7165
Aracati	2	0.005063	285	0.7215
Barro	2	0.005063	287	0.7266
Baturite	2	0.005063	289	0.7316
Beberibe	2	0.005063	291	0.7367
Bela Cruz	2	0.005063	293	0.7418
Camocim	2	0.005063	295	0.7468
Carire	2	0.005063	297	0.7519
Cedro	2	0.005063	299	0.757
Forquilha	2	0.005063	301	0.762
Guaraciaba Do Norte	2	0.005063	303	0.7671
Ibicuitinga	2	0.005063	305	0.7722
Ico	2	0.005063	307	0.7772
Ipu	2	0.005063	309	0.7823
Iracema	2	0.005063	311	0.7873
Itatira	2	0.005063	313	0.7924
Jaguaribe	2	0.005063	315	0.7975

level	freq	perc	cumfreq	cumperc
Madalena	2	0.005063	317	0.8025
Massape	2	0.005063	319	0.8076
Moraujo	2	0.005063	321	0.8127
Morava Nova	2	0.005063	323	0.8177
Ocara	2	0.005063	325	0.8228
Oros	2	0.005063	327	0.8278
Paraipaba	2	0.005063	329	0.8329
Pedra Branca	2	0.005063	331	0.838
Pentecoste	2	0.005063	333	0.843
Quiterianopolis	2	0.005063	335	0.8481
Quixeramobim	2	0.005063	337	0.8532
Reriutaba	2	0.005063	339	0.8582
Russas	2	0.005063	341	0.8633
Santana Do Acarau	2	0.005063	343	0.8684
Abaiara	1	0.002532	344	0.8709
Acarape	1	0.002532	345	0.8734
Acarau	1	0.002532	346	0.8759
Aiuaba	1	0.002532	347	0.8785
Alcantaras	1	0.002532	348	0.881
Alto Santo	1	0.002532	349	0.8835
Ararende	1	0.002532	350	0.8861
Aurora	1	0.002532	351	0.8886
Boa Viagem	1	0.002532	352	0.8911
Boa Vista	1	0.002532	353	0.8937
Cascavel	1	0.002532	354	0.8962
Catunda	1	0.002532	355	0.8987
Chaval	1	0.002532	356	0.9013
Crateus	1	0.002532	357	0.9038
Cruz	1	0.002532	358	0.9063
Farias Brito	1	0.002532	359	0.9089
Frecheirinha	1	0.002532	360	0.9114
Graca	1	0.002532	361	0.9139
Granja	1	0.002532	362	0.9165
Granjeiro	1	0.002532	363	0.919
Hidrolandia	1	0.002532	364	0.9215
Ibaretama	1	0.002532	365	0.9241
Icapui	1	0.002532	366	0.9266
Independencia	1	0.002532	367	0.9291
Iraucuba	1	0.002532	368	0.9316
Itaigaba	1	0.002532	369	0.9342
Jaguaruana	1	0.002532	370	0.9367
Juazeiro do Norte	1	0.002532	371	0.9392
Martinopole	1	0.002532	372	0.9418
Meruoca	1	0.002532	373	0.9443
Miraima	1	0.002532	374	0.9468
Missao Velha	1	0.002532	375	0.9494
Mombaca	1	0.002532	376	0.9519
Morada Nova	1	0.002532	377	0.9544
Mossoro	1	0.002532	378	0.957
Nova Olinda	1	0.002532	379	0.9595
Novo Oriente	1	0.002532	380	0.962
Pacoti	1	0.002532	381	0.9646

level	freq	perc	cumfreq	cumperc
Pacuja	1	0.002532	382	0.9671
Palmacia	1	0.002532	383	0.9696
Penaforte	1	0.002532	384	0.9722
Pires Ferreira	1	0.002532	385	0.9747
Santa Quiteria	1	0.002532	386	0.9772
Sao Luis Do Curu	1	0.002532	387	0.9797
Solonopole	1	0.002532	388	0.9823
Tabuleiro Do Norte	1	0.002532	389	0.9848
Tarauaca	1	0.002532	390	0.9873
Taua	1	0.002532	391	0.9899
Tairi	1	0.002532	392	0.9924
Tururu	1	0.002532	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