

# pacote Rcoisas: funções para aulas de epidemiologia e bioestatística

Fúlvio B. Nedel

2023-11-26

## Dados gerais

### Resumo/Objetivos

Pacote na linguagem R com funções para facilitar a análise básica de dados com ‘inputs’ e ‘outputs’ em português: funções para leitura, agregação ou formatação de dados do DATASUS; desenho de pirâmides populacionais e gráficos de indicadores de saúde como a Curva de Nelson de Moraes; análise estatística básica, como a descrição “completa” de uma variável numérica, o cálculo e gráfico da probabilidade de situar-se em uma área sob a curva normal, ou a toma de amostras aleatórias de um banco de dados para a organização de exercícios em aula.

### Palavras-chave

Ensino de bioestatística; ensino de epidemiologia; análise de dados em saúde

### Público alvo

Estudantes de graduação e pós-graduação em saúde; professores de epidemiologia e bioestatística em língua portuguesa

## Descrição

### Contexto

O “R” é uma linguagem computacional de manejo e análise estatística de dados de código aberto surgida nos anos 1990 e que tornou-se um dos *softwares* mais utilizados no ensino de epidemiologia e bioestatística atualmente. Entre suas vantagens estão: 1) acesso público, facilitando ao estudante ter em seu computador o mesmo programa do laboratório de informática; 2) grande quantidade de *pacotes* voltados a diferentes necessidades na análise de dados; 3) fácil produção de relatórios automatizados da análise realizada, com qualidade de publicação; 4) fácil criação de pacotes voltados a necessidades específicas do usuário, com funções novas ou modificadas de outros pacotes;<sup>1</sup> 5) está integrado ao TabWin, software de tabulação de dados do Ministério da Saúde Brasileiro.<sup>2</sup> Entre as desvantagens estão uma curva de aprendizagem de início lento e a produção de outputs em inglês, o que exige trabalho adicional para a produção de relatórios automatizados, tabelas e gráficos com números no formato latino.

O pacote Rcoisas é o agrupamento de uma série de funções desenvolvidas em R ao longo dos anos para aulas de epidemiologia e bioestatística, entre outras análises e apresentações de resultados.<sup>3</sup> Pode ser instalado pela plataforma GitHub, mas ainda não tem uma *release*. Exemplos de uso podem ser vistos no “Readme” (<https://github.com/fulvionedel/Rcoisas#readme>).

## Justificativa

O ensino de estatística nos cursos de graduação em saúde deve ir além do cálculo de medidas de tendência central e posição (percentis) realizados à mão ou em planilhas de cálculo. O Epi-Info, que foi um clássico nesse papel (e na análise de dados em estudos de campo) nas versões para o sistema operacional DOS, não manteve sua aplicabilidade (e popularidade) ao passar para o sistema Windows. Entre os diversos softwares de uso livre voltados ao manejo e análise estatística de dados em saúde, o R é talvez o mais completo na atualidade, competindo com linguagens como Python, por sua vez mais difícil e menos usada no campo da saúde. Entretanto, comparado ao Epi-Info, escrever em R exige conhecer e atentar a uma série de detalhes que tornam mais lenta sua aprendizagem, o que pode ser uma limitação importante em cursos rápidos ou disciplinas de graduação dos cursos da Saúde. O pacote procura facilitar a entrada de comandos (“*inputs*”) para a análise básica de dados por usuários de língua portuguesa e a leitura de “*outputs*”, além de apresentar amostras de dados de mortalidade e internações brasileiros, úteis para a construção de indicadores de saúde em exemplos didáticos.

## Objetivos

### Geral

Facilitar o uso do R em análises estatísticas básicas com *inputs* intuitivos ao usuário brasileiro e *outputs* em português, permitindo assim o uso da ferramenta sem grande desvio de foco nos conceitos da estatística e epidemiologia e a publicação de resultados automatizados em formato latino.

### Específicos

- Resolver definições ainda pendentes em algumas funções;
- Melhorar a documentação das funções existentes;
- Lançar a 1ª *release* do pacote.

## Metodologia

Em algumas funções simplesmente são realizadas traduções dos argumentos e modificações no formato ou lay-out do *output* de funções existentes em outros pacotes do R. Em outras são criados os comandos para a análise ou a criação de gráficos – como nas funções `plotZ`, para a probabilidade de uma área sob a curva Normal,<sup>4</sup> `ggplot_pir` para a criação de pirâmides populacionais ou `fxetarNM` para a agregação da idade nas faixas de Nelson de Moraes e desenho da curva.

São disponibilizadas amostras de bancos de dados públicos reais, como as bases do Sistema de Informação de Mortalidade (SIM) e Sistema de Informações Hospitalares do SUS (SIH-SUS),<sup>5</sup> o que facilita a visualização e uso desses dados, sem grande uso de memória do computador.

O trabalho é realizado no ambiente de desenvolvimento integrado (IDE) RStudio, com auxílio dos pacotes devtools e roxygen2. O controle de versão é feito pela plataforma GitHub. Todo o software necessário é open-source e está instalado em meus computadores, em ambiente Windows 10® e Linux (Ubuntu22.04). Não é necessária a alocação de espaço físico ou equipamento adicional na Universidade.

## Resultados esperados

Disponibilização à comunidade acadêmica de um pacote para o uso em R voltado principalmente ao ensino de bioestatística e epidemiologia com R no Brasil. O pacote deve ter três utilidades principais:

- apresentação de números em formato latino, para publicações em português (e outras línguas latinas);
- funções voltadas à construção de indicadores de saúde e descrição univariada de variáveis numéricas e categóricas;
- amostras de bancos de dados dos principais Sistemas de Informação em Saúde do Brasil.

## Referências

1. Nedel FB. Pacote csapAIH: a Lista Brasileira de Internações por Condições Sensíveis à Atenção Primária no programa R. *Epidemiologia e Serviços de Saúde* 2019; 28: e2019084.
2. Brasil. and Ministério da Saúde and DATASUS. Tab para Windows - TabWin. , <http://datasus.saude.gov.br/apresentacao> (2010).
3. Nedel FB. *Rcoisas: Minhas facilidades*, <https://github.com/fulvionedel/Rcoisas> (2023).
4. Martín M, Horna O, Nedel F, et al. *Fundamentos de estadística en ciencias de la salud*. Cerdanyola del Vallès: Servei de Publicacions, Universitat Autònoma de Barcelona.
5. DATASUS. SIHSUS – Arquivos públicos, <ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/>.