



DESENVOLVIMENTO DE NOVAS FUNCIONALIDADES NO PACOTE ESTATÍSTICO LECA

Ana Carolina Abreu Barbosa (PIBIC/FA), Deyse Márcia Pacheco Gebert, Airton Kist
(Orientador), e-mail: kist@uepg.br.

Universidade Estadual de Ponta Grossa/Departamento de Matemática e Estatística.

Probabilidade e Estatística, Probabilidade e Estatística aplicadas.

Palavras-chave: software R, pacote computacional, estatística básica.

Resumo

A acessibilidade ao desenvolvimento proporcionada pelo software gratuito e de código aberto R permite que pacotes sejam constantemente disponibilizados e atualizados em seu repositório. Inúmeros desses pacotes se destacam na área da estatística, porém exigindo um conhecimento prévio do usuário da linguagem R para a utilização de suas funções e interpretação de seus resultados, não abordando o tópico de forma educacional. Pacotes que oferecem uma interface gráfica ao usuário costumam ser mais acessíveis nesse cenário, uma vez que exigem pouco conhecimento adicional para sua utilização. O LECA, pacote em R desenvolvido nesse trabalho, oferece funções de distribuição de frequências para variáveis aleatórias, construção de diagramas de árvore a partir de cenários fornecidos pelo usuário, acesso ao software estatístico PQRS®, geração de testes de hipóteses ou intervalos de confiança e amostragem de uma variável. Essas funções são acessadas dentro de uma interface gráfica de usuário implementada em linguagem de programação Tcl. O resultado final facilita a manipulação e leitura de dados por seus usuários, mesmo os inexperientes com a área estatística ou com o software R.

Introdução

O R (R CORE TEAM, 2017) é um software gratuito e de código aberto que disponibiliza diversas funções estatísticas para seus usuários. O fácil acesso ao desenvolvimento proporcionado pela ferramenta estimula a criação e atualização constante de pacotes – que são grupos de uma ou mais funções executáveis dentro do ambiente – em seu repositório e permite que novas soluções e produções possam ser propostas por meio dela.

No Brasil, produções como os pacotes *ds* (ARNHOLD, 2014) e *ExpDes* (FERREIRA et al., 2014), oferecem, respectivamente, opções de estatística descritiva e análise de variância. Para a utilização desses pacotes é necessário um conhecimento prévio sobre a linguagem de programação R. Suas funções precisam ser invocadas manualmente por meio do terminal do software, com muitas delas exigindo um conhecimento mais avançado sobre a utilização de parâmetros dentro da ferramenta e de conceitos estatísticos. Em um ambiente educativo essas características podem ser custosas aos que possuem pouca ou nenhuma experiência com a ferramenta, principalmente para estudantes que não são da área da estatística.

Pacotes que incluem uma interface gráfica exigem pouco conhecimento adicional de seus usuários e costumam ser mais acessíveis e de fácil uso. O pacote *Rcmd* (FOX, 2005) é um exemplo de pacote estatístico em R que oferece interface gráfica para comunicação com seus usuários. O pacote se torna atrativo por fazer uso de diferentes tipos de objetos gráficos, facilitando o contato no ambiente.



Adicionalmente, softwares interativos se mostram ferramentas benéficas para o ensino da estatística, como abordado por Knypstra (1999). Estudantes mostram dificuldade na compreensão de termos que se diferem bastante de seus significados cotidianos (KNYPSTRA, 1999, apud HAWKINS et al., 1992).

O objetivo desse trabalho é a continuidade do desenvolvimento de um pacote em R, o LECA, que apresente suas funções em uma interface gráfica de usuário interativa capaz de autoajuste com foco na aprendizagem. Dessa maneira, será possível propiciar o acesso a estatística básica a todo tipo de usuário, bem como facilitar a realização de cálculos por meio das funcionalidades oferecidas pelo software R sem que o processo de aprendizagem de estudantes não familiares com essa ferramenta seja prejudicado.

Material e Métodos

Todo o embasamento teórico utilizado na implementação do LECA foi baseado nos seguintes livros: Bussab e Morettin (2013), Larson e Farber (2010) e Martins e Domingues (2014). Essas obras também apresentam modelos que influenciam a interface gráfica do pacote, como o formato da distribuição de frequências, seus gráficos, o diagrama de árvore de probabilidades, a apresentação de testes de hipótese e intervalos de confiança e o sistema de amostragem.

A linguagem de programação utilizada para a implementação da interface gráfica é a Tcl (OUTERSHOUT, 1989). O pacote *tcltk* (DALGARRD, 2001a; DALGARRD, 2001b) é capaz de reconhecer e executar a linguagem dentro do software R junto a sua extensão gráfica Tk, assim como o pacote *tcltk2* (GROSJEAN, 2014), e ambos são utilizados no desenvolvimento e utilização do LECA. Como a interface gráfica possui papel importante no pacote apresentado, foram realizadas pesquisas quanto a sua implementação no R com foco em usabilidade, assunto bastante discutido por Lawrence e Verzani (2014). O restante das opções oferecidas pelo LECA foram implementadas com o auxílio dos pacotes básicos do R.

O pacote foi desenvolvido e testado utilizando o R e sua versão gráfica, RStudio (RSTUDIO TEAM, 2016), ambos na versão 3.4.1 e em sistema operacional Windows nas versões 8.1 e 10.

Resultados e Discussão

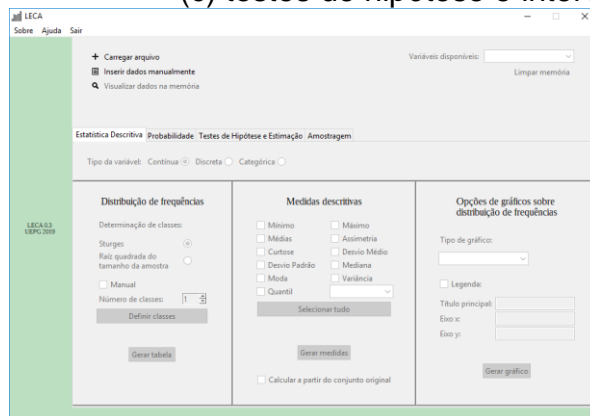
A tela principal do pacote LECA é apresentada na Figura 1, onde é possível observar as abas com suas funcionalidades. A manipulação de variáveis (inserção manual, carregamento por meio de um arquivo ou exclusão dos dados), pode ser feita por meio do cabeçalho do pacote, que pode ser acessado a qualquer momento.

Para auxiliar o estudante em sua utilização, o pacote conta com uma interface interativa: algumas opções são habilitadas apenas quando o usuário realiza uma ação. Por exemplo: a opção de gerar medidas descritivas de uma variável, disponível na área “Medidas Descritivas” da aba de estatística descritiva (Figura 1(a)), só se torna acessível após a inserção ou o carregamento de dados no pacote.

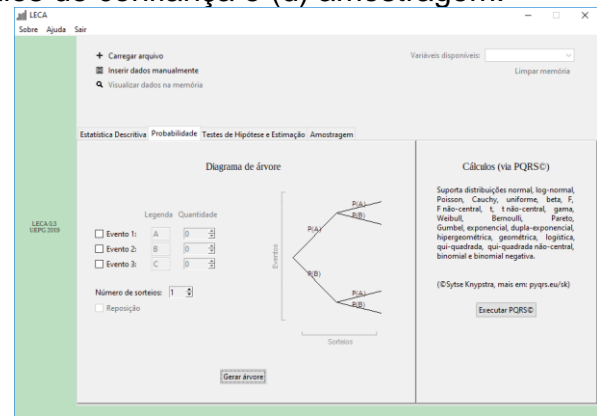
O software gratuito PQRS© foi desenvolvido por Sytse Knypstra (2000) e agrupa uma grande quantidade de modelos de distribuições de probabilidade para cálculos, podendo ser acessado por meio de seu botão de execução na aba de probabilidade, que simplesmente executa o software na máquina do usuário. O programa inclui gráficos interativos que facilitam a visualização de conceitos de estatística básica e apresenta resultados de forma prática. O programa não faz parte do desenvolvimento do LECA e todos os direitos autorais pertencem ao seu desenvolvedor. Para utilizar o PQRS© fora do ambiente R, basta fazer o download por meio do endereço pyqrs.eu/sk/.



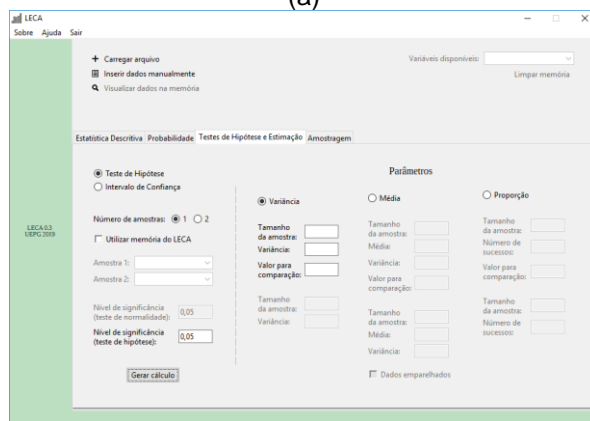
Figura 1 – Interface do LECA: opções de (a) estatística descritiva, (b) probabilidade, (c) testes de hipótese e intervalos de confiança e (d) amostragem.



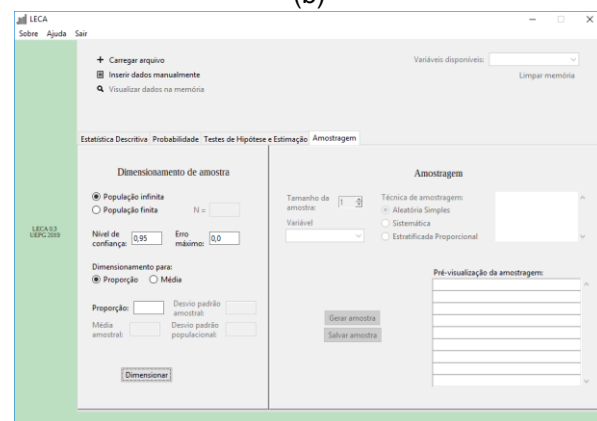
(a)



(b)



(c)



(d)

Fonte: os autores.

Um resultado parcial desse trabalho foi publicado em uma edição especial do periódico brasileiro *Sigmae* (BARBOSA, GEBERT e KIST, 2019).

Conclusões

O LECA oferece funções de estatística básica em uma interface gráfica interativa com foco na aprendizagem. Ele pode ser facilmente utilizado como material de apoio mesmo por aqueles que não possuem conhecimento sobre a linguagem R, o que facilita o contato de um estudante com um material virtual e com a estatística básica.

O pacote resultado desse trabalho ainda não se encontra no repositório oficial do R, pois ajustes em seu manual de utilização e a criação de novas funcionalidades estão sendo aplicados. Além disso, pretende-se realizar um teste prático com o pacote, o aplicando em aulas de estatística para a visualização do impacto de sua utilização em sala de aula.

Agradecimentos

A Universidade Estadual de Ponta Grossa e a Fundação Araucária pelo incentivo à pesquisa e pelas bolsas concedidas dentro do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) nos períodos de desenvolvimento do pacote.

Aos professores orientadores desse trabalho, Airon e Deyse, pelas discussões, ideias e conselhos compartilhados durante nossas reuniões, cujas conclusões levaram a concretização do LECA.



Referências

ARNHOLD, E. **Pacote em ambiente R para automatizar estatísticas descritivas**. Sigmae, 2014. v.3, n.1, p.36-42.

BARBOSA, A. C. A.; GEBERT, D. M. P.; KIST, A. **LECA**: Pacote educacional com interface gráfica de usuário para estatística descritiva e probabilidade no R. Sigmae, 2019. v. 8, p. 306-314.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 8ª ed., São Paulo: Saraiva, 2013. 548 p.

DALGARRD, P. The R-Tcl/Tk interface. **Proceedings of the 2nd International Workshop on Distributed Statistical Computing**, 2001a. ISSN 1609-395X.

DALGARRD, P. **A Primer on the R-Tcl/Tk Package**. R News, 2001b. v.1/3, p.27-31. ISSN 1609-3631.

FERREIRA, E. B.; CAVALCANTI, P. P.; NOGUEIRA, D. A. **ExpDes: An R Package for ANOVA and Experimental Designs**. Applied Mathematics, 2014. v.5, p.2952-2958.

FOX, J. **The R Commander**: A Basic Statistics Graphical User Interface to R. Journal of Statistical Software, 2005. v.14, n.9, p.1-42.

GROSJEAN, P. **SciViews**: A GUI API for R. UMONS, Mons, Belgium, 2014. Disponível em: <http://sciviews.org/SciViews-R>. Acesso em: 04 de março de 2019.

KNYPSTRA, S. **PyQRS**. 2000. Disponível em: <https://pyqrs.eu/sk/>. Acesso em: 10 de março de 2019.

KNYPSTRA, S. **Inference as a dynamic concept map**. 1999.

LARSON, R.; FARBER, B. **Estatística Aplicada**. 4ª ed., São Paulo: Pearson Prentice Hall, 2010. 637 p.

LAWRENCE, M.; VERZANI, J. **ProgGULinR**: support package for “Programming Graphical User Interfaces in R”. 2014. Pacote em R versão 0.0-4. Disponível em: <https://CRAN.R-project.org/package=ProgGULinR>. Acesso em: 04 de março de 2019.

MARTINS, G. A.; DOMINGUES, O. **Estatística Geral e Aplicada**. 5ª ed., São Paulo: Atlas, 2014. 416 p.

OUTERSHOUT, J. K. **TCL**: An Embeddable Command Language. Proceedings of the 1990 Winter USENIX Conference, 1989. p. 133-146.

R CORE TEAM. **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2017. Disponível em: <https://r-project.org>. Acesso em: 04 de março de 2019.

RSTUDIO TEAM. **RStudio**: Integrated Development for R. RStudio, Inc., Boston, 2016. Disponível em: <https://rstudio.com>. Acesso em: 04 de março de 2019.