



OBTENÇÃO DE DADOS NÃO AGREGADOS NO DATASUS: Análise das internações por asma no estado do Paraná de 1992 a 2017

Arthur Cesar de Moura Rocha¹, Paula Mitiko Heller¹ e Eniuce Menezes de Souza²

¹ Curso de Graduação em Estatística, Universidade Estadual de Maringá, Maringá, Paraná, Brasil

² Departamento de Estatística, Universidade Estadual de Maringá, Maringá, Paraná, Brasil

RESUMO

A base de dados do DATASUS é bastante rica e importante para o suporte de inúmeras pesquisas na área de saúde. Em geral, tais dados são obtidos online através do aplicativo TabWin o qual é muito dependente do usuário, dispendioso e restrito na seleção das variáveis, além de retornar apenas dados agregados, o que impossibilita a utilização de diversas metodologias estatísticas. Assim, este trabalho disponibiliza um passo-a-passo para obtenção de dados de internações não agregados no SUS de modo que facilite a manipulação desses dados e auxilie pesquisas da área de saúde. Questões relacionadas às mudanças de códigos e padronização das informações nesta plataforma serão consideradas para obtenção de dados que possam ser efetivamente analisadas e comparadas em localidades distintas e também ao longo do tempo. Em especial, será exemplificado como filtrar informações de interesse, no caso serão selecionadas internações por asma no estado do Paraná de janeiro de 1992 a setembro de 2017. Será apresentada a análise descritiva de tais dados, incluindo a geração de séries temporais e visualização em plataforma interativa em formato html. **Palavras chave:** DATASUS; Séries temporais; Asma.

1. INTRODUÇÃO

O Departamento de Informática do Sistema Único de Saúde (DATASUS) é um órgão do governo que tem como competência “prover os órgãos do SUS de sistemas de informação e suporte de informática, necessários ao processo de planejamento, operação e controle” (DATASUS, 2017). Dessa forma, uma de suas missões é a manutenção da base de dados necessária para o funcionamento integrado dos órgãos do SUS (DATASUS, 2017). Para o cumprimento dessa missão é gerado um banco de dados rico em informações sobre diversas condições de saúde da população brasileira, o qual está disponível para acesso ao público por meio online. Por causa de sua riqueza de dados e por ser uma fonte de coleta de dados confiável, dado que se trata de um órgão governamental, os dados disponibilizados pelo DATASUS são de fundamental importância para as pesquisas da área de saúde, tanto diretamente quanto indiretamente.

Por outro lado a dificuldade de manipulação de bancos de dados com grande volume de variáveis é uma reclamação frequente entre pesquisadores de diversas áreas do conhecimento. Essa limitação é visível em pesquisas na área da saúde como quando dito que

“A consulta aos dados disponibilizada, diretamente, no endereço eletrônico do DATASUS, tem algumas limitações. A principal é que se encontra separada por períodos determinados, sem a possibilidade de agregar diferentes períodos para o estudo de séries temporais. Além disso, as tabelas não permitem estratificação para a análise de características adicionais, como o tipo de hospital (psiquiátrico ou geral), ou por outros grupos de idade, por exemplo” (CANDIAGO, 2007).

A partir dessa dificuldade, este projeto se propõe a buscar métodos computacionais e estatísticos que facilitem a manipulação desses dados a fim de auxiliar pesquisas da área de saúde. Para isso serão utilizados o software R e a base de dados do DATASUS.

Após realizada a proposta, com o intuito de verificar o cumprimento do objetivo, será realizada uma aplicação do método para um estudo descritivo sobre a asma tendo em vista que a asma é um problema de saúde pública que, conforme a Pesquisa Nacional de Saúde (PNS) do Ministério da Saúde e Instituto Brasileiro de Geografia e Estatística (IBGE), atinge 6,4 milhões de brasileiros acima de 18 anos (BRASIL, 2015).

A Organização Mundial de Saúde (OMS) estima que existam cerca de 300 milhões de casos de asma pelo mundo. No Brasil, a asma é responsável pela internação de mais de 100 mil pessoas no SUS e segundo o Boletim Epidemiológico n. 18, v. 47, do Ministério da Saúde, Brasil (2016), a asma foi causa de 38% das internações por doenças respiratórias crônicas (DRC) no Brasil de 2003 a 2013. Um estudo importante efetuado em várias grandes cidades do Brasil, o ISAAC (*International Study of Asthma and Allergy in Childhood*), apontou que entre 20% a 30% das crianças e adolescentes apresentam sintomas indicativos da doença (PORTARIA 054-R, 2009).

A asma é uma doença inflamatória pulmonar crônica definida pela hiperreatividade das vias aéreas inferiores e limitação do fluxo de ar (SBPT, 2002). Mesmo sendo de natureza crônica, esta patologia pode ser controlada, mas a falta de um controle apropriado pode ocasionar um custo elevado em gastos com internações, medicamentos e serviços emergenciais, além de poder privar o indivíduo com asma de suas ações rotineiras, tais como emprego e estudos (BARNES, JONSSON e KLIM, 1996).

2. METODOLOGIA

A metodologia para o desenvolvimento deste projeto será baseada na utilização de recursos computacionais, os quais envolvem uma extensão dos conceitos aprendidos durante o Curso de Graduação em Estatística. A metodologia está organizada como segue.

2.1. Download dos dados

Embora seja comum obter dados agregados diretamente no site do DATASUS via aplicativo TabWin, é possível baixar os dados não agregados com as características de cada usuário, identificado por um código, para garantir o anonimato. Esse *download* de dados é realizado através de *softwares* diretamente do diretório. Devido a Classificação Internacional de Doenças e Problemas Relacionados com a Saúde, CID (em inglês: *International Statistical Classification of Diseases and Related Health Problems - ICD*) ter sido alterada em 2007, passando da versão 9 para versão 10, os dados de internação são disponibilizados em dois diretórios, os quais, na presente data são: ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/199201_200712/Dados/ e ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/200801_/Dados/, respectivamente. Nestes diretórios são disponibilizados dados referentes a Autorização de Internação Hospitalar (AIH) no formato reduzido (RD), rejeitadas (RJ) e atos médicos (SP). Os arquivos de interesse são os reduzidos, os quais são iniciados por RD. As duas próximas letras no nome de cada arquivo referem-se ao estado da internação. No caso deste trabalho, são de interesse as internações ocorridas no estado do Paraná, portanto, os arquivos devem iniciar com RDPR.

Os arquivos baixados estão no formato dbc, o qual é um arquivo de banco de dados criado pelo editor CANdb ou CANdb++. Existem diferentes formas possíveis para ler este tipo de arquivo nos *softwares* estatísticos, no caso do presente trabalho, o interesse é de fazê-lo no software R. Para isso foram utilizadas duas formas diferentes. A primeira consiste em obter o arquivo diretamente do diretório através de um laço de repetição para o *download* do arquivo em formato dbc. Com o auxílio da biblioteca “foreign” e de um executável disponível pelo DATASUS é possível realizar a conversão para dbf, tipo de arquivo de banco de dados criado pelo sistema dBase, e então a leitura do arquivo. Este método é limitado pois em computadores com sistema operacional diferente do Windows é necessário adaptar o executável, porém uma vez convertido o arquivo, este pode ser lido em outros *softwares* como o SAS. O segundo método, que é mais recente, segue a mesma sequência para a obtenção dos dados, porém utiliza a biblioteca recém desenvolvida “read.dbc”, a qual permite a leitura do arquivo sem a necessidade de convertê-lo para dbf, facilitando o processo para usuários de outros sistemas operacionais que não o Windows.

2.2. Armazenamento dos dados

Dada a grande quantidade de variáveis observadas nos arquivos obtidos pelo DATASUS, é necessário encontrar meios de armazenar estes dados de maneira que não haja confusão pelo sistema no momento de concatenar as colunas dos dados. Dessa forma, optou-se por criar um laço de repetição de leitura já selecionando

as colunas de interesse, que são em um primeiro momento, referentes ao diagnóstico, idade, sexo e município da hospitalização. Assim, utilizou-se uma estrutura de dados do tipo data frame para reunir todos os dados, uma vez que a estrutura de um data frame permite o armazenamento de dados em que cada linha representa uma observação.

2.3. Filtragem/Seleção dos dados de interesse

Para implementação dos filtros, são utilizados conceitos de Linguagem de Consulta Estruturada (Structured Query Language - SQL), cujas características originais foram inspiradas na álgebra relacional (SAS User's Guide, 2012). Assim, utilizando a biblioteca "sqldf", foram selecionadas das colunas desejadas os dados referentes a asma e armazenadas em outra estrutura de data frame. Para isso foi realizada uma seleção através dos códigos J45, J46 e 493, referentes ao CID 10 e CID 9, respectivamente, dessas colunas a fim de armazenar somente as observações de interesse.

Embora a CID forneça códigos de modo a padronizar e facilitar a obtenção de dados, alguns cuidados devem ser tomados na geração de séries temporais devido a alteração de tal codificação em 2007, passando de CID 9 para CID10. Assim, são criadas macros para possibilitar a obtenção automática dos códigos a partir de palavras chave específicas, tanto no CID9 quanto no CID10. Para averiguação e validação das implementações realizadas, os dados obtidos serão comparados com dados baixados diretamente pelo TabWin.

2.4. Apresentação dos dados

As análises descritivas serão realizadas de modo a permitir a identificação de diferentes características dos dados de um modo geral. Também serão criadas séries temporais, nas quais serão identificados padrões tais como, tendência e sazonalidade. Além das informações quantitativas, serão gerados gráficos, tais como box-plots, gráficos de linhas e mapas interativos em formato html para facilitar a visualização e posterior disponibilização à comunidade usuária. Nesta etapa, poderá ser utilizada a integração Shiny e RStudio em linguagem R (R Core Team, 2013).

3. RESULTADOS E DISCUSSÕES

Além da apresentação do passo a passo para obtenção dos dados do DATASUS, será apresentada a análise descritiva das internações por asma no estado do Paraná de 01 de janeiro de 1992 a 31 de dezembro de 2015, incluindo a geração de séries temporais e visualização em plataforma interativa em formato html, como ilustrado na figura abaixo..

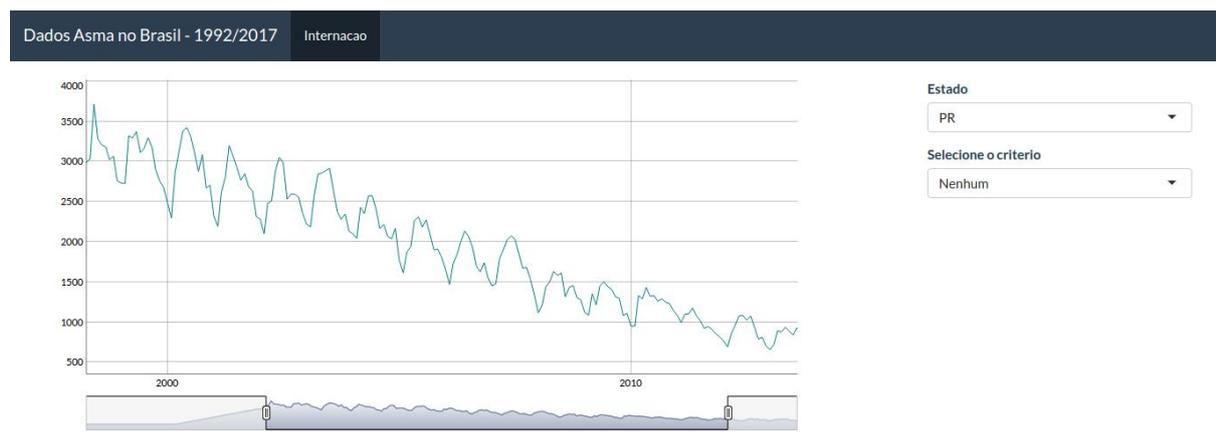


Figura 1: Plataforma interativa em html para visualização dos dados.

4. CONCLUSÃO

Para a realização de análises estatísticas é essencial que se tenha acesso aos dados não agregados, ou que se tenha liberdade de agregá-los de acordo com diferentes características de interesse, no caso do estudo de séries temporais. Além disso, o acesso direto a partir de softwares de ampla utilização em estatística, tais como, R e SAS, são facilitadores de um modo geral. Em continuidade a este trabalho, serão criadas macros em ambientes R de modo que a implementação seja mais automatizada e facilitada para o usuário.

5. AGRADECIMENTOS

Agradecemos a professora Eniuce Menezes de Souza por nos orientar e dar a oportunidade de desenvolvermos este projeto.

Referências

- [1] BARNES, P.J.; JONSSON, B.; KLIM, J.B. The costs of asthma. *European Respiratory Journal*, Copenhagen, v. 9, n. 4, p. 636-642, Abril 1996.
- [2] BRASIL. DATASUS. Histórico / Apresentação. Disponível em: <<http://datasus.saude.gov.br/datasus>>. Acesso em: 26 maio 2017.
- [3] BRASIL. MINISTÉRIO DA SAÚDE. Boletim Epidemiológico n. 18, v. 47. 2016. Disponível em: <<http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/secretarias/svs/noticiassvs/23618-boletim-epidemiologico-apresenta-perfil-da-morbimortalidade-por-doencasrespiratorias-cronicas>>. Acesso em: 27 maio 2017.
- [4] BRASIL. PORTAL BRASIL. Asma atinge 6,4 milhões de brasileiros. 2015. Disponível em: <<http://www.brasil.gov.br/saude/2015/01/asma-atinge-6-4-milhoes-de-brasileiros>>. Acesso em: 27 maio 2017.
- [5] CANDIAGO, R. H. Uso do DATASUS para avaliação de mudanças nos padrões das internações psiquiátricas no Brasil. 2007. 109 f. Dissertação (Mestrado) - Curso de Pós-graduação em Ciências Médicas: Psiquiatria, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/11406/000612639.pdf?sequence=1>>. Acesso em: 26 maio 2017.
- [6] PORTARIA 054-R. Diretrizes Terapêuticas para o Manejo da Asma não Controlada. Diário Oficial dos poderes do estado, 2009.
- [7] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. 2013.
- [8] SAS 9.3 SQL Procedure, User's Guide, 2012.
- [9] SBPT. Sociedade Brasileira de Pneumologia e Tisiologia. III Consenso brasileiro no manejo da asma. *Jornal de pneumologia*, São Paulo, v. 28,