

OBTENÇÃO, GERENCIAMENTO E VISUALIZAÇÃO DE SÉRIES TEMPORAIS DO DATASUS: Um estudo sobre a Asma no Paraná

Arthur Cesar de Moura Rocha, Paula Mitiko Heller, Eniuce Menezes de
Souza (Orientadora), e-mail: emsouza@uem.br.

Universidade Estadual de Maringá / Centro de Ciências Exatas/Maringá, PR.

PROBABILIDADE E ESTATÍSTICA ANÁLISE DE DADOS

Palavras-chave: DATASUS, Shiny, Asma

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. Assim, 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. Objetiva-se desenvolver e implementar um algoritmo para facilitar a obtenção de dados do DATASUS ao longo do tempo, geração de séries temporais, além da análise descritiva e visualização de tais séries, bem como de estatísticas descritivas em plataforma interativa em formato html. Em especial, serão analisados os dados de internações por asma disponível na base de dados do DATASUS a fim de exemplificar a utilização da aplicação.

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). Para o cumprimento desse objetivo, o mesmo gera 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.

Apesar disso, a dificuldade de manipulação de bancos de dados com grande volume de variáveis é uma reclamação frequente entre pesquisadores de diversas áreas do conhecimento. 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. Após realizada a proposta, com o intuito de verificar o cumprimento do

objetivo, foi realizada uma aplicação do método para um estudo descritivo sobre a asma.

Materiais e métodos

A metodologia para o desenvolvimento deste projeto foi baseada na utilização de recursos computacionais mais avançados, os quais envolvem uma extensão dos conceitos aprendidos durante o Curso de Graduação em Estatística. Neste sentido, com o auxílio do software R, foi desenvolvido um programa o qual realiza o download de forma automatizada dos bancos disponibilizados pelo DATASUS. Esses dados são convertidos de seu formato dbc em dbf com o auxílio do pacote "foreign".

Por meio do pacote "sqldf" foi possível o uso de funções e conceitos de Linguagem de Consulta Estruturada (Structured Query Language - SQL), para o agrupamento dos dados em um único conjunto macro por doença desejada. Para a aplicação deste trabalho, optou-se por utilizar as variáveis idade, sexo, cidade de internação e cidade de residência como covariáveis para as ocorrências de asma no estado do Paraná, uma vez que existe uma grande variação nas variáveis disponibilizadas pelo DATASUS ao longo dos anos.

Com a utilização dos pacotes "ggplot2", "leaflet" e "dygraphs" do R, foram gerados gráficos descritivos interativos das variáveis selecionadas. Por meio do Shiny foi possível a criação de um aplicativo que permite ao usuário o acesso a uma interface que apresenta três abas: na primeira, são apresentadas as informações de frequência de internações por asma por local de residência através de um mapa interativo; na segunda aba está apresentada a série temporal que informa a quantidade de internados conforme o passar do tempo, que pode ser apresentada levando em consideração a proporção de internados de cada sexo ou categoria de idade; e na terceira são apresentados dois gráficos de análise descritiva em que aparecem as proporções dos sexos e a quantidade de internados por categoria de idade, podendo ser para o estado do Paraná por inteiro ou levando em conta apenas uma cidade. Para que fosse possível o acesso remoto à essa aplicação foi utilizado o Shiny Server, o qual é um servidor desenvolvido especificamente para aplicativos do Shiny.

Resultados e Discussão

Como resultado final do trabalho desenvolvido, foram gerados mapas (Figura 1), gráficos descritivos e séries temporais de dados referentes a internações causadas por asma no estado do Paraná desde o ano de 1997 até o final do ano de 2017 (Figuras 2 a 4). As análises dos gráficos levaram à conclusão de que os locais de residência dos internados por asma são bem distribuídas através do estado considerando que, apesar do maior volume nas regiões metropolitanas, todas as regiões possui observações.

Além disso, tornou-se visível que a quantidade de internações por conta desta doença vem reduzindo de forma gradativa com o passar dos anos.

Figura 1 – Mapa de residência de internados por asma

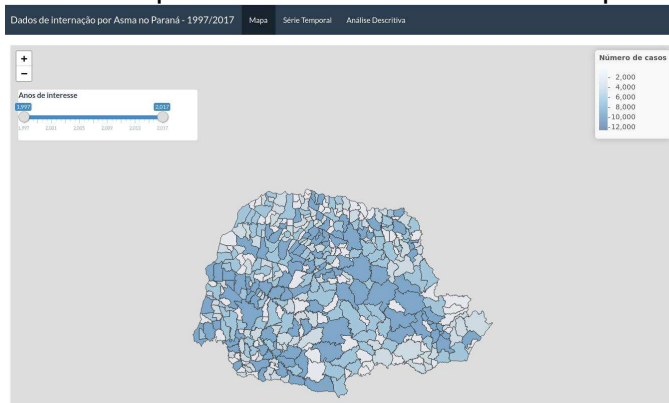


Figura 2 – Série temporal de internações por asma

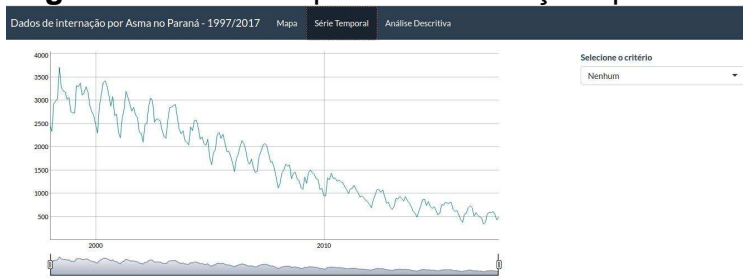


Figura 3 – Proporção de internações por sexo

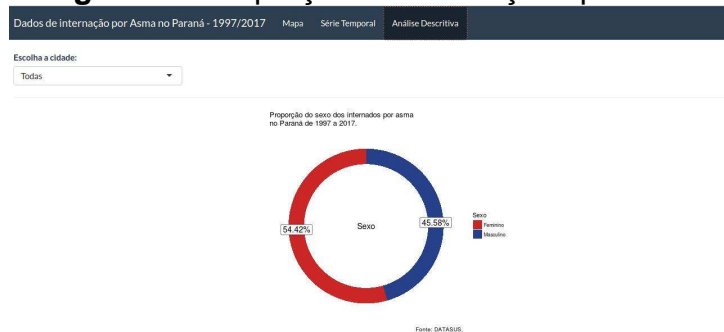
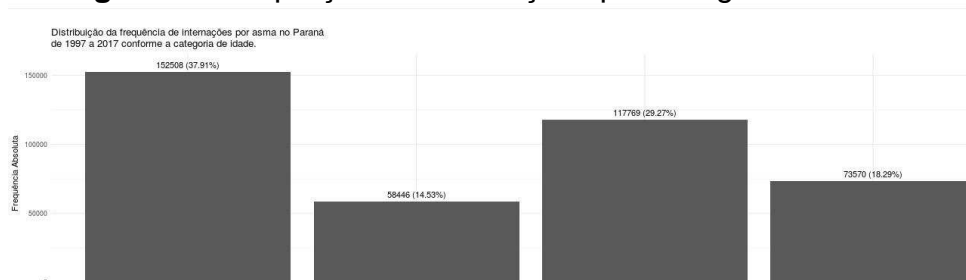


Figura 4 – Proporção de internações por categoria de idade



Apesar disso, é importante destacar que a doença apresenta um comportamento cíclico de picos de internações nos primeiros meses do ano.

Em acréscimo a isto, foi encontrado que a proporção de internações de mulheres é quase 10% maior que a de homens no estado do Paraná e que as crianças e os adultos são as categorias que mais são internadas por asma no Paraná, os quais representam mais de dois terços da população analisada.

Conclusões

O programa desenvolvido permite a utilização, de forma acessível, dos dados do DATASUS para análises descritivas. Dessa utilização para os casos de internação por asma no Paraná se concluiu que ao longo dos anos, o volume de internações vem diminuindo de forma quase constante. Apesar dessa redução, é visível existe um comportamento cíclico nas incidências de internação, onde há um pico nos primeiros meses do ano. Além disso, foi identificado que as mulheres são a maior parte dos internados pela doença no estado. Em acréscimo a isso, foi encontrado que as crianças representam mais de um terço das internações.

Este trabalho foi um passo inicial no sentido de aproximar os resultados de pesquisas realizadas na Universidade da sociedade, tanto em nível de graduação, como de pós graduação. Pretende-se que resultados de modelos desenvolvidos, predições e classificações, que são tão importantes para tomadas de decisão tanto de gestores quanto membros da comunidade de um modo geral, fiquem disponíveis de modo fácil, agradável e interativo. A partir do momento que se tem um servidor e todos os relatórios interativos são gerados diretamente do código em R, tem-se um facilitador que reduz o tempo do pesquisador e atende as demandas mais diversificadas da sociedade.

Além disso, o contato direto dos alunos com o download automático de dados, gerenciamento e geração de aplicativos interativos diretamente da linguagem R, vem de encontro com a atual tendência e expectativa do mercado de trabalho. Atualmente, as grandes empresas em nível nacional e internacional, usam o R/Shiny para acompanhamento contínuo e análise de dados, tais como, Google, Amazon, Boticário, Avon, Uber, Airbnb, dentre muitas outras. A busca pelo R/Shiny tem sido cada vez maior. Tais plataformas poderia ser desenvolvidas em SAS, entretanto, seria necessária a compra de diversos módulos extras, o que implicaria em altos custos para as empresas. Assim, o desenvolvimento do projeto trata-se de uma experiência importante aos alunos de graduação tanto para iniciação científica quanto para o desenvolvimento de habilidades para enfrentar o mercado de trabalho nesta era tecnológica.

Referências

BRASIL. DATASUS. Histórico / Apresentação. Disponível em: <<http://datasus.saude.gov.br/datasus>>. Acesso em: 26 maio 2017.