

EXPLORANDO R COMO FERRAMENTA AUXILIAR NOS CONTEÚDOS DE ESTATÍSTICA

EDUARDO DE SÁ BUENO NÓBREGA¹; **ELISIA RODRIGUES CORREA²**;
GISELDA MARIA PEREIRA³; **ANA RITA DE ASSUMPÇÃO MAZZINI⁴**

¹*Universidade Federal de Pelotas – eduardosbnobrega@gmail.com*

²*Universidade Federal de Pelotas – elisiarc@hotmail.com*

³*Universidade Federal de Pelotas – gmpereira08@gmail.com*

⁴*Universidade Federal de Pelotas – anarita.mazzini@gmail.com*

1. INTRODUÇÃO

A estatística como um todo vem crescendo e expandindo suas aplicações, seja no ramo da ciência, política ou economia. Segundo SOUSA (2002), os resultados obtidos com a aplicação dos métodos estatísticos na resolução de problemas dos diversos domínios do conhecimento, aliados à evolução tecnológica dos últimos anos, fizeram com que os conhecimentos estatísticos se tornassem indispensáveis em todos os domínios.

Hoje em dia a população está exposta a uma grande quantidade de informações sobre dados estatísticos, seja através da mídia ou das redes sociais. Estes dados, se interpretados de maneira inadequada, podem afetar de forma significativa as decisões tomadas por cada cidadão (CAZORLA, 2002). Neste contexto, o ensino de graduação tem papel fundamental na formação de profissionais com conhecimento na área da estatística, de forma a solucionar questões em suas áreas de atuação.

Na Universidade Federal de Pelotas, as disciplinas de estatística abrangem diversas áreas de ensino, tais como: ciências agrárias, exatas, humanas, biológicas, dentre outras. Nos diferentes cursos relacionados a estas áreas o principal objetivo das disciplinas de estatística são habilitar o estudante para compreensão da base conceitual e metodológica da estatística requerida no planejamento, análise de dados e interpretação dos resultados de pesquisa científica. Além disso, cabe ressaltar que o ensino da estatística é abordado de forma teórica na maioria das disciplinas ofertadas pelo Departamento de Matemática e Estatística da UFPEL.

É notório que um dos grandes impactos no desenvolvimento da estatística foi o surgimento dos recursos computacionais e o efeito que tiveram na prática desta disciplina. De acordo com IGNÁCIO (2010), a criação de computadores cada vez com maior capacidade de processamento foi decisiva e fez com que a estatística se tornasse mais acessível aos pesquisadores dos diferentes campos de atuação. Segundo AGUIAR et al. (2016) com avanço da tecnologia, e a grande utilização de softwares estatísticos para análise de dados, não são mais utilizadas planilhas manuais, que dispendem tempo para a produção de resultados. Dentre estes softwares, o R (CORE TEAM, 2019) tem sido muito utilizado, pois embora possua uma linguagem de programação, onde os comandos devem ser aprendidos pelo usuário, o domínio da técnica possibilita que o acadêmico trabalhe com elevada produtividade e eficácia (AQUINO, 2014).

Um dos desafios das universidades em geral é a compra anual da licença de softwares estatísticos, dificultando a utilização destes recursos em sala de aula. Desta forma, o ambiente R (CORE TEAM, 2019) é um excelente recurso a ser utilizado como ferramenta que complementa e fortalece os conceitos teóricos da área da estatística, permitindo o acesso gratuito para estudantes, professores, pesquisadores, comunidades acadêmicas e científicas.

Sendo assim, foi constituído um grupo de estudos com o objetivo de explorar o ambiente R (CORE TEAM, 2019), identificando e selecionando pacotes e rotinas que possam ser utilizadas na resolução de exercícios em sala de aula e extraclasse, visando motivar os alunos à prática de um recurso computacional durante sua vida acadêmica.

2. METODOLOGIA

No mês de maio de 2019, foram iniciadas as atividades do grupo de estudo. Para melhor aproveitamento e desenvolvimento das atividades propostas, estabeleceram-se metas semanais, que envolveram a busca de materiais de suporte ao ambiente R (CORE TEAM, 2019), referências e pacotes relevantes aos temas trabalhados além de aperfeiçoamentos na interface gráfica do software.

Para tanto, os seguintes conteúdos foram definidos: gráficos de colunas verticais, colunas horizontais, setores, diagrama de ramo e folhas, gráfico de caixas, tabelas de distribuição de frequências, histogramas, polígonos de frequências, gráfico de séries temporais e medidas descritivas. Estes materiais foram compilados e organizados no R (CORE TEAM, 2019), que consiste em um conjunto integrado de facilidades de software para manipulação de dados, cálculo e visualização gráfica (BEASLEY, 2004).

3. RESULTADOS E DISCUSSÃO

O *default* do R (CORE TEAM, 2019) foi utilizado para a obtenção das principais medidas de resumo: média aritmética, variância, desvio padrão, coeficiente de variação e quantis. Além das medidas alguns gráficos também foram produzidos pelo *default*, tais como: gráficos de colunas verticais, colunas horizontais, setores e diagrama de ramo e folhas. As funções disponíveis no *default* são obtidas com comandos simples o que facilita a utilização. A função *summary*, apresenta o resumo de algumas medidas descritivas. Para os quantis especificamente, o R (CORE TEAM, 2019) oferece alguns métodos de obtenção que permitem o usuário escolher o que julgar mais apropriado.

Na construção dos gráficos de setores, barras e *boxplot* foi utilizado o *default* do R (CORE TEAM, 2019) com os comandos *pie()* e *barplot()*, respectivamente. Entretanto, os gráficos produzidos pelo *default* muitas vezes não mostram todas informações necessárias para compreensão da figura, tais como: nome da variável, unidades e escalas dos eixos (Figura 1a). Para obtenção de gráficos com informações mais completas e melhor visualização é necessário o acréscimo de comandos conforme os apresentados para construção da Figura 1b.

Para histogramas (Figura 2a e 2b), polígono de frequência e tabelas de frequência foi empregado o pacote *fdth* (FARIA et al., 2019), e para séries temporais o pacote *lattice* (SARKAR; DEEPAYAN, 2008). Conforme GARFIELD (2002), é necessário auxiliar os estudantes a desenvolver seu raciocínio e o uso de softwares permitem fazer simulações e testar as modificações que ocorrem quando se trabalha com diferentes amostras.

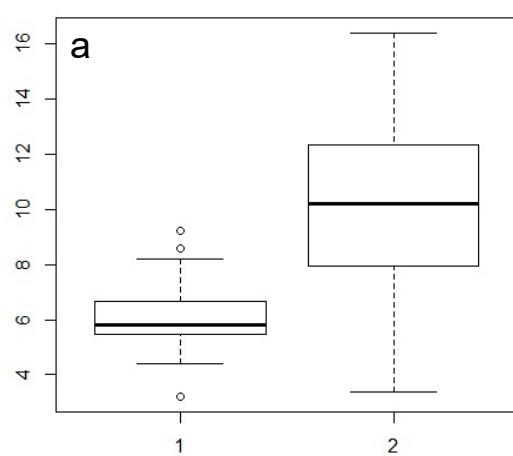


Figura 1 a – *Boxplot* gerado pelo comando do R: `boxplot(x, y)`.

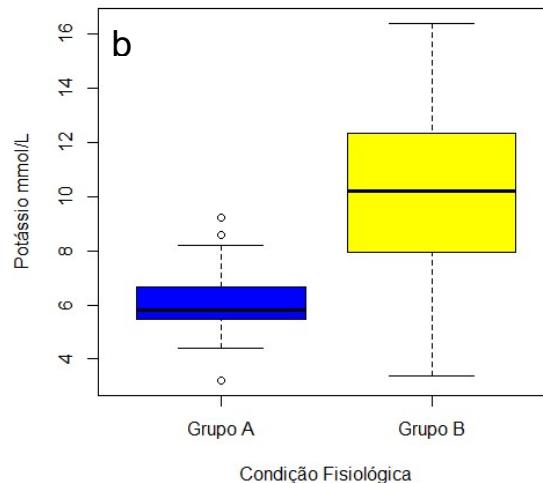


Figura 1 b – *Boxplot* gerado pelo comando do R: `boxplot(x, y, xlab="Condição Fisiológica", names=c("Grupo A", "Grupo B"), ylab="Potássio mmol/L", col=c("blue", "yellow"))`.

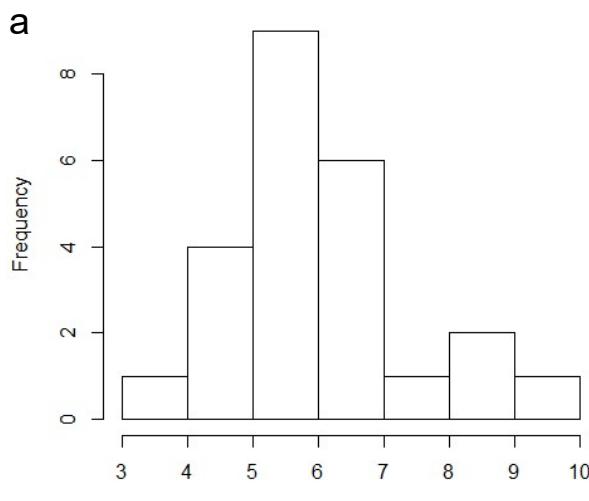


Figura 2 a – Histograma gerado pelo comando do R: `hist(x)`.

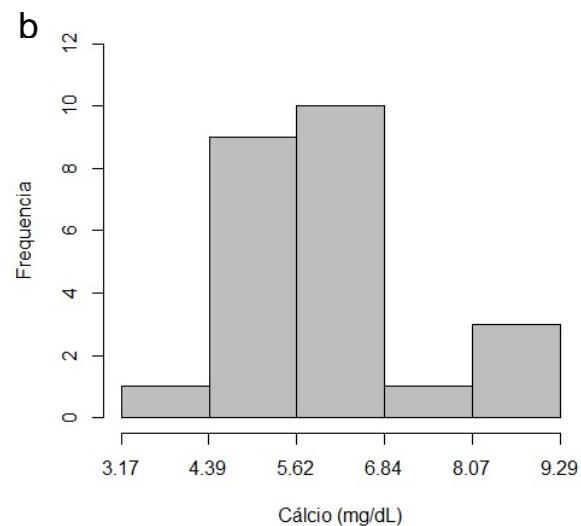


Figura 2 b – Histograma gerado pelo comando do R: `plot(fdt(x, 5), xlab="Cálcio (mg/dL)", ylab="Frequencia", col=8)`.

O software R (CORE TEAM, 2019) apresenta uma interface pouco intuitiva. Por esta razão, o grupo de estudos, a partir das funções e pacotes identificados, está elaborando um material escrito com o passo a passo de tudo o que foi construído até o presente momento com o intuito de apresentar para o usuário (aluno de graduação) opções simples e diretas para realização de tarefas. Conforme SILVA; SAMÁ (2017), a utilização de propostas pedagógicas que contemplem materiais concretos e ferramentas digitais podem potencializar o processo de ensino aprendizagem da Estatística.

4. CONCLUSÕES

Foram identificados e selecionados os pacotes *fdth* (FARIA et al., 2019) e *lattice* (SARKAR; DEEPAYAN, 2008), além do *default* do R (CORE TEAM, 2019). O grupo de estudo permitiu a troca de informações entre docentes e estudante, possibilitando a seleção e compilação dos diversos materiais encontrados, facilitando a elaboração de material didático a ser trabalhado nas disciplinas de estatística.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- AGUIAR, J. et al. Software R: Capacitação em análise estatística de dados utilizando um software livre. In: **SEMINÁRIO DE ENSINO, PESQUISA E EXTENSÃO (SEPE)**. Chapecó. Anais, n.1, 2016.
- AQUINO, J. A. R para cientistas sociais. Ilhéus: Editus, 2014.
- BEASLEY, C.R. Bioestatística usando R. **Universidade Federal do Pará – Campus de Bragança**, Bragança, 2004.
- CAZORLA, I.M. **A relação entre a habilidade viso-pictórica e o domínio de conceitos estatísticos na leitura de gráficos**. 2002. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade Estadual de Campinas.
- FARIA, J.C; Jelihovschi, E.G. & Allaman, I. B. *fdth: Frequency Distribution Tables, Histograms and Polygons*. UESC, Bahia, Brasil, 2019.
- GARFIELD, J. The challenge of developing statistical reasoning. In: **Journal of statistic education**, v.10, n.3. Acessado em 12 de setembro de 2019. Disponível em: www.amstat.org/publications/jse/v10n3/chance.html.
- IGNÁCIO, S.A. Importância da Estatística para o Processo de Conhecimento e Tomada de Decisão. **Instituto Paranaense de Desenvolvimento Econômico e Social - IPARDES**, Curitiba, n. 6, 2010.
- R CORE TEAM. **R: A languagean denvironment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2019.
- SARKAR, DEEPAYAN. *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5, 2008.
- SILVA, C. R.; SAMÁ, S. Material concreto e tecnologias digitais: possibilidades para o ensino e a aprendizagem da Estatística. **Polyphonía**, v. 28/2, 2017.
- SOUSA, O. **Investigações Estatísticas no segundo ciclo do ensino básico**. 2002. Dissertação de Mestrado, Universidade de Lisboa.