

# 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 14, 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(sex)),names.arg=c("masculino","feminino"),
xlab="Sexo",width=0.5,xlim=c(0,1.7),space=0.5,col=0)
```

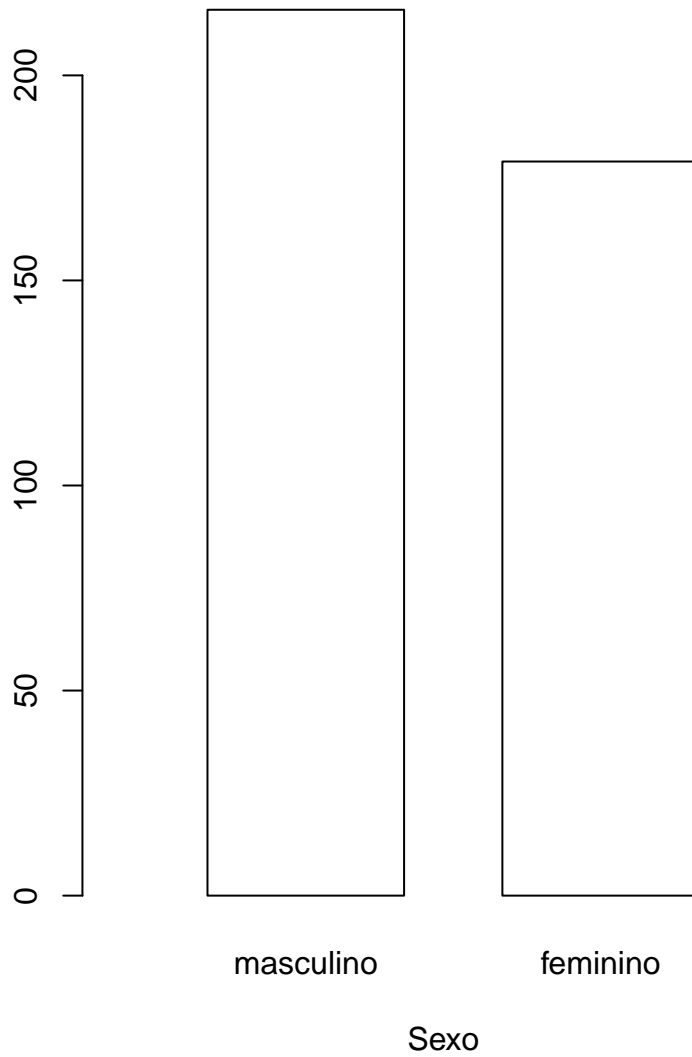


Figura 1: número de pacientes segundo o sexo.

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

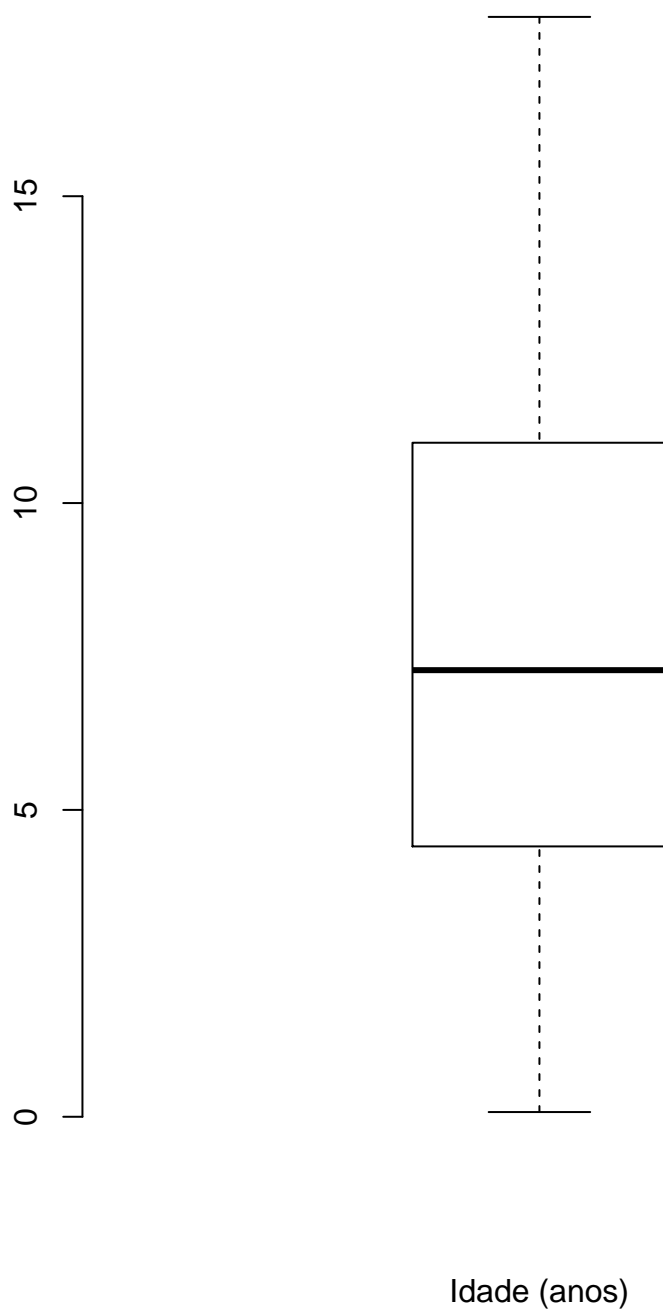


Figura 2: idade dos pacientes ao diagnóstico.

```
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
Juazeiro do Norte	5	0.01266	211	0.5342
Eusebio	4	0.01013	215	0.5443
Horizonte	4	0.01013	219	0.5544
Itarema	4	0.01013	223	0.5646
Maranguape	4	0.01013	227	0.5747
Pacajus	4	0.01013	231	0.5848
Sao Benedito	4	0.01013	235	0.5949
Tiangua	4	0.01013	239	0.6051
Aquiraz	3	0.007595	242	0.6127
Aracoiaba	3	0.007595	245	0.6203
Barbalha	3	0.007595	248	0.6278
Barreira	3	0.007595	251	0.6354
Capistrano	3	0.007595	254	0.643
Crato	3	0.007595	257	0.6506
Guaiuba	3	0.007595	260	0.6582
Jucas	3	0.007595	263	0.6658
Lavras Da Mangabeira	3	0.007595	266	0.6734
Limoeiro	3	0.007595	269	0.681
Morada Nova	3	0.007595	272	0.6886
Parambu	3	0.007595	275	0.6962
Pindoretama	3	0.007595	278	0.7038
Ubajara	3	0.007595	281	0.7114
Acopiara	2	0.005063	283	0.7165
Amontada	2	0.005063	285	0.7215
Apuiaries	2	0.005063	287	0.7266
Aracati	2	0.005063	289	0.7316
Barro	2	0.005063	291	0.7367
Baturite	2	0.005063	293	0.7418
Beberibe	2	0.005063	295	0.7468
Bela Cruz	2	0.005063	297	0.7519
Camocim	2	0.005063	299	0.757
Carire	2	0.005063	301	0.762
Cedro	2	0.005063	303	0.7671
Forquilha	2	0.005063	305	0.7722
Guaraciaba Do Norte	2	0.005063	307	0.7772

level	freq	perc	cumfreq	cumperc
Ibicuitinga	2	0.005063	309	0.7823
Ico	2	0.005063	311	0.7873
Ipu	2	0.005063	313	0.7924
Iracema	2	0.005063	315	0.7975
Itatira	2	0.005063	317	0.8025
Jaguaribe	2	0.005063	319	0.8076
Madalena	2	0.005063	321	0.8127
Massape	2	0.005063	323	0.8177
Moraujo	2	0.005063	325	0.8228
Ocara	2	0.005063	327	0.8278
Oros	2	0.005063	329	0.8329
Paraipaba	2	0.005063	331	0.838
Pedra Branca	2	0.005063	333	0.843
Pentecoste	2	0.005063	335	0.8481
Quiterianopolis	2	0.005063	337	0.8532
Quixeramobim	2	0.005063	339	0.8582
Reriutaba	2	0.005063	341	0.8633
Russas	2	0.005063	343	0.8684
Santana Do Acarau	2	0.005063	345	0.8734
Abaiara	1	0.002532	346	0.8759
Acarape	1	0.002532	347	0.8785
Acarau	1	0.002532	348	0.881
Aiuaba	1	0.002532	349	0.8835
Alcantaras	1	0.002532	350	0.8861
Alto Santo	1	0.002532	351	0.8886
Ararende	1	0.002532	352	0.8911
Aurora	1	0.002532	353	0.8937
Boa Viagem	1	0.002532	354	0.8962
Boa Vista	1	0.002532	355	0.8987
Cascavel	1	0.002532	356	0.9013
Catunda	1	0.002532	357	0.9038
Chaval	1	0.002532	358	0.9063
Crateus	1	0.002532	359	0.9089
Cruz	1	0.002532	360	0.9114
Farias Brito	1	0.002532	361	0.9139
Frecheirinha	1	0.002532	362	0.9165
Graca	1	0.002532	363	0.919
Granja	1	0.002532	364	0.9215
Granjeiro	1	0.002532	365	0.9241
Hidrolandia	1	0.002532	366	0.9266
Ibaretama	1	0.002532	367	0.9291
Icapui	1	0.002532	368	0.9316
Independencia	1	0.002532	369	0.9342
Iraucuba	1	0.002532	370	0.9367
Itaigaba	1	0.002532	371	0.9392
Jaguaruana	1	0.002532	372	0.9418
Martinopole	1	0.002532	373	0.9443
Meruoca	1	0.002532	374	0.9468
Miraima	1	0.002532	375	0.9494
Missao Velha	1	0.002532	376	0.9519
Mombaca	1	0.002532	377	0.9544
Mossoro	1	0.002532	378	0.957

level	freq	perc	cumfreq	cumperc
Nova Olinda	1	0.002532	379	0.9595
Novo Oriente	1	0.002532	380	0.962
Pacoti	1	0.002532	381	0.9646
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

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

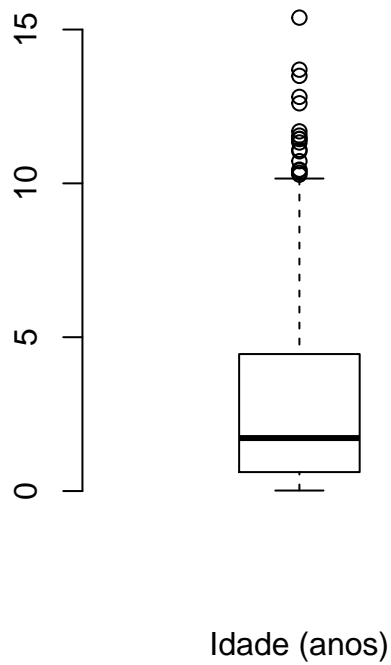


Figura 3: tempo de seguimento.

```
require(DescTools)
panderOptions('table.split.table', Inf)
set.caption("Histologia dos tumores")
pander(Freq(histo,ord='desc')[,-4], style = 'rmarkdown')
```

Table 2: Histologia dos tumores

level	freq	perc	cumperc
Unknown	105	0.2658	0.2658
Meduloblastoma classico	54	0.1367	0.4025
Astrocitoma pilocítico	30	0.07595	0.4785
Astrocitoma difuso/fibrilar	26	0.06582	0.5443
Ependimoma	26	0.06582	0.6101
Ependimoma anaplasico	17	0.04304	0.6532
Glioblastoma	16	0.04051	0.6937
Meduloblastoma desmoplasico/nodular	15	0.03797	0.7316
Craniofaringioma	11	0.02785	0.7595
Astrocitoma pilomixóide	10	0.02532	0.7848
Meduloblastoma	10	0.02532	0.8101
Tumor de células germinativas SOE	8	0.02025	0.8304
Astrocitoma anaplasico	6	0.01519	0.8456
Malformação cavernosa (cavernoma)	4	0.01013	0.8557
Oligodendroglioma	4	0.01013	0.8658
Tumor neuroectodérmico primitivo - PNET	4	0.01013	0.8759
Germinoma	3	0.007595	0.8835
Melanoma	3	0.007595	0.8911
Meningioma	3	0.007595	0.8987
Meningioma transicional	3	0.007595	0.9063
Papiloma de plexo coroide	3	0.007595	0.9139
Carcinoma de plexo coroide	2	0.005063	0.919
Ganglioglioma anaplasico	2	0.005063	0.9241
Ganglioglioma atípico	2	0.005063	0.9291
Ganglioglioma Infantil Desmoplasico	2	0.005063	0.9342
Glioma SOE	2	0.005063	0.9392
Meduloepitelioma	2	0.005063	0.9443
Meningioma atípico	2	0.005063	0.9494
Meningioma meningotelial	2	0.005063	0.9544
Pineoblastoma	2	0.005063	0.9595
Schwannoma	2	0.005063	0.9646
Tumor teratóide-rabdoide atípico	2	0.005063	0.9696
Adenoma de hipófise	1	0.002532	0.9722
Astrocitoma	1	0.002532	0.9747
Astrocitoma subependimário de células gigantes	1	0.002532	0.9772
Carcinoma embrionário	1	0.002532	0.9797
Coriocarcinoma	1	0.002532	0.9823
Disgerminoma	1	0.002532	0.9848
Fibrossarcoma infantil	1	0.002532	0.9873
Gangliocitoma	1	0.002532	0.9899
Meduloblastoma de grandes células/anaplasico	1	0.002532	0.9924
Oligodendroglioma anaplasico	1	0.002532	0.9949
Sarcoma granulocítico	1	0.002532	0.9975
Xantastrocitoma pleomorfo	1	0.002532	1

```
require(reshape)
```

```
## Loading required package: reshape
```

```
snc$histo.group=combine_factor(histo,c(1,2,3,2,2,2,2,4,5,5,1,5,6,7,4,8,8,8,8,5,3,8,1,9,9,9,9,10,4,1,1,1),  
levels(snc$histo.group)=c("Outros tumores benignos","Glioma de baixo grau","Glioma de alto grau","Outros  
barplot(sort(summary(snc$histo.group)*100/sum(summary(snc$histo.group))),decreasing=T,col="white")
```

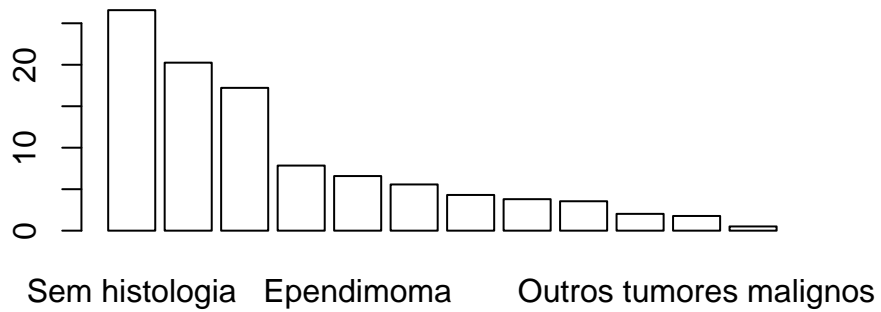


Figura 4: grupos segundo a histologia (%).

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