

Evento: XX Jornada de Extensão

**PROJETO DE EXTENSÃO EM ANÁLISE ESTATÍSTICA COM O SOFTWARE
LIVRE R: O CASO DA UFFS CAMPUS CERRO LARGO¹
EXTENSION PROJECT IN STATISTICAL ANALYSIS WITH FREE
SOFTWARE A: THE CASE OF UFFS CAMPUS CERRO LARGO**

**Letícia Slodkowski², Iara Denise Endruweit Battisti³, Felipe Micaíl Da Silva
Smolski⁴, Erikson Kaszubowski⁵, Tatiane Chassot⁶, Denize Ivete Reis⁷**

¹ Projeto de Extensão realizado na Universidade Federal da Fronteira Sul- Campus Cerro Largo.

² Acadêmica do curso de Engenharia Ambiental e Sanitária da Universidade Federal da Fronteira Sul - Campus Cerro Largo, bolsita Cnpq, leticiaslodkowski98@gmail.com.

³ Doutora em Epidemiologia, Orientadora, docente da Universidade Federal da Fronteira sul, campus Cerro Largo, iara.battisti@uffs.edu.br.

⁴ Mestre no Programa de Pós-Graduação em Desenvolvimento e Políticas Públicas da Universidade Federal da Fronteira Sul - campus Cerro Largo, felipesmolski@hotmail.com.

⁵ Doutor em Psicologia. erikson84@yahoo.com.br

⁶ Doutora em Engenharia Florestal, docente da Universidade Federal da Fronteira Sul campus Cerro Largo. tatianechassot@uffs.edu.br.

⁷ Doutora em Qualidade Ambiental, docente da Universidade Federal da Fronteira Sul- campus Cerro Largo. denizeir@uffs.edu.br

Introdução

Os avanços dos recursos computacionais demandados pelo incremento na produção de informações na atualidade resultam em desafios e oportunidades sem igual à pesquisa acadêmica. Por um lado, maiores pressões para a realização de produções científicas requerem o aprendizado contínuo para elevar a confiabilidade dos resultados. Em outro sentido, a melhora dos *softwares* e o acesso a diversas fontes de dados oferece à produção de conhecimento a chance de que os trabalhos sejam produzidos com maior colaboração e eficiência entre os pares, sendo criados estudos com maior transparência, replicabilidade e reprodutibilidade.

É desejável, portanto, que o ferramental do campo da estatística seja absorvido e aprimorado pelos indivíduos em qualquer das etapas de sua formação acadêmica, em especial ao longo de seu curso de graduação. Nem por isto exclui, por consequência, os interessados na atualização dos conhecimentos de referência ou mesmo aqueles que desejam um aprendizado inicial no campo de análise de dados.

Neste sentido, o projeto de extensão "*Software R: capacitação em análise estatística de dados utilizando um software livre*" buscou conciliar o ensino de estatística básica com a utilização de tecnologias, no caso a linguagem de programação R e seu console RStudio. O projeto foi aprovado no Edital de Apoio a Programas de Extensão (Nº 522/GR/UFFS/2016) da Universidade Federal da Fronteira Sul (UFFS). Como premissas principais, contou com a integração de colaboradores de

Evento: XX Jornada de Extensão

diversas áreas na produção e disseminação do curso, bem como oportunizou atender a demanda interna e externa de pessoas interessadas neste curso gratuito, pois atendeu a alunos de graduação, pós-graduação, bolsistas e comunidade em geral. Contando com oferta anual dos módulos Básico e Avançado no laboratório de informática do Câmpus Cerro Largo (RS), teve sua primeira oferta em 2016.

Metodologia

Este curso de extensão objetiva introduzir e/ou aprimorar os conhecimentos sobre estatística utilizando o *software* R, com adesão aberta e gratuita à comunidade interna e externa. Para isto, o conteúdo do curso foi elaborado pelos colaboradores responsáveis (professores, alunos, bolsistas e colaboradores externos), sendo reunidos em forma de livro e digitalmente disponibilizados aos participantes ao longo das aulas. Os encontros expositivos contam com certificação para os participantes com frequência mínima de 75%, contabilizando 40 horas para os cursos Básico e Avançado. O requisito para participação foi o conhecimento de estatística básica por parte dos inscritos. O módulo Básico divide-se em: Módulo 1 - Introdução ao R; Módulo 2 - Estatística Descritiva; Módulo 3 - Inferência Estatística; Módulo 4 - Teste de Qui-Quadrado; Módulo 5 - Modelos de Regressão. Já o módulo avançado conta com os conteúdos: Módulo 6 - Experimentação; Módulo 7 - Análise Fatorial; Módulo 8 - Modelos de Regressão Múltipla; Módulo 9 - Modelos de Regressão Logística; Módulo 10 - Rmarkdown. O material criado para o curso está em constante aprimoramento e está sendo encaminhado para publicação subsequente em forma de um livro.

Resultados e discussão

A linguagem de programação R foi criada na década de 1990 pelos estatísticos da Universidade de Auckland (Nova Zelândia), Ross Ihaka e Robert Gentleman. Tiveram como intuito inicial desenvolver um programa acessível para suas necessidades ao passo que poderiam de forma autônoma efetuar análises estatísticas. Com a expansão da internet, sua solução atraiu cientistas de todo o mundo, pois a ferramenta além de ser de código aberto disponibiliza ao pesquisador a programação de rotinas e a incorporação de diversos pacotes, com funções específicas (TIPPMANN, 2015). Em 2018 já haviam mais de 12.700 pacotes disponíveis aos utilizadores, sendo que possui uma comunidade ativa de desenvolvedores (GENTLEMAN; IHAKA; BATES, 2018). Dentre seus benefícios estão a extensa gama de modelos estatísticos que o programa trabalha (GANDRUD, 2013) e a produtividade e flexibilidade auferida pelo aprendizado dos comandos de programação (AQUINO, 2014). Ainda, o *software* R contempla as análises gráficas, manipulação de dados, tratamento, cálculo, armazenamento e carregamento de diversas bases, cálculos numéricos em vetores e matrizes, exibição dos resultados na tela ou no console (FERREIRA, 2018).

Evento: XX Jornada de Extensão

Já o programa RStudio se constitui de um console amigável com versão *open source*, ou seja, um ambiente de desenvolvimento aprimorado que integra a linguagem R (GANDRUD, 2013). É necessária a instalação dos dois programas (R e RStudio) no computador do usuário. Foi desenvolvido em 2010 e facilitou a execução dos códigos de programação, o acesso às definições das funções, a plotagem de gráficos, a administração de múltiplos diretórios, a depuração dos resultados e do histórico (RSTUDIO, 2017). Além disso, combina outras linguagens de marcação para a saída de diferentes documentos como resultado final (LaTeX, Markdown, HTML, PDF, MS Word, txt, Libre Office, por exemplo) e utiliza também o pacote RMarkdown para isto (AQUINO, 2014). Estimula assim a cooperação entre os pares, visto que os trabalhos são facilmente compartilháveis entre os pesquisadores.

A utilização destas ferramentas para análises estatísticas contribuem para que as pesquisas possam ser replicadas, ou seja, repetidas de forma intencional para que seus resultados sejam corroborados ou não por outras análises (MAKEL, 2014). Já o conceito de reprodutibilidade atua no sentido de repetir os experimentos para que sejam verificados se foram efetuados os controles legitimados e os testes estatísticos apropriados, fazendo com que o experimento científico seja validado ou não (BEGLEY; IOANNIDIS, 2015).

Além do aperfeiçoamento da utilização das técnicas estatísticas, o projeto de extensão proporciona o cumprimento do tripé ensino-pesquisa-extensão atribuídos às Universidades Federais (SLEUTJES, 1999), sendo relevante a democratização da extensão e a cooperação com a sociedade (TREVISOL, 2010).

Desta forma, após a divulgação do projeto de extensão nas dependências da UFFS, na internet (<https://smolski.github.io/softwarelivrer/>) e por e-mail para a comunidade externa, abrindo-se o período de inscrições disponibilizando 30 vagas para cada módulo básico, sendo ofertado no primeiro semestre ano de 2019. Inscreveram-se 30 alunos para o curso. Dentre os inscritos ao curso básico 58,5% consiste do sexo feminino e o restante do sexo masculino, destes 35,8% são do curso de Administração, 26,4% do curso de Engenharia Ambiental e Sanitária, 18,9% do curso de Agronomia, e os demais são discentes do curso de Ciências Biológicas.

Ao final dos módulos do curso de extensão estão sendo disponibilizadas enquetes, respondidas digitalmente, para avaliar as opiniões sobre o alcance dos objetivos do projeto pelos estudantes. A seguir seguem as respostas dos alunos que realizaram o curso básico de 2019, sendo que se constituíram 15 respondentes (o módulo avançado será trabalhado em seguida).

Ao serem questionados se “os objetivos do curso foram alcançados”, 46,7% concordaram totalmente e 46,7% concordaram parcialmente. Quando questionados se “o conteúdo abordado no curso permite estabelecer relações entre teoria e prática”, 66,7% concordaram totalmente e 33,3% concordaram parcialmente.

Em “o horário do curso e seu turno foram adequados”, 33,3% responderam que concordam totalmente e 33,3% concordam parcialmente. Quando questionados se “os recursos físicos

Evento: XX Jornada de Extensão

(laboratório de informática, computadores, projetores) foram adequados”, 53,3% concordaram totalmente e 26,7% concordaram parcialmente.

A produção do material surpreendeu positivamente os colaboradores, pois a afirmação “o material didático disponibilizado foi de acordo com as aulas do curso” obteve 86,7% de concordância total e 13,3% concordam parcialmente. Demonstra-se assim a elevada qualidade da produção realizada.

Quanto à autoanálise dos estudantes, a afirmação “revisei e estudei os conteúdos abordados no curso” foi respondida por 46,7% como concordo parcialmente, e 33,3% como concordo totalmente. Já 46,7,3% concordaram totalmente com a assertiva “fui assíduo aos encontros do curso”. Quando responderam à questão “expressei minhas dúvidas e opiniões sobre os conteúdos abordados”, 50% concordaram parcialmente e 46,4% concordaram totalmente.

Uma questão relevante de avaliação dos conteúdos trabalhados foi “procurei estabelecer relações entre os conteúdos abordados no curso com os problemas e outros conhecimentos abordados ao longo da vida acadêmica e profissional”. Em 60% das respostas os alunos concordaram parcialmente, 26,7% concordaram totalmente e 13,3 % nem concordaram nem discordaram.

Considerações finais

Os *softwares* estatísticos apresentam-se com grande importância na aquisição de conhecimento, bem como a colaboração científica entre pesquisadores possibilitados através das ferramentas livres. Outra característica consiste na maior praticidade, agilidade e robustez nas análises em pesquisa acadêmicas. Neste sentido, o curso de extensão universitária aqui realizado é importante, pois dissemina saberes assim incorporando a comunidade interna e externa, alunos e professores, sendo um dos pilares das Universidades Federais brasileiras. Auxiliam na produção de conhecimento realizada nas pesquisas da UFFS Campus Cerro Largo, oportunizando a constante qualificação do capital social local e o desenvolvimento.

Palavras-chave: Softwares estatísticos; R,RStudio; Análise e interpretação de dados.

Referências

AQUINO, J. A. DE. **R para cientistas sociais**. Ilhéus, BA: Editus, 2014.

BEGLEY, C.; IOANNIDIS, J. Reproducibility in science Reprodutibilidade. **Circulation research**, 2015.

FERREIRA, P. G. C. **Análise de Séries Temporais em R: curso introdutório**. 1. ed. Rio de Janeiro: Elsevier, 2018.

GANDRUD, C. **Reproducible Research with R and R Studio**. 2. ed. London: CRC The R Series,

Evento: XX Jornada de Extensão

2013.

GENTLEMAN, R.; IHAKA, R.; BATES, D. **The R project for statistical computing**. Disponível em: . Acesso em: 10 jun. 2017.

MAKEL, M. C. The empirical march: Making science better at self-correction. **Psychology of Aesthetics, Creativity, and the Arts**, v. 8, n. 1, p. 2-7, 2014.

RSTUDIO. **Take control of your R code**. Disponível em: . Acesso em: 10 jun. 2017.

SLEUTJES, M. H. S. C. Refletindo sobre os três pilares de sustentação das universidades : ensino-pesquisa-extensão. **Rev. Adm. Pública**, v. 33, n. 3, p. 99-111, 1999.

TIPPMANN, S. Programing Tools: Adventures with R. **Nature**, v. 157, p. 109-110, 2015.

TREVISOL, J. V. **A UNIVERSIDADE E O DESENVOLVIMENTO REGIONAL: subsídios para conceber as políticas de extensão e de pesquisa da UFFS**. I Seminário de Interação entre Universidade e Comunidade: a UFFS e o desenvolvimento regiona. **Anais...2010** Disponível em: . Acesso em: 4 jul. 2017