

O Problema do Aniversário, o pacote IPSUR e seu *plugin* para o R Commander: uma possibilidade para sala de aula

Felipe Rafael Ribeiro Melo

(DMQ - UNIRIO)

VII Seminário Internacional de Estatística com R
25 de maio de 2023

1. Apresentando o problema e suposições

O Problema do Aniversário

Em uma sala com n pessoas, qual a probabilidade de termos pelo menos um par de pessoas fazendo aniversário no mesmo dia?

Suposições:

- Desconsiderar 29 de fevereiro.
- Assumir a mesma probabilidade de aniversário para cada um dos 365 dias, ou seja: $P(\text{"01/jan"}) = P(\text{"02/jan"}) = \dots = P(\text{"30/dez"}) = P(\text{"31/dez"}) = 1/365$.
- Não ter gêmeos na sala.

2. Solução do Problema do Aniversário

Vamos denotar o evento de interesse pela letra B , ou seja,

$B = \{\text{pelo menos um par de pessoas aniversariar no mesmo dia}\}.$

A solução deste problema não é difícil se “atacarmos” pelo complementar do evento B , ou seja,

$B^C = \{\text{as } n \text{ pessoas fazem aniversários em dias diferentes}\}.$

Por meio da propriedade que relaciona a probabilidade de um evento com a probabilidade do seu evento complementar, podemos obter a probabilidade de interesse, $P(B)$, por meio da expressão: $P(B) = 1 - P(B^C).$

- Pelo Princípio de Dirichlet (ou Princípio das casas dos pombos), $P(B^C) = 0$ quando $n > 365$. Portanto, voltemos para o cálculo de $P(B^C)$ quando $n \leq 365$.
- Pela suposição de mesma probabilidade de aniversário para cada um dos 365 dias, cada n -upla de aniversários tem a mesma probabilidade de ocorrência: $1/365^n$.
- Dessa forma, podemos usar o conceito clássico da Probabilidade para calcular $P(B^C)$:

$$P(B^C) = \frac{\#B^C}{\#\Omega}.$$

As cardinalidades acima são dadas por $\#\Omega = 365^n$ e $\#B^C = A_{365}^n = C_{365}^n \times n!$

- Portanto, para $n \in \{1, 2, \dots, 365\}$,

$$\begin{aligned} P(B^C) &= \frac{\#B^C}{\#\Omega} = \frac{n! \times 365! / [n! \times (365 - n)!]}{365^n} \\ &= \frac{365!}{365^n \times (365 - n)!}. \end{aligned}$$

- Logo, em uma sala com n pessoas ($n \leq 365$), a probabilidade de termos pelo menos um par de pessoas fazendo aniversário no mesmo dia, sob os pressupostos explicitados, é

$$P(B) = 1 - \frac{365!}{365^n \times (365 - n)!}.$$

- Note que, se $n = 1$, então $P(B) = 1 - 1 = 0$.

- Para $n > 365$, temos $P(B^C) = 0$, conforme já mencionado, o que implica $P(B) = 1$.
- Portanto, de maneira mais abrangente, a probabilidade abordada no Problema do Aniversário é dada por

$$P(B) = \begin{cases} 1 - \frac{365!}{365^n \times (365 - n)!} & , \text{ se } n \leq 365, \\ 1 & , \text{ se } n > 365, \end{cases}$$

ou ainda, utilizando notação de número de arranjos simples,

$$P(B) = \begin{cases} 1 - \frac{A_{365}^n}{365^n} & , \text{ se } n \leq 365, \\ 1 & , \text{ se } n > 365. \end{cases}$$

3. O Problema do Aniversário no R

- Há (pelo menos) dois pacotes no R que trazem funções relacionados ao Problema do Aniversário:
 - ➊ IPSUR
 - ➋ RcmdrPlugin.IPSUR
- A expressão IPSUR provém de *I*n*t*ro*d*uction to *P*ro*b*ability and *S*t*a*tistics *U*sing *R*.
- Infelizmente, estes dois pacotes não estão mais disponíveis no repositório CRAN. Para contornar isto, instalaremos tais pacotes por meio da função `install_github` do pacote `devtools`. Para usuários do Windows, em particular, é necessária a instalação prévia do **Rtools**, conforme explicitado no Passo 2 em <https://www.r-project.org/nosvn/pandoc/devtools.html>.

3.1. O pacote IPSUR

O pacote IPSUR não está mais disponível no repositório CRAN desde 30/05/2019, impedindo assim sua instalação via método convencional. Uma alternativa para a sua instalação segue no passo a passo disponível em <https://ipsur.org/>.

```
install.packages(c("actuar", "aplpack", "binom", "boot", "coin", "diagram",  
                  "distrEx", "e1071", "emdbook", "ggplot2", "HH", "Hmisc",  
                  "lmtest", "mvtnorm", "qcc", "reshape", "RcmdrMisc",  
                  "scatterplot3d", "TeachingDemos", "vcd"))
```

```
install.packages("combinat") # the prob package depends on combinat  
install.packages("devtools") # if you don't have it already  
devtools::install_github("gjkerns/prob")
```

```
devtools::install_github("gjkerns/IPSUR")
```


- Uma vez instalado o pacote IPSUR, carregue-o por meio dos comandos `library(IPSUR)` ou `require(IPSUR)`.
- Feito isto, a execução do comando

`read(IPSUR)`

abre o arquivo *IPSUR.pdf*, que é a edição mais recente (tomando por base o momento de instalação do pacote) do livro ***Introduction to Probability and Statistics Using R***, do mesmo autor dos pacotes IPSUR e RcmdrPlugin.IPSUR: Gary Jay Kerns.

- No Windows, o arquivo *IPSUR.pdf* está diretório `AppData\Local\R\win-library\4.3\IPSUR\doc` da sua pasta de usuário. Também estão alocados nesta pasta os arquivos *IPSUR.R* e *IPSUR.Rnw*.

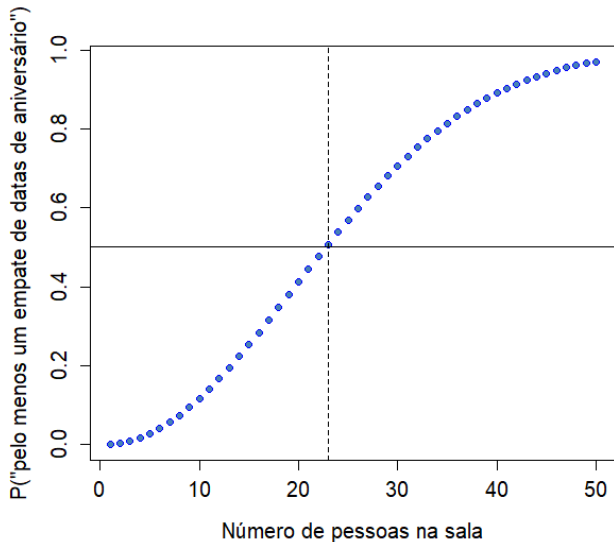
Além dos arquivos mencionados acima e da função `read`, o pacote IPSUR traz as funções `pbirthday.ipsur` e `qbirthday.ipsur`, detalhadas a seguir com seus valores *default*.

- `pbirthday.ipsur(n, classes=365, coincident=2)`: calcula a probabilidade de, numa sala com n pessoas, ao menos 2 fazerem aniversário no mesmo dia, considerando 365 possibilidades de dias.
- `qbirthday.ipsur(prob=0.5, classes=365, coincident=2)`: retorna o menor valor n cuja probabilidade de, numa sala com n pessoas, ao menos 2 fazerem aniversário no mesmo dia, considerando 365 possibilidades de dias, seja maior ou igual a `prob`.

```
> pbirthday.ipsur(10)
[1] 0.1169482
> pbirthday.ipsur(23)
[1] 0.5072972
> pbirthday.ipsur(30)
[1] 0.7063162
> pbirthday.ipsur(40)
[1] 0.8912318
> pbirthday.ipsur(50)
[1] 0.9703736
> pbirthday.ipsur(100)
[1] 0.9999997
> pbirthday.ipsur(120)
[1] 1
> options(digits=22)
> pbirthday.ipsur(120)
[1] 0.9999999997560852227352
> |
```

```
> qbirthday.ipsur(0.25)
[1] 15
> qbirthday.ipsur(0.50)
[1] 23
> qbirthday.ipsur(0.75)
[1] 32
> qbirthday.ipsur(0.90)
[1] 41
> qbirthday.ipsur(0.95)
[1] 47
> qbirthday.ipsur(0.99)
[1] 57
> qbirthday.ipsur(1)
[1] 366
> |
```

A 4ª edição do livro IPSUR aborda o Problema do Aniversário no seu Exemplo 4.27, que traz o *script* deste gráfico.



3.2. O pacote RcmdrPlugin.IPSUR

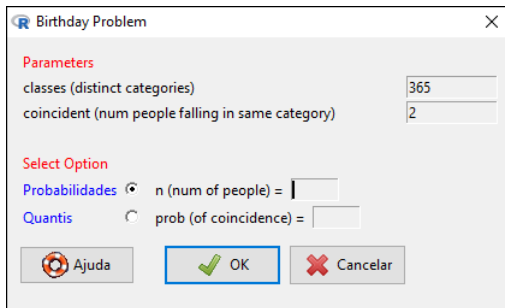
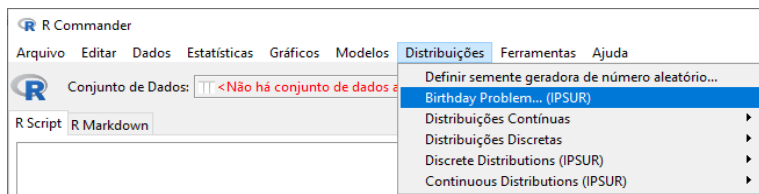
- Para os usuários da interface *R Commander*, que provém do carregamento do pacote Rcmdr, há a alternativa de se trabalhar com o pacote RcmdrPlugin.IPSUR, que é um *plugin* do pacote IPSUR dentro desta interface.
- Trabalhar no *R Commander* com turmas que tenham pouca ou nenhuma experiência em programação deixa os discentes mais à vontade, por se tratar de uma interface amigável, cujas funcionalidades provém de menus.
- *Plugins* do *R Commander* também são pacotes, e devem ser instalados via repositório CRAN (método mais tradicional) ou por formas alternativas, como o uso da função `install_github` do pacote devtools, quando da impossibilidade do uso do método mais tradicional.

- O pacote RcmdrPlugin.IPSUR não está mais disponível no repositório CRAN desde 10/08/2022. Entretanto, podemos instalá-lo por meio da linha de comando

```
devtools::install_github("gjkerns/RcmdrPlugin.IPSUR")
```

- Ao carregar o pacote RcmdrPlugin.IPSUR diretamente no *R Console*, a janela *R Commander* será aberta com seus menus incluindo funcionalidades provenientes deste *plugin*, indicadas pelo rótulo “(IPSUR)”.
- Observação: Caso a janela *R Commander* esteja aberta antes do carregamento do pacote RcmdrPlugin.IPSUR, este pode ser carregado no menu *Ferramentas > Carregar plug-in(s) do Rcmdr*.

Menu *Distribuições* > *Birthday Problem* utiliza *pbirthday.ipsur* ou *qbirthday.ipsur*, a depender da opção selecionada.



- Além deste menu voltado ao Problema do Aniversário, o pacote RcmdrPlugin.IPSUR adiciona ao *R Commander* outras funcionalidades nos menus *Distribuições*, *Gráficos* e *Estatísticas*, além de três conjuntos de dados que podem ser carregados no menu *Dados > Conjunto de dados em pacotes > Ler dados de pacote “attachado”*.
- Mais detalhes podem ser vistos no manual do pacote, disponível em <https://mran.microsoft.com/snapshot/2015-01-06/web/packages/RcmdrPlugin.IPSUR/RcmdrPlugin.IPSUR.pdf>.

4. Curiosidades

- O Problema do Aniversário é geralmente atribuído a Harold Davenport, por volta de 1927. Embora ele não o tenha publicado na época, Davenport não reivindicou ser seu descobridor “porque ele não podia acreditar que não havia sido estabelecido anteriormente”. A primeira publicação de uma versão do Problema do Aniversário foi feita por Richard von Mises, em 1939.
- Frequentemente, o termo Paradoxo do Aniversário (*Birthday Paradox*) é associado ao Problema do Aniversário, em menção ao fato de que bastam 23 pessoas escolhidas aleatoriamente para que a probabilidade de que pelo menos duas delas façam aniversário no mesmo dia exceda 50%.

Referências Bibliográficas



Diaconis, P. & Mosteller, F. (1989). *Methods for studying coincidences*. JASA 84:853-861. Disponível em: https://www.stat.berkeley.edu/~aldous/157/Papers/diaconis_mosteller.pdf.



Kerns, G.J. (2022). **Introduction to Probability and Statistics Using R**. Fourth Edition. GNU Free Documentation License. Disponível em <https://github.com/gjkerns/IPSUR/blob/master/inst/doc/IPSUR.pdf>.



Kerns, G.J. (2023). *IPSUR: Introduction to Probability and Statistics Using R*. R package version 4.0. <https://ipsur.org>.



Kerns, G.J. (2019). *RcmdrPlugin.IPSUR: An IPSUR Plugin for the R Commander*. R package version 0.2-1.1. Manual disponível em <https://mran.microsoft.com/snapshot/2015-01-06/web/packages/RcmdrPlugin.IPSUR/RcmdrPlugin.IPSUR.pdf>.



R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.



Ross, S. (2010). **A First Course In Probability**. 8th Edition. Prentice Hall.



von Mises, R. (1964). *Selected Papers of Richard von Mises*, vol. 2, pp. 313–334. Providence. American Mathematical Society. Disponível em: http://alexander.shen.free.fr/vonMises_64_SelectedPapersVol2OCR.pdf.

*MUITO
OBRIGADO !!!*