

EVOLUINDO NA APRENDIZAGEM DO AMBIENTE R: INFERÊNCIA ESTATÍSTICA E ESTATÍSTICA EXPERIMENTAL

SILVIA NAIANE JAPPE¹; ISADORA MOREIRA DA LUZ REAL²; LUÃ IEPSSEN³;
GISELDA MARIA PEREIRA⁴; ELISIA RODRIGUES CORRÊA⁵; ANA RITA DE
ASSUMÇÃO MAZZINI⁶

¹Universidade Federal de Pelotas – jappesilvia@gmail.com

²Universidade Federal de Pelotas – isadora.real18@hotmail.com

³Universidade Federal de Pelotas – iepsen.lua@gmail.com

⁴Universidade Federal de Pelotas – gmpereira08@gmail.com

⁵Universidade Federal de Pelotas – elisiarc@hotmail.com

⁶Universidade Federal de Pelotas – anarita.mazzini@gmail.com

1. INTRODUÇÃO

A estatística é uma importante ferramenta usada para a compreensão da ciência e suas especificidades visando especialmente a busca de soluções para coletar, organizar, analisar e compreender dados em cada área do conhecimento. Isto é possível a partir de um conjunto de métodos que possibilitam tirar conclusões e tomar decisões válidas a respeito de um conjunto de dados (COSTA NETO, 2002).

Em 2019, teve início o projeto de ensino “Explorando R como ferramenta auxiliar nos conteúdos de estatística”, constituído por uma equipe multidisciplinar de docentes do Departamento de Matemática e Estatística do Instituto de Física e Matemática (DME/IFM) e acadêmicos dos cursos de graduação em Agronomia, Ciências Biológicas, Química de Alimentos, Matemática e Medicina Veterinária da Universidade Federal de Pelotas (UFPEL). O escopo de atuação desse grupo de estudos é centrado na introdução do ambiente R (R CORE TEAM, 2021) para ampliar as habilidades estatísticas dos discentes. Os materiais didáticos são organizados condizentes com as aulas teóricas das disciplinas de Estatística. O grupo também busca a seleção e identificação de pacotes como ferramentas para resolução de exercícios relacionados aos conteúdos das disciplinas (NÓBREGA et al., 2019).

Além disso, em um processo contínuo de ensino-aprendizagem, foi realizada, com sucesso, a primeira oficina piloto presencial do projeto, em dezembro de 2019, conforme descrito por JAPPE et al. (2020). Nessa oficina houve a participação de discentes de diferentes cursos de graduação da UFPEL, com interesse na aprendizagem de um ambiente computacional com diversas funcionalidades estatísticas.

Devido a pandemia da COVID-19 e a impossibilidade da realização de atividades presenciais, o grupo adaptou-se ao novo formato (ALVES et al., 2020), adicionando ao seu itinerário a realização de oficinas didáticas remotas no RStudio. Essas oficinas foram formuladas e postas em prática com uma primeira edição em 2020 e uma segunda edição em 2021. Na primeira foram abordados conteúdos de estatística descritiva e na segunda, adicionados tópicos de inferência estatística.

A partir da demanda dos alunos participantes das oficinas remotas já realizadas, aliado ao propósito de ampliar o número de estudantes com acesso ao ambiente computacional, o projeto pretende realizar duas novas oficinas no ano de 2021. Diante disso, o objetivo do presente trabalho é apresentar a construção, organização, previsão e aplicabilidade das próximas ações de oficinas didáticas

remotas no RStudio, evoluindo da estatística descritiva para a inferência estatística e estatística experimental.

2. METODOLOGIA

Desde março de 2020, com a paralisação das atividades acadêmicas presenciais na UFPEL, foi preciso reprogramar as ações do projeto, através de reuniões semanais do grupo de estudos por videoconferência e a realização de oficinas didáticas remotas no ambiente R.

Em 2020, a equipe optou por utilizar a interface RStudio, para realização das oficinas remotas, para isso desenvolveu e adaptou bancos de dados no excel para realizar as aplicações estatísticas. Os bancos objetivam a aplicabilidade em diferentes áreas do conhecimento. Para as próximas oficinas estão sendo adaptados novos exemplos de aplicação compostos por variáveis quantitativas e qualitativas.

A equipe está organizando e preparando os novos materiais que incluem a seleção de novos pacotes para a realização de duas novas oficinas no segundo semestre de 2021, as quais serão realizadas pelo *Google Meet* de forma síncrona. Tendo em vista o grande número de pacotes que abordam os conteúdos de estatística, a equipe seleciona os que melhor se aproximam da forma trabalhada em sala de aula. A primeira oficina será com as aplicações de inferência estatística, vistas nas disciplinas de estatística básica do DME da UFPEL, com previsão de ocorrência para o final do mês de setembro. A segunda oficina será com aplicações de estatística experimental, abordando os conteúdos trabalhados na disciplina de Estatística Experimental I, ofertada pelo DME para o curso de Agronomia, com previsão de ocorrer na primeira semana de novembro.

3. RESULTADOS E DISCUSSÃO

Até o presente momento a equipe desenvolveu quatro bancos de dados englobando diferentes áreas do conhecimento (administração, agronomia, biologia e química), visto que a oficina atende alunos de diferentes cursos de graduação. Além dos bancos já construídos, está sendo realizada a busca e a construção de novos exemplos e aplicações que se ajustem às disciplinas de Estatística.

O roteiro da oficina de estatística descritiva já possui os seguintes tópicos de inferência estatística: teste F para homogeneidade de variâncias e teste t para duas amostras independentes, pressupondo distribuição normal e variâncias homogêneas ou heterogêneas. Estão sendo implementados no roteiro de inferência estatística: intervalos de confiança (para média, variância e proporção), testes de médias, variância, proporção, qui-quadrado, correlação e regressão.

Em relação aos testes de hipóteses, já foram desenvolvidos os procedimentos tanto para os testes paramétricos quanto para os não paramétricos. Junto aos testes paramétricos, os intervalos de confiança são dados automaticamente, sendo calculados dentro dos testes em si. Ao estudar o teste Z para uma e duas proporções, encontramos duas formas de realizá-lo: utilizando uma função com os próprios comandos do R ou através do uso do pacote *distributions3* (HAYES et al., 2021), pois este possibilita operar e criar distribuições de probabilidade.

Para o teste de variância para uma população foi selecionado o pacote *EnvStats* (MILLARD et al., 2021). Na realização do teste t pareado, foi necessário ajuste dos dados para obtenção de medidas anteriores e posteriores de um mesmo

indivíduo, possibilitando a análise. Nas análises de correlação e regressão, optou-se por utilizar o pacote *Agricolae* (MENDIBURU, 2021), já utilizado nas oficinas anteriores.

Em relação aos testes não paramétricos, estão inclusos o teste de qui-quadrado de aderência para frequências esperadas diferentes e a tabela de contingência, com teste de independência. Para essas análises, verificou-se que sua realização independe da utilização de pacotes específicos.

Quanto à estatística experimental, serão abordados os seguintes tópicos: Delineamento Inteiramente Casualizado (DIC) e Delineamento em Blocos Casualizados (DBC); ensaios fatoriais e testes de comparações múltiplas: Tukey, Duncan, t e SNK. Para realizar as aplicações de estatística experimental foi selecionado o pacote *ExpDes.pt*. (FERREIRA et al., 2021). Outra possibilidade encontrada para realizar aplicações de estatística experimental é o pacote *Agricolae*, contudo a apresentação dos resultados é menos organizada em comparação ao pacote *ExpDes.pt*. Quanto à estatística experimental, o grupo de estudos ainda está evoluindo na busca das aplicações para esquema fatorial e testes de comparações múltiplas.

É importante ressaltar que as aplicações descritas são umas das possíveis formas de realização das análises para obtenção dos resultados.

4. CONCLUSÕES

A realização das duas oficinas está programada para o segundo semestre deste ano. Os pacotes necessários para fazer as análises já foram selecionados e os bancos de dados seguem em construção. O conteúdo para a oficina de inferência estatística encontra-se organizado. O conteúdo para a oficina de estatística experimental segue evoluindo.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, K. S. et al. Ações no Cenário Remoto do Projeto de Ensino “Explorando R Como Ferramenta Auxiliar nos Conteúdos de Estatística”. In: **CONGRESSO DE ENSINO DE GRADUAÇÃO**, 6., Pelotas, 2020. ANAIS CEG 2020, Área Ciências Exatas e da Terra. UFPEL, 2020.

COSTA NETO, P. L. O. **Estatística**. 3ª ed. São Paulo: Blucher, 2002.

FERREIRA, E.B.; CAVALCANTI, P. P.; NOGUEIRA, D.A. (2021). **ExpDes.pt**: Pacote Experimental Designs (Portugues). R package version 1.2.1. Acessado em 23 jul. 2021. Online. Disponível em: <https://cran.r-project.org/web/packages/ExpDes.pt/index.html>

HAYES, A. et al. (2021). **Distributions3**: Probability distributions as S3 objects. R package version 0.1.1. Acessado em 23 jul. 2021. Online. Disponível em: <https://cran.r-project.org/web/packages/distributions3/index.html>

JAPPE, S. N. et al. Relato da Oficina Piloto Estatística Descritiva Usando o R. In: **CONGRESSO DE ENSINO DE GRADUAÇÃO**, 6., Pelotas, 2020. ANAIS CEG 2020, Área Ciências Exatas e da Terra. UFPEL, 2020.

MENDIBURU, F. (2021). **Agricolae**: Statistical Procedures for Agricultural Research. R package version 1,3-5. Acessado em 22 jul. 2021. Online. Disponível em: <https://CRAN.R-project.org/package=agricolae>

MILLARD, S. P.; KOWARIK, A. (2021). **EnvStats**: Package for Environmental Statistics, Including US EPA Guidance. R package version 2.4.0. Acessado em 23 jul. 2021. Online. Disponível em: <https://cran.r-project.org/web/packages/EnvStats/index.html>

NÓBREGA, E. S. B. et al. Explorando o R como Ferramenta Auxiliar nos Conteúdos de Estatística. In: **CONGRESSO DE ENSINO DE GRADUAÇÃO**, 5., Pelotas, 2019. ANAIS CEG 2019, Área Ciências Exatas e da Terra. UFPEL, 2019.

R Core Team (2021). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Acessado em 22 jul. 2021. Disponível em: <https://www.R-project.org/>.