



CAPACITAÇÃO EM ANÁLISE ESTATÍSTICA DE DADOS COM USO DO SOFTWARE LIVRE R

*Felipe Micaíl da Silva Smolski¹
Iara Endruweit Battisti
Tatiane Chassot,
Denize Ivete Reis
Erikson Kaszubowski
Djaina Sibiani Rieger*

RESUMO

O conhecimento gerado pelo desenvolvimento da ciência elevou, e muito, a resolução de problemas e a capacidade de melhorias nas condições de vida. Ao mesmo tempo, o incremento no volume de informações disponíveis cresce exponencialmente com o advento das tecnologias e inovações. Os estudos acadêmicos, por sua vez, demandam cada vez mais transparência, a contar pela necessidade de que os resultados das descobertas científicas sejam plenamente reproduzidos e replicados por outros pesquisadores, validando a robustez dos seus conteúdos. O objetivo deste projeto de extensão foi disseminar os saberes a respeito da utilização da linguagem de programação livre R e seu console RStudio, voltados aos conhecimentos de Estatística Básica e estudos reprodutíveis. Voltou-se para alunos, professores e comunidade em geral da Universidade Federal da Fronteira Sul Câmpus Cerro Largo - RS. Como metodologia, elaboraram-se oficinas no laboratório da Universidade, com prévia criação e disponibilização de material dos conteúdos abordados. Quanto aos resultados, observou-se que a atuação da Universidade, no sistema de extensão universitária, possui papel fundamental para a formação do conhecimento, visto complementar a consistência dos saberes entre discentes e docentes, efetuar integração da sociedade com o ensino superior, em seu papel social por excelência, e contribuir para o aperfeiçoamento da utilização de técnicas estatísticas proeminentes no cenário acadêmico pelas diversas áreas do conhecimento.

Palavras-chave: R. Software Estatístico. RStudio. Extensão Universitária.

TRAINING IN STATISTICAL DATA ANALYSIS USING THE FREE SOFTWARE R

ABSTRACT

The knowledge generated by scientific developments has resolved many problems and increased the capacity for living condition improvements. Concurrently, the volume of information available grows exponentially with the advent of technologies and innovations. Academic studies demand more and more transparency due to the need for results of scientific discoveries to be fully reproduced and replicated by other researchers, validating the robustness of their contents. The objective of this extension project was to disseminate

¹ Universidade Federal da Fronteira Sul (UFFS) Câmpus Cerro Largo. Contato: felipesmolski@hotmail.com

knowledge about the use of the free programming language R and its RStudio console to students, professors, and community members at the Federal University of Fronteira Sul–Cerro Largo Campus in Rio Grande do Sul. Workshops focused on basic statistical knowledge and reproducible studies and were prepared in the University's laboratory. It was observed that the University's performance in the university extension system has a fundamental role to contribute in the formation of knowledge while complementing the consistency of knowledge among students and professors. Extension projects integrate society with higher education and contribute to the improvement of the use and accessibility of prominent statistical techniques in the academic setting.

Keywords: R. Statistical Software. RStudio. University Extension.

CAPACITACIÓN EN ANÁLISIS ESTADÍSTICA DE DATOS UTILIZANDO EL SOFTWARE LIBRE R

RESUMEN

El conocimiento generado por el desarrollo de la ciencia elevó muchísimo la resolución de problemas y la capacidad de mejoras en las condiciones de vida. Al mismo tiempo, el incremento en el volumen de información disponible crece exponencialmente con el advenimiento de las tecnologías e innovaciones. Los estudios académicos, a su vez, demandan cada vez más transparencia, a contar por la necesidad de que los resultados de los descubrimientos científicos sean plenamente reproducidos y replicados por otros investigadores, validando la robustez de sus contenidos. El objetivo de este proyecto de extensión fue diseminar los saberes acerca de la utilización del lenguaje de programación libre R y su consola RStudio, volcados a los conocimientos de Estadística Básica y estudios reproducibles. Se destinó a alumnos, profesores y comunidad en general de la Universidade Federal da Fronteira Sul - Campus Cerro Largo - RS. Como metodología, se elaboraron talleres en el laboratorio de la Universidad, con previa creación y disposición de material de los contenidos abordados. En cuanto a los resultados, se observó que la actuación de la Universidad en el sistema de extensión universitaria tiene un papel fundamental para contribuir a la formación del conocimiento, mientras que complementa la consistencia de los saberes entre discentes y docentes, se realiza integración de la sociedad en la enseñanza superior, en su papel social por excelencia, y contribuye al perfeccionamiento de la utilización de técnicas estadísticas prominentes en el escenario académico por las diversas áreas del conocimiento.

Palabras clave: R. Software Estadístico. RStudio. Extensión Universitaria.

INTRODUÇÃO

A crescente demanda por maior transparência e rigor nos estudos científicos exige constante atualização acerca dos processos estatísticos de referência e dos recursos computacionais de destaque na comunidade acadêmica. Também, ensinam a utilização de programas que não só permitam agilidade, segurança e robustez e possibilitem incrementos na cooperação com a produção científica entre os pesquisadores, mas também facilitem a sua divulgação ao público.

A estatística, por sua vez, auxilia nos métodos para a coleta, organização, descrição, análise e interpretação de dados, propiciando sua utilização na tomada de decisões (CORREA, 2003). De forma complementar, o aumento do poder de processamento e memória dos computadores, nos últimos anos, favoreceu fortemente a análise estatística, sendo ao mesmo tempo a linguagem R a maior ferramenta criada pela estatística nos últimos vinte anos, pois explora todo o poder computacional em um único ambiente de programação para várias necessidades (JELIHOVSCHI, 2014).

Na atualidade, dispõe-se de uma variedade de *softwares* estatísticos, proprietários e livres. Os *softwares* proprietários possuem um custo para adquirir a licença para o uso, ao passo que os ambientes de programação livres, como o R e seu console RStudio, facilitam o acesso do público. Ainda, são ferramentas promissoras quanto a sua robustez e abrangência estatística, facilidade na distribuição dos códigos utilizados e na elevação dos parâmetros de reprodutibilidade e replicabilidade dos resultados, visto que todos os passos efetuados pelo pesquisador em sua análise podem ser acompanhados.

O importante papel da extensão universitária no Brasil se alinha à necessidade de maior interação entre o conhecimento gerado pelos pesquisadores e sua apropriação por parte da sociedade, seja o público externo seja o interno. No primeiro caso, ela está voltada a espalhar as possíveis atividades que vão ao encontro das soluções e da capacitação desejada pelo público local, e, no segundo caso, amplia a disseminação das técnicas e práticas mais atualizadas para as demais áreas do conhecimento.

O presente projeto “*Software R: capacitação em análise estatística de dados utilizando um software livre*” tem por objetivo conciliar o ensino de análises estatísticas e o uso das tecnologias, por meio da linguagem R, utilizando o console RStudio. Além disso, busca a integração entre os colaboradores do projeto nas suas diversas áreas, valorizando os pressupostos de extensão universitária para a sociedade e disseminando o conhecimento gerado pelo capital social da academia. O projeto originou-se da sua aprovação em Edital de Apoio a Programas de Extensão (Nº 522/GR/UFFS/2016) da Universidade Federal da Fronteira Sul (UFFS), ofertando anualmente os módulos básico e avançado no laboratório de informática do Câmpus Cerro Largo – RS. Sua primeira oferta ocorreu no ano de 2016, dirigida a um público de alunos e professores de graduação e pós-graduação, bolsistas e comunidade em geral.

METODOLOGIA

Para cumprir os objetivos deste curso de extensão, realizou-se a criação de material, a disponibilização e a execução de aulas expositivas, pelos professores, alunos e bolsistas da UFFS Câmpus Cerro Largo - RS, com certificação para os participantes com frequência mínima de 75% e computando-se 40 horas de atividades. As oficinas tiveram o propósito de abranger alunos e professores de graduação e pós-graduação, bolsistas, comunidade acadêmica e regional. Na segunda edição do curso, as aulas foram realizadas entre os dias 25/05/2017 e 12/06/2017, no laboratório de informática do Câmpus Cerro Largo, sendo o objetivo do projeto de extensão ofertar anualmente dois módulos: Básico e Avançado. Além disso, o material com o conteúdo formulado foi disponibilizado eletronicamente aos participantes, já que o objetivo posterior era a publicação dos conteúdos fundamentais, em forma de livro, para cada módulo. Dessa forma, o material poderá ser utilizado pelos discentes nas aulas com as disciplinas de estatística e correlacionadas, bem como contribuirá para a disseminação do conhecimento efetuado pela universidade.

O papel da extensão universitária na sociedade

A Constituição Federal de 1988, no seu artigo 207, além de atribuir autonomia didático-científica, administrativa, financeira e patrimonial às universidades, determina como indissociáveis o ensino, a pesquisa e a extensão. Esse tripé torna-se, portanto, essencial para ressaltar a relevância da universidade frente aos valores fragmentados que emergiram com a pós-modernidade ([SLEUTJES, 1999](#)).

A extensão universitária, para [Oliveira e Goulart \(2015\)](#), possui a função institucional de elo entre o ensino e a pesquisa científica. Quanto ao seu percurso histórico, a extensão sofreu alterações em suas características constituintes, originadas com atributos de prestador de serviços à sociedade; num segundo momento, ela foi marcada com predicados assistencialistas e, finalmente, transformou-se em uma função dialógica. Esta última, por sua vez, constitui-se no processo atual da extensão universitária e se mostra avessa à simples transmissão “bancária” de conhecimentos. Entretanto, devem-se ressaltar a interdisciplinaridade, a interprofissionalidade e a notoriedade que a caracterizam na à formação do estudante e também o impacto na transformação social do indivíduo.

De outro ponto de vista, a relação entre a extensão universitária e a cultura é explicitada por [Arruda \(2010\)](#), o qual reconhece primeiramente que a prática científica exerce, na atualidade, é papel principal da academia, graças a conhecimentos disponibilizados por suas instituições e pesquisadores. No entanto, nem sempre essa prática científica ocorre em locais onde há iniciativas culturais, sendo necessário uma aliança entre ciência e cultura nos quadros universitários, sem se deixarem de lado a formação estudantil e a produção científica; deve-se, pelo contrário, ultrapassar a simples divulgação de conhecimento ou a prestação de serviços no atendimento de demandas, para haver o compartilhamento dos avanços alcançados.

O domínio do conhecimento científico e tecnológico é fundamental para contribuir com melhorias das condições de vida da sociedade. Nesse entendimento, as ações de extensão das universidades e a interação com os agentes locais são primordiais, pois exaltam a divulgação científica para cultivar uma maior cultura científica; fazem circular o conhecimento dentro do ambiente das universidades e difundi-lo, envolvendo os diversos atores da sociedade na produção dos grupos de trabalho ([MARANDINO, 2013](#)). Em complemento, a extensão universitária é capaz de transmitir os conhecimentos previamente gerados para o âmbito da sociedade, influenciando significativamente na vida das pessoas que, de forma genuína, se interessarem por essa transferência de saberes ([SLEUTJES, 1999](#)).

A execução das ações e cursos de extensão universitária, por outro lado, deve atentar para as dificuldades enfrentadas pelos estudantes, bolsistas e público em geral em seu desenvolvimento (as quais constituem-se obstáculos em seu percurso). Algumas das dificuldades descritas pelos agentes que participam dos cursos de extensão são: a preparação para o novo; o encontro e o contato do aluno com a sociedade, para além das aulas teóricas das disciplinas; o despreparo na transferência de determinado conhecimento ou mesmo o ajuste da linguagem acadêmica ao público externo em seu processo de comunicação; a falta de recursos; a desatenção do público participante; a falta de continuidade dos cursos de extensão na comunidade ([OLIVEIRA; ALMEIDA JÚNIOR; SILVA, 2016](#)).

Os programas de extensão universitária são marcados pelo desenvolvimento das seguintes condições: (a) os *meios/recursos* para implementação da extensão, como os extensionistas envolvidos (corpo docente, discentes, técnicos-administrativos, alunos da

pós-graduação), as bolsas e recursos financeiros, os conhecimentos e infraestrutura disponibilizada pelas universidades; (b) as *tarefas/atividades* elaboradas, representadas por treinamentos, cursos, visitas técnicas, palestras e atendimentos; (c) a criação dos *produtos* da extensão pelos participantes, bens tangíveis ou intangíveis de que se beneficiam os públicos interno e externo, como livros, manuais, monografias, dissertações, eventos, anais, artigos, *softwares*, publicações em revistas; (d) os benefícios que esses novos conhecimentos trazem para o público participante; (e) os *impactos* ou as mudanças de longo prazo que ocorrem devido à utilização dos saberes induzidos pelas ações de extensão ([BUVINICH; AMORIM, 2013](#)).

No caso no Câmpus Cerro Largo, no estado do Rio Grande do Sul, a instalação da Universidade Federal da Fronteira Sul (UFFS) vem contribuir fortemente para o desenvolvimento humano, científico e tecnológico da região. Dessa forma, no entendimento de [Trevisol \(2010\)](#), as pesquisas e os cursos de extensão revelam a concepção que a Universidade possui de si própria nesse contexto sociocultural. Ainda, segundo o autor, a extensão como objetivo-fim deve ser operacionalizada para livrar-se de seu caráter fragmentário e mercantilista, que por vezes marca as instituições, democratizando seu acesso e aumentando a interação e a cooperação com a sociedade.

A linguagem de programação R e o RStudio

O projeto de extensão, aqui proposto, de qualificação da comunidade para a utilização de tecnologias voltadas à estatística em várias disciplinas da academia, alinha-se com os objetivos universitários da disseminação do conhecimento para a sociedade. Para tanto, é preciso que se aproveitem, desde logo, as possibilidades disponibilizadas pelas novas ferramentas que facilitam as análises estatísticas, neste caso, os ambientes de programação R e o RStudio.

Na atual era tecnológica, as análises estatísticas de dados, que anteriormente eram desenvolvidas em planilhas manuais, não mais necessitam de longos períodos de tempo para coleta e apreciação, e para a consequente produção de resultados, devido ao uso de melhores *softwares* estatísticos ([AGUIAR et al., 2016](#)). A ciência da computação, em seus avanços recentes, tendo de enfrentar uma complexidade cada vez maior das bases de dados, ainda possui limitações para avaliar seus resultados e publicá-los nas revistas científicas ([PENG, 2011](#)). Contudo, o extraordinário desenvolvimento recente dos *softwares*, também na comunidade de código aberto, possibilita que as pesquisas sejam classificadas como reprodutíveis, quando e se os autores do projeto conseguirem disponibilizar todo o material para outros autores replicarem o trabalho sem outras informações adicionais ([OROZCO et al., 2018](#)).

A linguagem R possui uma extensa gama de modelos estatísticos (modelagem linear e não-linear, testes estatísticos clássicos, análise de séries temporais, agrupamento, classificação, etc.) e técnicas gráficas. É preciso efetuar uma primeira distinção acerca de ambos os programas utilizados neste curso de extensão, pois o programa R representa linguagem de programação primária para estatísticas e gráficos, enquanto o RStudio é um console amigável para a utilização desta linguagem ([GANDRUD, 2013](#)). Embora seja uma linguagem de programação, em que os comandos devem ser aprendidos pelo usuário, o domínio da técnica possibilita que o estudante trabalhe com elevada produtividade e eficácia ([AQUINO, 2014](#)).

Uma característica importante do R é o fato de ele possuir uma comunidade ativa de desenvolvedores, que está se expandindo regularmente, ao mesmo tempo que abrange

uma gama de disciplinas para analisar dados e rodar diversas análises estatísticas ([GANDRUD, 2013](#)). O *software* R contempla a manipulação de dados, cálculo e exibição gráfica, possibilitando eficiência para tratamento e armazenamento de dados, cálculos numéricos em vetores e matrizes, ferramentas intermediárias para análise, exibição das análises na tela ou em console, e linguagem de programação bem desenvolvida ([FERREIRA, 2018](#)).

Esta linguagem de programação foi criada na década de 1990 pelos desenvolvedores Ross Ihaka e Robert Gentleman, estatísticos da Universidade de Auckland, na Nova Zelândia, com o intuito de torná-lo um *software* prático para as suas necessidades, para poderem analisar dados de forma autônoma. Com o incremento no acesso à rede mundial de computadores, esta linguagem atraiu a atenção dos pesquisadores e cientistas em vários países, pois poderiam tornar seus códigos acessíveis, ao mesmo tempo que pré-programavam suas rotinas ou mesmo contribuíam com a criação de diversos pacotes ([TIPPMANN, 2015](#)).

Na atualidade (2018), já são utilizados mais de 12.700 pacotes ([GENTLEMAN; IHAKA; BATES, 2018](#)) para diferentes propósitos, desde a comparação de genomas humanos, passando pela modelagem do crescimento da população, predição de preços, criação de gráficos, escrita e publicação de manuscritos, etc. As ciências biológicas e a agricultura apresentaram importantes incrementos na produção científica, por utilizarem o ferramental da linguagem R desde os anos 2000, seguidas pela ciência ambiental e pela área da bioquímica, genética e biologia molecular ([TIPPMANN, 2015](#)). Outra vantagem do R é que o usuário encontra diversas funções estatísticas para utilizar, sem precisar saber como elas foram codificadas. É uma plataforma de sucesso no meio acadêmico e empresarial por possibilitar a combinação da estatística e da mineração de dados no mesmo ambiente ([AMARAL, 2016](#)).

O *software* RStudio, por sua vez, é um ambiente de desenvolvimento integrado que combina R com outras linguagens de marcação (LaTeX, Markdown e HTML, por exemplo) e diversos pacotes que executam várias funções ([GANDRUD, 2013](#)). Em seu console suporta diretamente a execução dos códigos de programação, faz a plotagem dos dados e gráficos, guarda o histórico, depura e gerencia o espaço de trabalho. Em sua versão *open source* ainda possibilita o rápido acesso às definições das funções, bem como a administração de múltiplos diretórios de trabalho e o acesso a uma extensa lista de pacotes de desenvolvimento ([RSTUDIO, 2017](#)).

Além de gratuito, o RStudio, que começou a ser desenvolvido em 2010, facilita a reutilização de códigos de programação e plotagens de resultados e uma rápida alteração ou reorganização das variáveis utilizadas em cada base de dados. Em adição, possui uma versão *Server*, em que o pesquisador utiliza um *login* e senha para acesso via web por meio de um navegador de *internet*. A colaboração entre os pares é estimulada, pois os projetos tornam-se facilmente compartilháveis entre os pesquisadores, fazendo com que cada um contribua para a produção dos trabalhos de forma concomitante, aumentando a eficiência da resolução de problemas. Ainda, a ferramenta RMarkdown, dentro do RStudio, contribui para que sejam realizados trabalhos mais ricos em seus detalhes, documentados por completo, também com a possibilidade de inclusão de comentários dos programadores, para posterior apresentação final em formatos como HTML, PDF, Word ou LaTeX ([AQUINO, 2014](#)).

Estudos reprodutíveis e replicáveis com a linguagem de programação R

Embora pareçam semelhantes, a reprodutibilidade e a replicabilidade possuem sutis diferenças. A replicação de um estudo objetiva a repetição de forma intencional, em outro contexto, da pesquisa realizada, para que sejam corroborados ou não os resultados anteriores ([MAKEL; PLUCKER, 2014](#)). Já a reprodutibilidade tem como objetivo central repetir os experimentos efetuados inicialmente, utilizando e verificando se foram utilizados os controles legitimados e testes estatísticos apropriados. Faz, portanto, com que o experimento científico revele sua validade e, por consequência, a ciência avance ([BEGLEY; IOANNIDIS, 2015](#)).

Com relação aos trabalhos científicos de cunho empírico e experimental, ainda existem intensos desafios a serem superados quanto à reprodução e replicação dos trabalhos de maneira geral, o que compromete a confiabilidade deles na comunidade científica, gerando retrabalho e desconfiança. No campo da educação, esforços têm sido empregados para melhorar as práticas científicas no sentido de que sejam encontradas as evidências do que realmente funciona nessa área do conhecimento, potencializando a confiabilidade dos estudos ([MAKEL; PLUCKER, 2014](#)).

Na área da saúde, têm-se reconhecido as fraquezas existentes acerca do sistema de pesquisas básica e pré-clínica, pois destaca-se a incapacidade de replicação de grande parte dos trabalhos em pesquisas publicadas em revistas importantes, que se estima estarem entre 75 e 90% do total dos artigos científicos. Ainda, a não-utilização de técnicas apropriadas para a experiência desejada contribui para que o experimento não seja replicado, e a conclusão principal do trabalho não tenha sido fundamentada ([BEGLEY; IOANNIDIS, 2015](#)). Em relação às pesquisas no campo da econometria, os desafios quanto à análise estatística e sua reprodutibilidade são enormes, com vistas a incentivar uma melhor distribuição de seus resultados e o enriquecimento dos detalhes disponibilizados dos trabalhos, bem como a utilização de ferramentas mais eficazes quanto à programação e arquivamento dos dados ([KOENKER; ZEILEIS, 2009](#)).

Uma vez que nenhum processo de pesquisa atual é completamente linear, pois há que se conseguir novos dados e alterar variáveis ou técnicas estatísticas, os documentos dinâmicos das pesquisas reproduzíveis deixam os processos mais fáceis de serem efetuados. Além disso, as pesquisas reproduzíveis são mais prováveis de serem utilizadas por outros pesquisadores do que as demais ([GANDRUD, 2013](#)). São necessários incrementos na reprodutibilidade dos trabalhos para que sejam elevados os padrões de julgamento científico; uma barreira é a disponibilização dos códigos utilizados no caso da ciência da computação. Há que se cultivar portanto uma "cultura da reprodutibilidade" ([PENG, 2011](#)).

A replicação de estudos de referência torna-se essencial para o desenvolvimento de políticas baseadas em evidências e práticas científicas melhores, visto elevar a confiabilidade dos ambientes educacionais ([MAKEL; PLUCKER, 2014](#)). Além disso, os diversos atores das instituições possuem responsabilidades, como os funcionários, alunos e bolsistas pós-doutorados, tornando-se a atualização anual dos principais pesquisadores de projetos experimentais importante, inclusive para pesquisadores seniores de instituições conhecidas ([BEGLEY; IOANNIDIS, 2015](#)).

Os pesquisadores que efetuam experimentos precisam estar aptos para rodá-los novamente, validando novos resultados de outros estudos e comparando com os originais. Assim, nas ciências computacionais e naquelas que utilizam o ferramental quantitativo empírico, a pesquisa é replicável se outros pesquisadores podem seguir os mesmos

procedimentos originalmente utilizados, com os mesmos dados e executando o código usado para a análise. É preciso apreender as ferramentas para que a pesquisa seja reprodutível, sendo o resultado um processo de pesquisa mais efetivo e, em última instância, mais fácil, por agregar hábitos de trabalho incrementados, ao mesmo tempo que torna a pesquisa acessível para outros ([GANDRUD, 2013](#)).

A linguagem estatística do R, juntamente com o console RStudio, fornecem, portanto, o ferramental necessário para a construção de pesquisas reprodutíveis de projetos inteiros, partindo da coleta dos dados ou sua inclusão posterior, passando pela análise estatística e chegando à conseqüente apresentação dos resultados encontrados.

O Projeto Software R

O presente projeto de extensão caracterizou-se, no primeiro momento, pela criação e disponibilização de apostilas sobre o R (<https://softwarelivrer.wordpress.com/>), em formato digital. Foram realizados estudos e manipulações do *software* utilizado, buscando-se o desenvolvimento do saber pessoal em relação ao programa adotado. Pesquisaram-se materiais bibliográficos e disseminou-se o conhecimento entre os colaboradores envolvidos. Esses são compostos por professores atuantes na graduação, que utilizam estatística, alunos da graduação e da pós-graduação.

Entre os módulos básico e avançado, 32 pessoas demonstraram interesse, chegando-se, porém, ao final do curso com 17 participantes (cumprindo os requisitos de entrega dos trabalhos e frequência nas oficinas), 59% do sexo feminino e 41% do masculino. Todos os extensionistas são graduandos da UFFS, 35% provenientes do curso de graduação em Administração, 35% de Engenharia Ambiental e Sanitária, 24% de Agronomia e 6% de Ciências Biológicas. Ao se inscreverem no projeto, os alunos declararam sua experiência com relação ao *software* do projeto; informaram ter baixo nível de conhecimento na utilização do *software*/linguagem R 29% e 71% ter uma experiência em nível médio.

Para o desenvolvimento das aulas, foram elaboradas apostilas dinâmicas, utilizando-se o RMarkdown (recurso de produção de texto disponibilizado pelo próprio *software* R), a fim de que os participantes interagissem ativamente e desenvolvessem um conhecimento inicial das interfaces do *software*RStudio. A administração das aulas ocorreu alternadamente entre os colaboradores do projeto – bolsistas, professores e alunos de pós-graduação.

As aulas subdividiram-se nos conteúdos: Módulo 1 - Introdução ao R; Módulo 2 - Estatística Descritiva; Módulo 3 - Inferência Estatística; Módulo 4 - Teste Qui-Quadrado; Módulo 5 - Modelos de Regressão. Como o entendimento dos conceitos estatísticos é primordial para a sua aplicação, fez-se necessário um embasamento inicial, para que fosse possível não só a manipulação e obtenção dos resultados, mas também a compreensão de cada um deles.

Com a finalidade de difundir o projeto à comunidade acadêmica e regional, divulgou-se a proposta de aulas através de panfletos, isto é, por meio físico e também com o uso de publicações *online* em redes sociais e em páginas administradas pela Universidade. O objetivo foi promover aqueles que participassem da capacitação do uso do *software* livre R para análise estatística de dados, como o próprio nome do projeto sugere.

Para nivelar o conhecimento sobre os principais recursos do RStudio – alguns não possuíam as habilidades básicas na sua utilização - e devido à realização da aula inicial para um número considerável de alunos, fez-se necessária a realização de uma breve

introdução sobre o *software* e seus recursos. Assim, surgiram dúvidas pontuais e dificuldades iniciais que foram esclarecidas ao longo das aulas e da prática das atividades, com a ajuda dos ministrantes.

Ao final do curso de extensão foi realizada uma enquete entre os participantes com relação aos conteúdos aprendidos. Os resultados mostraram que 63% concordaram totalmente que “alcançaram o estabelecimento de relações entre os conteúdos abordados no curso com outros conhecimentos abordados nos componentes curriculares de seu respectivo curso de graduação”, e 37% concordaram parcialmente com a assertiva acima. Isso demonstra que o fato de os professores conteudistas do projeto de extensão utilizarem o *software*RStudio, para suas aulas de graduação, contribuiu fortemente para a consolidação da utilização da ferramenta pelos alunos no projeto. Eles puderam assim dirimir dúvidas e desenvolver habilidades adicionais, as quais não tinham sido trabalhadas no ambiente de sua graduação contribuindo para o sucesso do projeto de extensão. Adicionalmente, o projeto de extensão aprimorou as habilidades dos alunos de graduação e pós-graduação em análise estatística, a ser utilizada nos trabalhos de conclusão de curso e nas dissertações.

Com relação aos desafios relatados pelos colaboradores, destacam-se as dificuldades de definir a melhor agenda para as aulas do projeto – mesmo que estas ocorram em horário diferente das aulas da graduação. Por outro lado, a definição do cronograma do curso passa pela definição da agenda de cada colaborador. Também, a desistência de alunos inscritos consome as vagas de outros interessados em participar dos encontros do projeto de extensão, uma vez que é limitado o número de inscritos, em razão da capacidade do laboratório de informática. Ainda, é um desafio a participação do público externo. Esses são problemas que devem ser observados por outros projetos de extensão, com o objetivo de aprimorar sua dinâmica.

Em suma, o objetivo de transmitir os conhecimentos e as habilidades necessárias sobre o *software* foi atendido, mais nitidamente verificado nas atividades encaminhadas aos participantes e retornadas por eles, sendo realizadas com o uso dos comandos do RStudio. A expectativa é de que os participantes tenham conhecimento suficiente para assimilar o aprendizado, e para enfrentar os problemas que surgem no decorrer da vida acadêmica e profissional, capacitando-se para um julgamento mais crítico em suas interpretações nas leituras de estudos científicos ou relatórios técnicos ([AGUIAR et al., 2016](#)). Sobretudo, o ganho de conhecimento pessoal e acadêmico é de suma importância, visto que poderá ser aplicado não só no decorrer do trabalho exposto, mas também em trabalhos futuros, os quais possam exigir a manipulação e análise estatística de dados, nas demandas das diversas áreas da ciência.

CONSIDERAÇÕES FINAIS

A extensão universitária no Brasil, como visto, possui um papel ímpar na divulgação dos conhecimentos gerados na academia para atender as demandas, cada vez mais prementes, de qualificação externa, da democratização, da transformação e da cooperação da universidade com a sociedade, bem como da disseminação dos saberes e ferramentas científicas. De igual forma, a extensão universitária faculta uma integração de forma mais completa com o seu público interno, representado pelos discentes, docentes, técnicos de administração e pesquisadores das mais diversas áreas do conhecimento.

O objetivo deste projeto de extensão foi levar o conhecimento teórico e prático da linguagem de programação R e de seu console RStudio, aplicados à análise estatística, para a comunidade interna e externa da UFFS Câmpus Cerro Largo. Acredita-se que este projeto tenha contribuído para melhorar as habilidades dos discentes nas análises estatísticas, mediante o uso da linguagem de programação livre e a apropriação, pelos estudantes, das informações necessárias para a continuidade do aprendizado das diversas técnicas avançadas de análises de dados. Em adição, o projeto proporcionou a elevação da interação dos conhecimentos dos docentes dos programas de graduação e pós-graduação com alunos e bolsistas, podendo a continuidade do projeto abranger, em outras edições, a disseminação do ferramental estatístico para novos públicos interessados, bem como diminuir as desistências. Isso faz com que a Universidade continue cumprindo seu papel na comunidade, qualificando o capital social local e promovendo o desenvolvimento.

Em virtude do que foi mencionado, os *softwares* estatísticos, sejam eles livres ou proprietários, são de grande valor para a realização de análises estatísticas de dados, visto que agregam mais praticidade ao desenvolvimento do trabalho. O projeto de extensão aqui proposto contribui com os interessados nesse tipo de análise, pois leva conhecimentos importantes à qualificação e ao avanço da ciência, com a correta utilização das análises estatísticas. Ao mesmo tempo, esse ferramental contribui para maior transparência nas pesquisas, incrementa o poder de análise, ajuda a disseminação dos achados científicos realizados na Universidade e eleva as capacidades individuais de utilização de programas e técnicas que, além de reconhecidas internacionalmente, tornaram-se referência no âmbito científico.

A disseminação do conhecimento através de trabalhos de extensão entre os discentes, docentes, e o envolvimento da comunidade em geral é um dos pilares das universidades brasileiras, posto que potencializa a consequente produção de conhecimento das pesquisas realizadas localmente pela UFFS Câmpus Cerro Largo. Assim, além da constante qualificação do capital social local, este projeto oportuniza a futura continuidade da utilização e aprendizagem de outras ferramentas livres em projetos de extensão, bem como contribui para a criação de uma cultura, nas aulas das graduações, de utilização de ferramentas mais avançadas e colaborativas de análise estatística.

Submetido em 26/08/17

Aceito em 12/08/18

REFERÊNCIAS

[AGUIAR, J. et al.](#) Software R:capacitação em análise estatística de dados utilizando um software livre. In: SEMINÁRIO DE ENSINO, PESQUISA E EXTENSÃO (SEPE), 6., 2016. Chapecó. **Anais**, n.1, 2016.

[AQUINO, J. A.](#) **R para cientistas sociais**. Ilhéus: Editus, 2014.

[AMARAL, F.](#) **Introdução à ciência de dados: mineração de dados e big data**. Rio de Janeiro: Alta Books, 2016.

[ARRUDA, M. A. N.](#) Políticas públicas de cultura e extensão universitária. **Revista de Cultura e Extensão USP**, São Paulo, v. 4, p. 9-14, 2010.

BEGLEY, C. G.; IOANNIDIS, J. P. A. Reproducibility in science. **Circulation Research**, Dallas, v. 116, n. 1, p. 116-126, 2015.

BUVINICH, M. J. R.; AMORIM, J. M. Sistema de indicadores para o monitoramento e avaliação das ações de extensão: o caso da UFPB. **Revista Ciência em Extensão**, São Paulo, v. 9, n. 1, p. 9-34, 2013.

CORREA, S. **Probabilidade e estatística**. Belo Horizonte: PUC Minas Gerais, 2003.

FERREIRA, P. G. C. **Análise de séries temporais em R: curso introdutório**. Rio de Janeiro: Elsevier, 2018.

GANDRUD, C. **Reproducible research with R and R Studio**. 2nd ed. London: CRC The R Series, 2013.

GENTLEMAN, R.; IHAKA, R.; BATES, D. **The R project for statistical computing**. Viena: The R Foundation, 2018. Disponível em: [<https://www.r-project.org/>](https://www.r-project.org/). Acesso em: 14 jul. 2018.

JELIHOVSCHI, E. **Análise exploratória de dados usando o R**. Ilhéus: Editus, 2014.

KOENKER, R.; ZEILEIS, A. On reproducible econometric research. **Journal of Applied Econometrics**, Chichester, v. 24, n. 5, p. 833-847, 2009.

MAKEL, M. C.; PLUCKER, J. A. Facts are more important than novelty: replication in the education sciences. **Educational Researcher**, Washington, v. 43, n. 6, p. 304-316, 2014.

MARANDINO, M. Educação, ciência e extensão: a necessária promoção. **Revista de Cultura e Extensão USP**, São Paulo, v. 9, p. 89-100, 2013.

OLIVEIRA, F.; GOULART, P. M. Fases e faces da extensão universitária: rotas e concepções. **Revista Ciência em Extensão**, São Paulo, v. 11, n. 3, p. 8-27, 2015.

OLIVEIRA, F. L. B.; ALMEIDA JÚNIOR, J. J.; SILVA, M. L. P. Percepção dos acadêmicos em relação às dificuldades desenvolvimento de projetos de extensão universitária. **Revista Ciência em Extensão**, São Paulo, v. 12, n. 2, p. 18-25, 2016.

OROZCO, V. et al. How to make a pie: reproducible research for empirical economics and econometrics. **Toulouse School of Economics - Working Paper 933**, 2018.

PENG, R. D. Reproducible research in computing science. **Science**, New York, v. 334, n. 6060, p. 1226-1227, 2011.

RSTUDIO. **Take control of your R code**. Northern Ave: RSTUDIO, 2017. Disponível em: [<www.rstudio.com/products/RStudio/>](http://www.rstudio.com/products/RStudio/). Acesso em: 10 jun. 2017.

SLEUTJES, M. H. S. C. Refletindo sobre os três pilares de sustentação das universidades :

ensino-pesquisa-extensão. **Revista de Administração Pública**, Rio de Janeiro, v. 33, n. 3, p. 99-111, 1999.

[TIPPMANN, S.](#) Programing tools: adventures with R. **Nature**, London, 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. In: Seminário de Interação entre Universidade e Comunidade: a UFFS e o desenvolvimento regional, 1. 2010. **Anais eletrônicos...** Chapecó: UFFS, 2010. Disponível em: <<https://goo.gl/pY6N4s>>. Acesso em: 4 jul. 2017.