

Ferramentas Práticas para o Ensino de Estatística

Darah Aparecida Pires Moreira ¹, Fernando de Souza Bastos²
Departamento de Estatística, UFV, Viçosa, MG

Resumo. Destacamos alguns dos resultados do projeto “Pesquisa e Criação de Ferramentas Práticas para o Ensino de Estatística”, aprovado no Edital de Bolsas de Iniciação a Docência da UFV de 2023, com apoio do CNPq, que teve como objetivo desenvolver ferramentas práticas para o ensino de Estatística básica em cursos de graduação, utilizando o *software* R e pacotes como *tidyverse*, *shiny-dashboard*, *ggplot* e *learnR*, entre outras ferramentas tecnológicas. A Estatística é uma disciplina fundamental em diversas áreas do conhecimento, e o aprendizado efetivo dos conceitos estatísticos pode ser desafiador para os estudantes. Nesse sentido, a utilização de ferramentas computacionais modernas pode auxiliar no ensino e na compreensão dos conceitos estatísticos, tornando o aprendizado mais prático e interativo. Acreditamos que as ferramentas criadas podem contribuir para tornar o aprendizado mais envolvente e efetivo, proporcionando aos estudantes uma compreensão mais profunda dos conceitos estatísticos e fortalecendo suas habilidades analíticas.

Palavras-chave. *software* R, Educação Estatística, Ensino, Graduação.

1 Introdução

O ensino efetivo de Estatística básica em cursos de graduação desempenha um papel crucial no desenvolvimento de habilidades analíticas e na formação de profissionais capacitados para lidar com a crescente quantidade de dados disponíveis em diversas áreas do conhecimento. Nesse contexto, as ferramentas práticas para o ensino de Estatística têm o potencial de transformar a forma como os estudantes interagem com os conceitos estatísticos, tornando-os mais acessíveis e aplicáveis ao mundo real.

A criação de ferramentas práticas não apenas estimula o interesse dos estudantes e aprimora o aprendizado dos conceitos estatísticos, mas também contribui para a formação de profissionais mais preparados para enfrentar os desafios do mundo moderno. Além disso, tais ferramentas práticas e computacionais podem gerar um significativo impacto científico, pois a disponibilidade dos produtos deste projeto permitem que pesquisadores e professores desenvolvam novas abordagens de ensino, explorem conceitos estatísticos complexos e realizem experimentos virtuais, ampliando assim o conhecimento científico e possibilitando a disseminação de práticas inovadoras.

Dentre os conteúdos que podem ser abordados, destacam-se a estatística descritiva, que permite a organização e resumo dos dados; as distribuições de probabilidade, que são fundamentais para a compreensão da incerteza e da aleatoriedade; os intervalos de confiança e os testes de hipóteses, que auxiliam na tomada de decisões baseadas em evidências estatísticas, entre outros conceitos. Através do uso do *software* R, do GeoGebra e de pacotes como *ggplot2* [5], *tidyverse* [6], *shiny* [2], *flexdashboard* [4], *learnr* [1], entre outros, é possível proporcionar uma experiência de aprendizado interativa e dinâmica.

O *software* R [3] é amplamente utilizado na área de Estatística e oferece uma vasta gama de pacotes e funcionalidades para análise de dados. Neste trabalho destacamos alguns resultados e os

¹Discente do Curso de Matemática da UFV Email: darah.moreira@ufv.br

²Docente do Departamento de Estatística Email: fernando.bastos@ufv.br

objetivos do projeto “Pesquisa e Criação de Ferramentas Práticas para o Ensino de Estatística”, aprovado no Edital de Bolsas Institucionais de Iniciação a Científica da UFV de 2023, com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), em que exploramos o potencial do R, combinado com os recursos de alguns pacotes, para facilitar o ensino de Estatística básica para a graduação. Através destes pacotes, podemos fornecer uma interface mais amigável para a manipulação e visualização de dados, simplificando a introdução de conceitos estatísticos.

2 Metodologia

O projeto foi concebido com a proposta de desenvolvimento em etapas, com o objetivo de criar e implementar ferramentas práticas para o ensino de Estatística Básica no nível de graduação. A metodologia inicialmente planejada foi estruturada em cinco fases principais:

Levantamento de materiais: Inicialmente, foi realizada uma discussão entre a bolsista do projeto e o orientador e também professor da disciplina de Estatística I (EST 106), ofertada para alunos dos cursos de Ciências Exatas da UFV, sobre requisitos e necessidades específicas dos estudantes e professores na área de Estatística básica. Pesquisamos, sites de professores no Brasil e no mundo de diversas instituições e elaboramos uma lista com roteiros de aulas, apostilas, livros utilizados na graduação, *applets*, aplicativos e ferramentas práticas para identificar materiais estatísticos relevantes e as possibilidades de uso, transformação de materiais ou criação de novas ferramentas para o ensino.

Desenvolvimento das ferramentas: Com base nos materiais identificados, as ferramentas práticas seriam desenvolvidas utilizando o *software* R, o GeoGebra e pacotes do R. A ideia era criar interfaces interativas e dinâmicas, com recursos de manipulação de dados, visualizações gráficas, cálculos estatísticos e exemplos práticos que ilustrassem os conceitos estatísticos abordados.

Testes e avaliação: As ferramentas desenvolvidas seriam submetidas a avaliação por parte de professores e estudantes por meio de atividades práticas, como a resolução de problemas, a exploração de dados reais e a realização de experimentos virtuais, a fim de verificar a eficácia das ferramentas no processo de aprendizado e a compreensão dos conceitos estatísticos pelos estudantes.

Ajustes e aprimoramentos: Com base nos resultados dos testes e na retroalimentação recebida, seriam realizados ajustes e aprimoramentos nas ferramentas desenvolvidas considerando sugestões dos usuários, correção de eventuais falhas e otimização da usabilidade e da experiência do usuário.

Disseminação e documentação: Por fim, as ferramentas práticas seriam documentadas e disponibilizadas para uso por parte dos professores e estudantes.

No entanto, devido a greve dos professores do ano de 2024, nosso projeto foi reavaliado e readaptado, após a primeira etapa metodológica de pesquisa de materiais, resolvemos criar um código em linguagem R que automaticamente faz a leitura e a criação de uma apresentação em formato *slidy*, formato de apresentação criado em *Rmarkdown*, para a exemplificação de tabelas e gráficos construídos a partir da aplicação de um *google forms* nas turmas de Estatística e uma apostila interativa criada em linguagem *markdown* utilizando o pacote *learnr*.

O formulário contém perguntas que não identificam o estudante, mas foi elaborado de forma a ressaltar algumas das dificuldades da limpeza e organização de um banco de dados. Foi perguntado, por exemplo, qual era a cidade de residência do respondente, alguns respondiam Belo Horizonte - MG, outros respondiam BH, outros BH-MG, outros BH-mg, entre outras formas. Portanto, o código que faz a criação dos slides para apresentação dos resultados já corrige alguns erros na leitura dos dados. Ele faz a limpeza, organiza em caixa alta, faz algumas transformações dos dados e cria uma apresentação dos resultados, conforme pode ser visualizado no seguinte endereço <https://rpubs.com/fsbmat/CensoEst105>. Esse endereço apresenta os dados observados de uma turma de Iniciação Estatística (EST 105) do segundo semestre de 2024.

Após a criação dos códigos para a leitura do banco de dados, começamos a desenvolver a apostila interativa do curso de Estatística I baseada na apostila do curso, de autoria do professor Luiz Alexandre Peternelli, do Departamento de Estatística da UFV, com o pacote `learnr` [1]. Nosso objetivo aqui foi de disponibilizar aos estudantes um material útil para o estudo de Estatística acompanhado do estudo da linguagem R, fundamental para o desenvolvimento de análises Estatísticas e visualização de tabelas e gráficos.

O material, ainda está recendo alguns ajustes, mas já está disponível no seguinte endereço <https://pibicest.shinyapps.io/ApostilaInterativa/>.

3 Resultados e Discussão

A partir da análise das respostas coletadas por meio do formulário elaborado no Google Forms, identificamos algumas dificuldades na criação de tabelas e gráficos, que merecem ser discutidas com os estudantes. Entre essas dificuldades, destacam-se a aplicação adequada do questionário, os desafios na avaliação de questões abertas, a inconsistência na padronização das respostas e a necessidade de realizar uma limpeza e organização eficaz dos dados. Observamos, ainda, a frequente classificação incorreta de respostas que deveriam ser quantitativas como qualitativas. Para tornar o ambiente de discussão mais dinâmico e menos rígido, foram incluídas perguntas que proporcionavam uma leveza à apresentação dos resultados. Um exemplo disso foi a indagação sobre a expectativa de rendimento na disciplina, cujas opções de resposta eram: “SIM”, “NÃO” e “Dependerá do Professor da Disciplina”. As respostas dos alunos dos cursos de Ciências Exatas podem ser observadas na Figura 1.

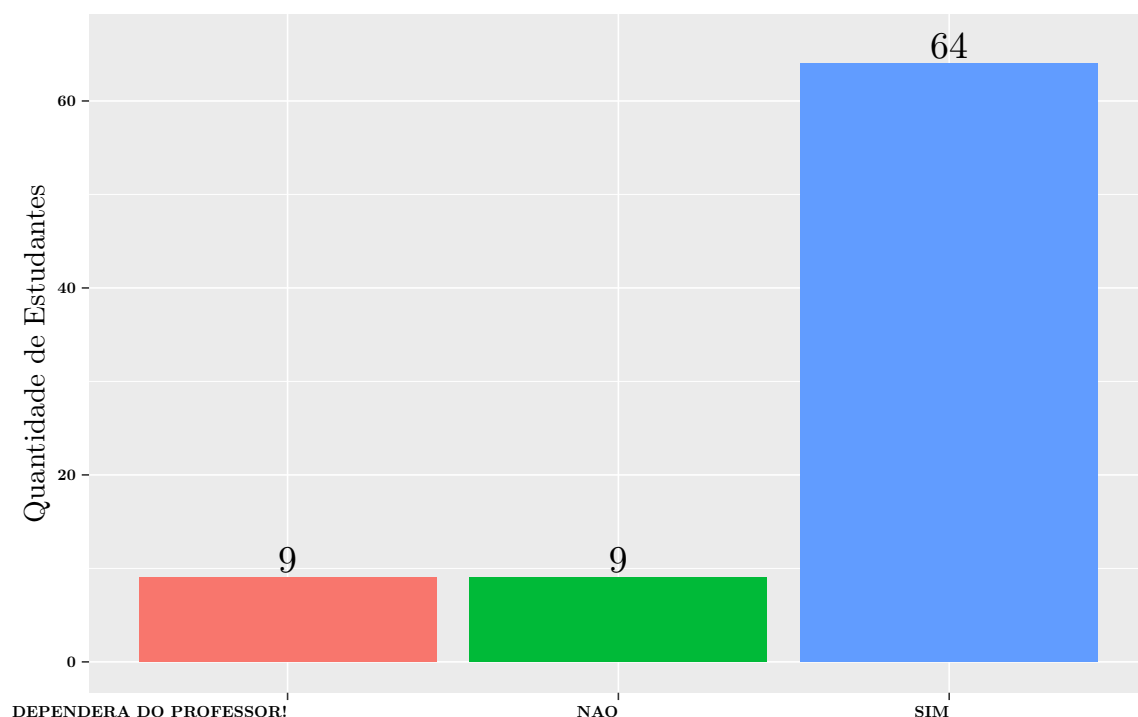


Figura 1: Você Acredita que Terá um Bom Rendimento na Disciplina?. **Fonte:** Elaboração Própria.

Observou-se que, mesmo entre os alunos matriculados em cursos da área de Ciências Exatas, uma parcela significativa respondeu que seu rendimento na disciplina dependeria do professor, enquanto outros indicaram que não teriam um bom desempenho. Essas respostas suscitaram debates em sala de aula, levando a diferentes interpretações. Alguns estudantes sugeriram que tais respostas poderiam refletir a insatisfação de alunos que consideram trocar de curso ou que ingressaram na área com a expectativa de possuir facilidade com disciplinas de exatas, mas posteriormente perceberam que não tinham. Além disso, em tom descontraído, alguns mencionaram que reprovar ao menos uma vez na disciplina poderia ser visto como parte da experiência acadêmica.

Agora observem a resposta a pergunta “Você tem Facilidade com Disciplinas de Exatas?”, apresentada aos alunos de Estatística I (EST 106), disciplina ofertada a cursos de Engenharia, Matemática e Ciências Econômicas. A resposta a essa última pergunta levou a uma boa discussão em sala de aula e agregou muito na condução da disciplina e na apresentação dos resultados. É importante destacar que esses resultados são obtidos no início da disciplina, o que nos leva a questionar as motivações por trás das respostas dos alunos. A pesquisa realizada com estudantes de cursos de ciências exatas revelou que 43,9% afirmaram não ter facilidade com disciplinas da área. Esse dado pode refletir, em parte, as dificuldades reais que alguns alunos enfrentam no aprendizado dessas matérias. No entanto, também é possível que uma parcela dos respondentes esteja tentando sensibilizar os professores para receberem uma avaliação mais branda, considerando as dificuldades iniciais do curso.

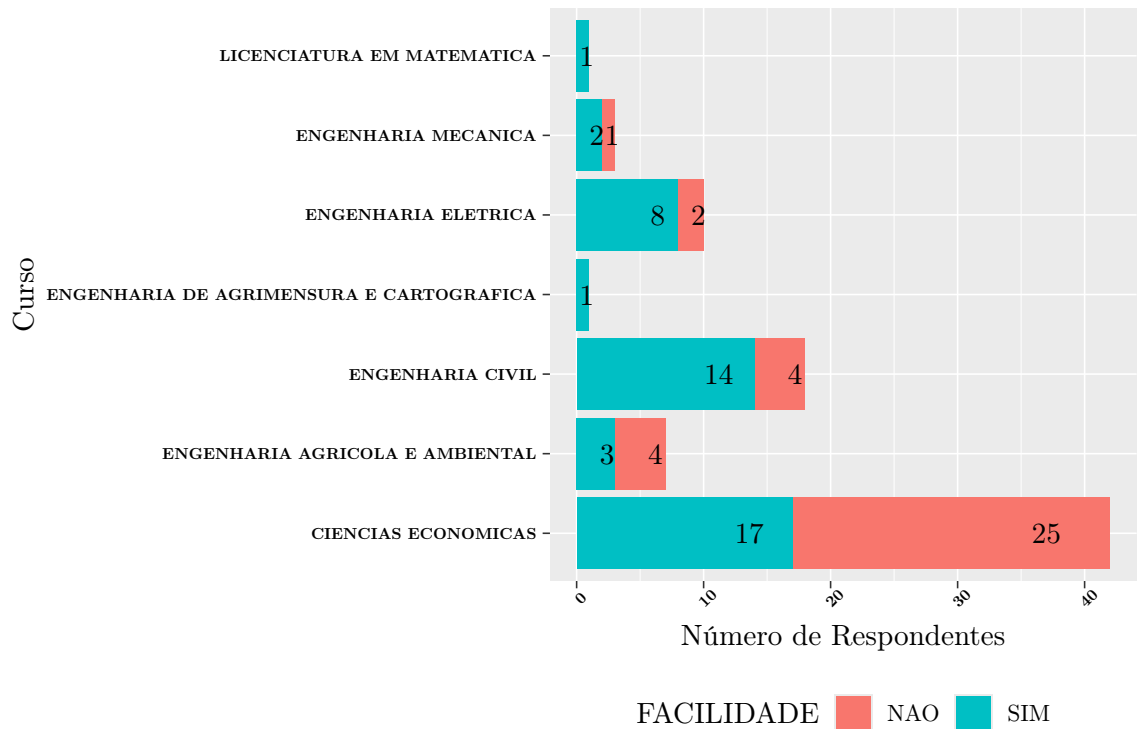


Figura 2: Você Tem Facilidade com Disciplinas de Exatas?. **Fonte:** Elaboração Própria.

Por fim, apresentamos dois exercícios contidos na apostila interativa, o primeiro ilustrado na Figura 3, onde o estudante é solicitado a compilar o código e verificar os resultados gerados. O pacote *learnr* do R oferece um ambiente interativo que permite ao aluno não apenas compilar o código diretamente na plataforma, mas também modificá-lo, acrescentar novas linhas de código e experimentar diferentes abordagens. Essa interatividade proporciona uma experiência de aprendizado muito mais rica, na qual o estudante tem a oportunidade de explorar conceitos de Estatística e de programação de forma prática e autônoma. O *learnr* é uma ferramenta extremamente poderosa ao possibilitar o desenvolvimento de múltiplas habilidades, promovendo tanto a compreensão teórica quanto a aplicabilidade dos conhecimentos adquiridos, ampliando significativamente as oportunidades de aprendizado ativo e personalizado.

```

Código R  Reiniciar  Executar código
1 # Vetores
2
3 x <- c(90, 95, 97, 98, 100, 60)
4 y <- c(60, 70, 80, 60, 90, 75)
5
6 #Soma Simples
7 sum(x)
8
9 #Soma de Quadrados
10 sum(x^2)
11
12 #Quadrado da Soma
13 (sum(x))^2
14
15 #Soma de Produtos

```

Figura 3: Exercício sobre Somatório Presente na Apostila Interativa de Estatística I. **Fonte:** Elaboração Própria.

Na Figura 4, apresentamos o processo de leitura de um banco de dados e a criação de uma tabela de frequências no R, após a devida compilação do código. Para um aluno de graduação, entender como construir uma tabela de frequências em um *software* computacional é uma habilidade fundamental. Esse conhecimento não apenas facilita a organização e a interpretação de dados, mas também capacita o estudante a lidar com grandes volumes de informação de maneira eficiente. O uso de ferramentas como o R, que automatizam esses processos, permite que o aluno se concentre na análise crítica dos dados, aprimorando sua capacidade de tomada de decisões informadas. Além disso, o domínio de técnicas computacionais para a construção de tabelas é um diferencial no mercado de trabalho, onde a habilidade de manipular dados e realizar análises estatísticas é cada vez mais valorizada.

```

Código R  Reiniciar  Executar código
1 url <- "https://raw.githubusercontent.com/ufvest/ufvest.github.io/master/Aulas_EST105/Curso_R/Dados/CompanhiaMB_clear.txt"
2 tab2_1 <- fread(url, sep = "\t", header = FALSE)
3 colnames(tab2_1) <- c("Registro",
4                       "Estado_Civil",
5                       "Grau",
6                       "N_filhos",
7                       "Salario",
8                       "Id_Anos",
9                       "Id_Meses",
10                      "Procedencia")
11 ni<-table(tab2_1$Grau) # Calcula a tabela de frequências absolutas e armazena o resultado em 'mytab'
12 fi<-prop.table(ni) # Tabela de frequências relativas (f_i)
13 p_fi<-100*prop.table(ni) # Porcentagem (100 f_i)
14
15 # Adiciona linhas de total
16

```

Figura 4: Exercício sobre Tabela de Frequências Presente na Apostila Interativa de Estatística I. **Fonte:** Elaboração Própria.

4 Considerações Finais

Conclui-se que a criação de atividades práticas e visualizações gráficas interativas desempenha um papel crucial no ensino e aprendizado de Estatística Básica em qualquer nível. A compreensão de conceitos estatísticos abstratos se torna mais acessível quando acompanhada de ferramentas que possibilitam a experimentação e a exploração de dados em tempo real. Neste contexto, os produtos desenvolvidos ao longo deste trabalho reforçam essa abordagem. O código em R criado

para a rápida limpeza e organização das respostas de questionários do Google Docs, aliado à geração automática de tabelas e gráficos, contribui significativamente para o ensino de Estatística Descritiva, facilitando tanto o aprendizado quanto a prática dos estudantes. Além disso, a apostila interativa desenvolvida com o pacote `learnr` do R possibilita que os alunos não apenas aprendam os conceitos estatísticos, mas também desenvolvam habilidades na linguagem R, tornando o processo de aprendizagem mais dinâmico e alinhado às demandas modernas da área. Tais recursos demonstram o impacto positivo da integração entre tecnologia e educação, promovendo uma formação estatística mais sólida e aplicada.

Agradecimentos

O presente trabalho foi realizado com o apoio do CNPq por meio da disponibilidade de bolsas ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC/CNPq 2023-2024). Os autores agradecem, portanto, a Universidade Federal de Viçosa e ao CNPq pelo apoio.

Referências

- [1] Garrick Aden-Buie, Barret Schloerke, JJ Allaire e Alexander Rossell Hayes. **learnr: Interactive Tutorials for R**. R package version 0.11.5. 2023. URL: <https://CRAN.R-project.org/package=learnr>.
- [2] Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert e Barbara Borges. **shiny: Web Application Framework for R**. R package version 1.7.4. 2022. URL: <https://CRAN.R-project.org/package=shiny>.
- [3] R Core Team. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing. Vienna, Austria, 2023. URL: <https://www.R-project.org/>.
- [4] Carson Sievert, Richard Iannone, JJ Allaire e Barbara Borges. **flexdashboard: R Markdown Format for Flexible Dashboards**. R package version 0.6.1. 2023. URL: <https://CRAN.R-project.org/package=flexdashboard>.
- [5] Hadley Wickham. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: <https://ggplot2.tidyverse.org>.
- [6] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo e Hiroaki Yutani. “Welcome to the tidyverse”. Em: **Journal of Open Source Software** 4.43 (2019), p. 1686. DOI: 10.21105/joss.01686.