

doi 10.46943/X.CONEDU.2024.GT20.034

GESTÃO DE SAÚDE PÚBLICA BASEADA EM EVIDÊNCIAS: UMA INTERSEÇÃO ENTRE CIÊNCIA DE DADOS COM LINGUAGEM R, ESTATÍSTICA E COMUNICAÇÃO EM UM CURSO EAD NO ESTADO DO ESPÍRITO SANTO

Marcos Vinícius Carneiro Vital¹
Isaac Negretto Schrarstzhaupt²
Adelmo Inácio Bertolde³
Marcelo Alves de Souza Bragatte⁴

RESUMO

As tomadas de decisão relacionadas à políticas públicas para Saúde afetam sua efetividade prática, influenciando o impacto de epidemias e pandemias, por exemplo. Para que estas decisões sejam efetivas e façam bom uso dos recursos públicos, é fundamental serem baseadas em evidências - o que pode ser alcançado com boas práticas de análise de dados. Entretanto, nem sempre os servidores possuem conhecimentos adequados e/ou atualizados para conduzir estas análises. Neste trabalho, relatamos o planejamento, execução e resultados do curso “Capacitação em Análise de Dados para a Qualificação da Gestão em Saúde”, realizado em formato EAD para agentes públicos da saúde do Estado do Espírito Santo via ICEPi, bem como lições aprendidas no percurso. O curso foi executado a partir de videoaulas diárias, encontros síncronos regulares, tutoria em fórum e atividades autodirigidas, aten-

- 1 Professor do curso de Ciências Biológicas na Universidade Federal de Alagoas (UFAL) e no Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi), marcosvital@gmail.com;
- 2 Doutorando pelo Curso de Epidemiologia da Universidade de São Paulo (USP) e Tutor no Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi), isaacjesus@gmail.com;
- 3 Professor do curso de Estatística da Universidade Federal do Espírito Santo (UFES) e no Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi), adelmoib@gmail.com;
- 4 Cientista no Instituto Todos pela Saúde (ITpS) e Coordenador de curso no Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi), marcelobragatte@gmail.com;

dendo cursistas com diferentes formações. A execução aconteceu em três módulos e foi finalizada com projetos individuais de coleta e análise de dados relacionados às práticas dos cursistas. Os resultados e efetividade foram observados a partir da entrega de tarefas, desempenho, relatos dos cursistas e questionários de avaliação. O curso teve 166 participantes com 18 conclusões. Uma dificuldade relacionada à retenção dos participantes foi a não adaptação do curso com sua rotina de trabalho. A heterogeneidade dos conhecimentos prévios foi desafiadora, mas contornada com tutoria, uso dos encontros síncronos e direcionamento dos projetos de acordo com o potencial individual dos cursistas. Os projetos finais evidenciaram excelente apropriação do conhecimento e direcionamento para áreas de interesse específicas, associados à avaliações positivas dos participantes. Uma nova edição está sendo construída com base nestes resultados, com foco em vídeos diários menores e acompanhamento com atividades semanais simples para atingir maior retenção. O EAD se mostrou um excelente aliado para conduzir um treinamento descentralizado com resultados concretos para a gestão da Saúde Pública.

Palavras-chave: Linguagem R, Ciência de Dados, Estatística, Saúde Pública, Ensino à Distância.

INTRODUÇÃO

Nas últimas décadas, o desenvolvimento de novos métodos de análise, os avanços tecnológicos e a crescente disponibilidade de bases de dados digitalizadas passaram a exigir que profissionais de diversas áreas dominem minimamente a capacidade de ler, visualizar, analisar e interpretar conjuntos de dados (PROVOST; FAWCETT, 2013). Esta necessidade de “alfabetização estatística” (SHARMA, 2017) afeta desde a capacidade de se desenvolver pesquisas científicas, realizar trabalhos técnicos, tomar decisões baseadas em evidências em políticas públicas e até mesmo leituras cotidianas de notícias envolvendo temas de interesse geral como saúde, economia e meio ambiente (TONI; DORNELES, 2022).

Na área da saúde, em particular, o domínio de métodos de análise e de comunicação de dados (como *storytelling* de dados) se mostrou essencial durante a pandemia de COVID-19, por exemplo (ROSA; SILVA; PAVÃO, 2023). Mesmo quando decisões eram tomadas a partir de análise rigorosa dos dados, nem sempre a forma como os dados, gráficos e conclusões das análises eram apresentados pelos veículos de mídia e interpretados pela população era adequada (SILVA, 2022), demonstrando a necessidade de se comunicar resultados de uma forma clara e eficiente, mesmo quando o público alvo tem pouca ou nenhuma base de conhecimento formal em estatística e análise de dados..

Entretanto, obter este domínio de métodos de análise, visualização e comunicação de dados pode ser desafiador ao considerarmos pesquisadores, técnicos e tomadores de decisão que possuem diferentes formações e competências. Além dos aspectos matemáticos envolvidos, a implementação dos métodos dependem de aplicações computacionais com diferentes níveis de exigência técnica para seu uso. Em particular, as implementações computacionais de métodos atuais é normalmente realizada pelo uso de linguagens de programação, pois estas permitem tanto a implementação rápida de novos métodos na medida em que eles são desenvolvidos quanto a reprodutibilidade destes para diferentes usuários (JUAVINETT, 2022; NORDMANN et al., 2022).

A principal linguagem de programação voltada para o campo da Ciência de Dados e utilizada por diversos campos de pesquisa e aplicação é a linguagem R. Atualmente o R é utilizado em publicações acadêmico-científicas, empresas e por tomadores de decisão (GIORGI; CERAOLO; MERCATELLI, 2022; MAIR et al., 2015). Em relação aos desafios técnicos que o domínio de uma linguagem

de programação podem trazer, o R possui uma vantagem por ter sido projetado para um uso muito específico (análise e visualização de dados) e ter sido projetado como linguagem de “alto nível”. Isto faz que um usuário final possa ter pouco ou nenhum domínio de lógica de programação e da linguagem em si ao começar a usá-lo, e mesmo assim realizar tarefas e aprender conforme necessita de novas aplicações para a linguagem. Finalmente, o fato da linguagem ser de código aberto possibilitou reunir uma grande e ativa comunidade de usuários e desenvolvedores, muitas vezes ligados ao meio acadêmico, fazendo com que o R seja amplamente usado, possua muitos canais de aprendizagem e resolução de problemas e, por fim, tenha um desenvolvimento e expansão constantes (AUKER; BARTHELMESS, 2020; TUCKER et al., 2023).

Considerando este contexto desafiador, entre Agosto de 2022 e Dezembro de 2023 oferecemos, a partir do Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi), o curso “Capacitação em Análise de Dados para a Qualificação da Gestão em Saúde” em formato EAD para um conjunto inicial de 263 Agentes Públicos da Saúde do Estado do Espírito Santo. Nossa meta foi capacitar agentes do Sistema Único de Saúde (SUS) do estado para tomadas de decisão e a avaliação das políticas públicas de saúde, em especial as relacionadas à vigilância em saúde, a partir de treinamento em ferramentas de análise de dados e estatística e da realização de estudos e análises de dados epidemiológicos sob demanda relacionadas ao projeto.

Neste trabalho, relatamos o planejamento, execução, resultados e aproveitamento deste curso, bem como nossas reflexões e lições aprendidas durante esta trajetória. O curso foi estruturado e aplicado em ambiente EAD, com aulas pré-gravadas, uso de fóruns e encontros síncronos regulares. O conteúdo foi distribuído em três módulos, iniciando com temas e conceitos básicos de análise e visualização de dados, passando por métodos de análise e inferência e finalizando com o desenvolvimento de projetos individuais. Apesar de uma quantidade relativamente pequena de concluintes do último módulo em relação aos matriculados no primeiro, consideramos que o curso teve grande sucesso em sua meta de treinar e instrumentalizar agentes de saúde nos processos de tratamento, exploração, análise e comunicação de dados. Consideramos que nossa visão que respeitou os desafios e dificuldades de uma turma heterogênea, o uso eficiente e ágil de tutoria e a organização como um todo do conteúdo (incluindo a flexibilidade em rever os conteúdos ao longo do trajeto) contribuíram fortemente para os bons resultados.

METODOLOGIA

O curso foi estruturado na plataforma *Moodle*, em versão disponibilizada pelo ICEPi, com a disponibilização de vídeos com aulas para serem assistidas de forma assíncrona, material para *download*, atividades autodirigidas e tutoria tanto pelo uso de fórum quanto por envio de mensagens privadas. Encontros ao vivo síncronos foram realizados pela plataforma *Zoom*, também disponibilizada pelo ICEPi, tendo sido gravados em sua totalidade e disponibilizados para acesso com as demais aulas, para que pudessem ser revistas se necessário e/ou acessadas pelos participantes que não puderam comparecer ao vivo. Todas as vídeo aulas e gravações de encontros ao vivo foram disponibilizados utilizando plataforma *Youtube* com o formato de “vídeos não listados”, que eram então incorporados no *Moodle* dentro da estrutura do curso (de forma que os links originais dos vídeos das aulas poderiam ser usados, por exemplo, para se assistir aulas em um dispositivo móvel que não esteja conectado ao *Moodle* do curso). As aulas e atividades foram organizadas em três módulos sequenciais, cada um abordando uma fase específica de aprendizado em ciência de dados, estatística e comunicação aplicados à saúde pública usando a linguagem de programação R como ferramenta para análise e visualização de dados.

Além do relato de conteúdos, métodos e formatos utilizados, os resultados apresentados aqui em relação à participação, aproveitamento e aprendizagem dos participantes são baseados nos seguintes instrumentos: nossas observações como docentes do curso, debatidas em reuniões de regularidades semanal e derivadas principalmente dos momentos de interação com a turma encontros síncronos; o registro de dúvidas e outras interações relacionadas à tutoria por escrito via fórum e mensagens no *Moodle*; a avaliação das atividades entregues pelos participantes e a adequação destas atividades em relação ao conteúdo; formulários anônimos de avaliação preenchidos pelos participantes ao final de cada módulo.

Todos estes dados foram avaliados de forma quali-quantitativa, conforme adequado à sua natureza. As respostas dos participantes em questionários avaliativos a respeito da estrutura, formato e conteúdos do curso foram analisadas com base nas etapas propostas pela Análise de Conteúdo de Bardin (BARDIN, 2020). No decorrer do curso, os momentos de diálogo e *feedback* dos participantes juntamente com os formulários de avaliação também nos permitiu flexibilizar entregas previstas, rever conteúdos planejados (e ainda não grava-

dos) para os módulos 2 e 3 e, no geral, reorganizar atividades de forma a melhor adaptar a experiência para todos.

RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta o quantitativo total de participação do curso, que iniciou com 263 participantes matriculados. A cada módulo, apesar da demonstração de interesse ao realizar a matrícula, tivemos participantes que não acessaram a plataforma e, desta forma, não participaram de qualquer atividade prevista. Assim, consideramos a taxa de conclusão utilizando o número de aprovações apenas em relação ao total de participantes ativos, que de fato acessaram a plataforma e participaram, mesmo que de forma incompleta, daquele módulo.

Tabela 1: Quantidade de matrículas, participação, aprovação, envio de projetos de taxa de conclusão do curso.

Módulo	Matrículas	Participantes ativos	Aprovados	Projetos enviados	Taxa de conclusão
1	263	166	84	176	50,6%
2	70	58	30	30	51,7%
3	30	24	18	18	75%

Fonte: autores

Os Módulos 1 e 2 foram compostos de aulas diárias pré-gravadas (disponibilizadas de segunda a sexta, exceto feriados) com duração aproximada de 20 minutos cada, alternando entre conteúdos teóricos e aulas teórico-práticas utilizando a linguagem de programação R a partir do uso da plataforma *RStudio* (que incluíram aulas iniciais dando instruções de instalação do software e uso de uma versão disponível na nuvem, que pode ser utilizada sem instalação a partir do uso de um navegador de *Internet*). As atividades práticas e autodirigidas guiavam os cursistas na aplicação em dados relacionados às temáticas abordadas, e tinham como principal objetivo a prática na linguagem R dos métodos apresentados. Ao longo de todo o curso foram realizados encontros síncronos nos quais os participantes podiam interagir por áudio e vídeo, tendo a oportunidade e sendo estimulados a sanarem dúvidas e compartilharem suas dificuldades, inquietações e expectativas de aprendizagem. Ao final de cada módulo projetos finais relacionados a seu conteúdo deveriam ser entregues pelos participantes para sua conclusão, com diferentes opções de envio visando participantes com

diferentes perfis de formação e de aprendizagem. Finalmente, o módulo 3 foi estruturado para permitir o desenvolvimento e execução de projetos de pesquisa com uso dos conhecimentos adquiridos nas etapas anteriores, sem o uso de aulas pré-gravadas e se apoiando no uso de encontros síncronos em grupo e tutoriais dos participantes com ministrantes-orientadores designados a eles de forma individual (incluindo reuniões individuais de orientação para os projetos ao longo do período do módulo).

O Módulo 1 do curso buscou introduzir os conceitos básicos de estatística e ciência de dados, utilizando a linguagem R como ferramenta principal para as atividades práticas (com aulas dedicadas à compreensão desta linguagem, utilizando a plataforma RStudio), e foi executado entre 15/08/2022 a 16/12/2022 em uma estrutura de 16 semanas letivas. Nele foram transmitidos os conceitos, os pré-requisitos e os passos a serem trilhados ao longo do curso. Conceitos como Estatística, *Data Science*, Epidemiologia, dado, informação e suas inter-relações e importância na sociedade foram apresentados junto com exemplos de situações reais - sempre com ênfase em exemplos e aplicações voltadas para a área da saúde, considerando o público alvo participante.

A execução deste módulo exigiu flexibilização e revisão da estrutura e planejamento do curso, na medida em que notamos uma grande heterogeneidade de conhecimentos prévios e habilidades dos participantes no uso das ferramentas, tanto pelas dificuldades e dúvidas levantadas no uso do fórum quanto (e principalmente) nos diálogos estabelecidos durante os encontros síncronos - que desempenharam um papel fundamental neste processo, uma vez que alguns participantes se sentiram à vontade para expor suas inquietações e dificuldades durante estas conversas, a na medida em que os estimulamos a fazer isso. Uma modificação que imediatamente demonstrou efetividade no desempenho e aproveitamento do curso foi uma semana de nivelamento realizada durante a semana oito (batizada de Nivelar, em uma referência à linguagem R), na qual os participantes com maior dificuldade tiveram a chance de avançar com aulas e atividades atrasadas, além de sanar dúvidas antes que o conteúdo continuasse a avançar.

Também considerando a heterogeneidade da turma e em estimular o aperfeiçoamento sem exigir um domínio avançado de todos os participantes, ao final deste módulo os projetos finais foram apresentados com opções de envio: cada participante poderia escolher entre cinco propostas de projetos (com diferentes níveis de dificuldade), dos quais o envio de apenas um era suficiente

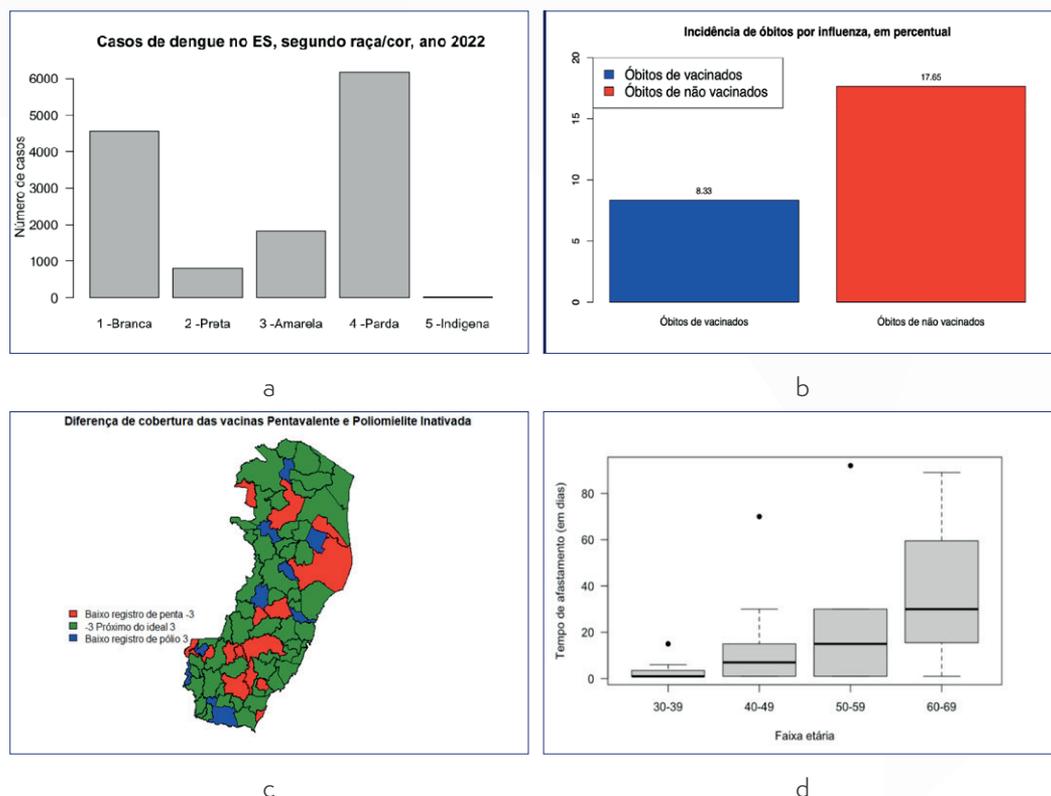
para obter aprovação (possibilitando a obtenção do certificado de conclusão e a solicitação de matrícula no Módulo 2). Para estimular o envio de mais projetos, utilizamos o sistema de emblemas da plataforma *Moodle*, conferindo uma “medalha” simbólica de bronze aos cursistas que enviassem um projeto, uma de prata aos que enviassem dois e, finalmente, uma de ouro para os que enviassem três ou mais. Por fim, participantes que criassem projetos de autoria própria com análise de dados reais tiveram também a atribuição da medalha de ouro. Pudemos notar que uma parte dos cursistas se sentiu estimulada pelo desafio de obter medalhas, de forma que um total de 176 projetos foram enviados por 84 participantes. Este quantitativo de envios nos mostrou que pequenos estímulos e processos simples de gamificação como este podem ser surpreendentemente eficazes para estimular participação e engajamento, mesmo em uma situação na qual não há atribuição de notas. Além disso, diversos projetos individualizados bastante relevantes dentro do contexto de tema e conteúdos do curso foram criados, dos quais destacamos alguns como exemplo (com imagens selecionadas para quatro deles, na Figura 1):

- Um estudo de prevalência trazendo a frequência de casos de dengue no estado do ES, estratificados por raça/cor, escolaridade, sexo e classificação do caso (Figura 1a);
- Incidência de óbitos por influenza em vacinados e não vacinados, para levantamento de hipótese (Figura 1b);
- Prevalência das coberturas vacinais da vacina pentavalente e contra poliomielite no estado do ES, para melhor distribuição das doses em estoque (Figura 1c);
- Incidência de doença meningocócica nos estados do Brasil, demonstrando que o estado do ES tem a maior incidência do país no período estudado;
- Análise de mandados judiciais para atendimento de saúde por tipo de solicitação, para melhor tomada de decisão dos gestores (Figura 1d).

Ao longo deste módulo também notamos problemas práticos que fugiam da estrutura didática proposta, como dificuldades técnicas de instalação de ferramentas nos locais de trabalho (que muitas vezes dependiam de respostas de setores de Tecnologia da Informação (TI), levando um tempo considerável para serem resolvidas) e demandas urgentes que surgiam ligadas aos órgãos de vín-

culo dos participantes, que acabavam por comprometer sua participação (mesmo considerando que os servidores deveriam ter o tempo cedido para o curso pelas suas chefias imediatas). Apesar de não ser possível quantificar precisamente, entendemos que uma parte das desistências antes do final do curso foram causadas por fatores externos como estes, e estas dificuldades foram comunicadas ao ICEPi para que possam ser amenizadas durante a próxima edição.

Figura 1: Imagens que mostram partes dos resultados dos exemplos seleccionados dos projetos de conclusão do Módulo 1.



Fonte: autores

O Módulo 2 abordou temas mais avançados a partir da base conceitual e prática adquirida no primeiro, e foi executado entre 27/02/2023 a 26/06/2023. Temas como Análise Espacial, Inferência Estatística, métodos de *Machine Learning* e Comunicação de Dados (com técnicas para construir relatórios vinculados à linguagem R) foram abordados e praticados nas aulas e atividades, novamente mantendo o foco em práticas e exemplos ligados à área de saúde, epidemiologia e saúde pública. A partir da experiência com o módulo anterior

e seu aproveitamento geral, neste módulo utilizamos uma avaliação intermediária dentro de um período de duas semanas nas quais não foram lançadas novas aulas de conteúdo, com o objetivo de ter uma visão intermediária do rendimento da turma ao mesmo tempo que os participantes pudessem ter tempo para alcançar conteúdos atrasados (de forma semelhante à semana “nívelaR” do módulo 1). Por fim, os projetos finais deste módulo foram novamente escolhidos entre diferentes opções com variados graus de desafio - mas sem a opção de envios múltiplos.

De uma forma geral, a participação neste último módulo exigiu o estudo de conceitos e métodos mais complexos, como as técnicas de *Machine Learning*. Apesar dos participantes terem acompanhado as aulas a respeito destes temas e consumido todo o material proposto, a execução dos projetos ao final do módulo 3 (descrito a seguir) nos permitiu notar que eles quase não foram utilizados. Os cursistas fizeram mais uso de métodos de exploração e visualização de dados, e é difícil inferir se conseguirão lançar mão das ferramentas mais sofisticadas que foram apresentadas. Em uma reflexão após o término de todo o curso, compreendemos que com um público tão heterogêneo é mais vantajoso focar em dominar com muita segurança métodos e conceitos fundamentais, permitindo que os participantes explorem outros conteúdos mais avançados de forma autônoma quando julgarem interessante e/ou necessário.

O Módulo 3 foi conduzido entre 16/08/2023 e 10/12/2023 em uma estrutura voltada para o desenvolvimento e execução de projetos, com um formato intercalando encontros síncronos com toda turma e encontros individualizados com professores-orientadores que foram designados para cada cursista matriculado. A proposta geral deste módulo foi de que cada participante iria, com base nos conhecimentos adquiridos até então, desenvolver projetos para solucionar problemas ou perguntas gerais associados à sua rotina ou área de atuação, sempre partindo de dados reais e solucionando questões concretas que pudessem trazer um retorno ou melhoria ao seu trabalho. O grau de complexidade dos problemas e dos métodos envolvidos ficou à cargo dos próprios cursistas, partindo tanto do seu interesse quanto de sua segurança. Nós, como professores-orientadores nesta etapa, fornecemos o apoio no acompanhamento dos projetos e solução das questões relacionadas tanto à elaboração conceitual quanto da execução prática das análises dos dados. Todos os participantes que finalizaram este módulo realizaram uma entrega de relatório e *scripts* na linguagem R.

Os 18 projetos entregues trataram dos seguintes temas:

1. Traçar o perfil dos pacientes que acessaram os prontos atendimentos de Vitória-ES, num determinado período, coletando informações quanto ao sexo, idade, UBS de origem e micro área a que pertencem.
2. Estudar a relação de acidentes com animais peçonhentos com acidentes de trabalho em dois municípios do interior do Estado do ES de um determinado ano.
3. Analisar o estoque de demandas judiciais em trâmite na SESA segundo o tipo de solicitação e procedimento.
4. Coletar os dados de atestados médicos do Hospital Dório Silva, município de Serra-ES, e analisar o quantitativo por mês, por CID e quantidade de dias de licença.
5. Análise exploratória dos dados das cirurgias eletivas do estado do ES
6. Analisar o perfil epidemiológico da Comissão de Controle de Infecção Hospitalar (CCIH) nos anos de 2021/22 no Centro de Atendimento Psiquiátrico Aristides Alexandre Campos (CAPAAC) em Cachoeiro de Itapemirim-ES, para identificar quais os patógenos predominantes, identificando a evolução mensal do tipo de topografia da infecção no período.
7. Estimar a variação real do PIB do Estado do Espírito Santo para o ano de 2023 e para os dois anos seguintes em base em dados do indicador antecedente de PIB trimestral, calculado pelo Instituto Jones dos Santos Neves.
8. Estudar os estabelecimentos do Estado do Espírito Santo que mais notificam queixas técnicas e eventos adversos de produtos para saúde a partir do banco de dados da Agência Nacional de Vigilância Sanitária (sistema NOTIVISA).
9. Estudar os efeitos dos fatores político-institucionais sobre a probabilidade de descumprimento do limite de despesa com pessoal estabelecido na LRF no Estado do ES. Os dados são provenientes da Siconfi/STN, SIDRA, SEFAZ-ES e IBGE.
10. Estudar dados de Ingresso e conclusão no ensino superior no RS, investigando o total de estudantes que concluíram os estudos por sexo e por ano e verificando o impacto dos os anos de pandemia (2019, 2020 e 2021) no ingresso nas taxas de conclusão.
11. Estudar a taxa de positividade e a distribuição geográfica das amostras de água coletadas pelas vigilâncias municipais no Brasil no que tange à

- contaminação por agrotóxicos, bem como buscar as principais classes de agrotóxicos encontrados nessas análises.
12. Analisar, a partir dos registros do setor de Assistência Farmacêutica Estadual no período de 2021 a 2022 o cenário de distribuição do medicamento rivaroxabana no estado do ES para pacientes com diagnóstico de fibrilação atrial.
 13. Agrupar as escolas estaduais de Ensino Fundamental - anos finais - no ES a partir do uso de técnica multivariada de Análise de Cluster, partindo de dados quantitativos, demográficos e de funcionamento destas instituições.
 14. Análise das correlações dos dados de indivíduos amputados do setor de Próteses e Órteses do Centro de Reabilitação do Espírito Santo, investigando laudos, tempo de espera para recebimento de próteses e taxas de amputação por município.
 15. Estudo da incidência de agravos por meses do ano a partir de dados abertos de Agravos da secretaria de saúde do estado do Espírito Santo, para responder quais meses do ano temos mais agravos e como foi a evolução através dos anos.
 16. Levantamento de afastamento de servidores por licença médica para tratamento de saúde em uma Unidade Hospitalar, entre o período de janeiro de 2022 a agosto de 2023.
 17. Estudo da capacidade de atendimento às demandas de exames laboratoriais do Laboratório Central de Saúde Pública do Espírito Santos (LACEN/ES), investigando quais ensaios ou procedimentos têm maior demanda operacional e mapeando a capacidade de cada setor, bem como a capacidade total do LACEN.
 18. Estudo da demanda do setor de regulação de leitos da SESA/ES quanto aos principais motivos de cancelamento de vagas e quais os estabelecimentos com maior frequência de cancelamentos.

Podemos notar que os temas explorados nos projetos e sua complexidade técnica variaram amplamente, mas quase sempre envolveram uma relação direta com o setor e área de atuação dos participantes. Em alguns casos, a proposta de projeto partiu de um entendimento dos cursistas das necessidades de seus órgãos de vínculo, e muitos tinham como meta dar continuidade ao que foi realizado na conclusão do módulo, visando melhorar a eficiência de seu trabalho

individual, automatizar processos manuais e/ou solucionar problemas conhecidos de forma a beneficiar os órgãos ou o sistema de saúde como um todo. Nesta etapa, os participantes demonstraram compreender a utilidade prática dos conhecimentos e métodos explorados até então ao longo dos dois módulos anteriores. Mesmo com uma taxa final de conclusão bastante pequena em relação à demanda inicial de matrículas, estes projetos e o *feedback* dado pelos participantes demonstrou que a proposta geral do curso foi bastante acertada, mesmo que algumas mudanças sejam necessárias para melhor aproveitamento em edições futuras (como uma segunda edição sendo iniciada no momento de redação deste trabalho).

Os participantes relataram que o uso e aplicação do conhecimento em projetos que eram relevantes diretamente em sua atuação foi um ponto extremamente positivo ao final do curso, e que se sentiram satisfeitos com esta oportunidade que possibilitou a aquisição de novos conhecimentos e a aplicação de forma verdadeiramente prática. Muitos relataram que dariam continuidade aos projetos iniciados ao fim do módulo 3, e que continuariam a utilizar o R e suas funcionalidades em Ciências de Dados daqui para frente. Estes relatos nos levam a acreditar que o resultado do curso vai além dos 18 concluintes deste último módulo, uma vez que estes participantes podem alterar e otimizar rotinas de seus ambientes de trabalho, estimular o interesse de colegas pelos temas e centralizar, a partir de sua atuação, um melhor uso de dados para tomadas de decisão nestes órgãos públicos.

Além das observações descritas até agora feitas com base no comportamento, entregas, perguntas e participações dos cursistas, o uso de questionários de avaliação do curso e dos módulos nos permitiu compreender tanto importantes acertos quanto pontos que demandam melhorias, que podem então ser incorporadas em experiências futuras. As sugestões e relatos de problemas foram usados ao longo do curso com alterações e flexibilizações do planejamento (como a semana de nivelamento do Módulo 1, por exemplo), além de serem levadas para nosso planejamento da segunda edição do curso, que acontecerá com grandes alterações com base nesta experiência. A partir das etapas de pré-análise e exploração (com categorização) da análise de conteúdo (BARDIN, 2020) podemos ressaltar os principais resultados e aspectos positivos inferidos destes dados:

- Didática dos professores: a capacidade de explicar conteúdos complexos de forma simples foi bastante elogiada, principalmente para o Professor Marcos Vital;
- Relevância profissional: a maioria dos cursistas reconheceu o conteúdo como importante para suas atividades profissionais e seu desempenho no trabalho;
- Agilidade nas respostas às dúvidas: a agilidade do retorno nas dúvidas e avaliações no processo de tutoria do curso foi citada como um ponto significativamente positivo para a continuidade, especialmente entre os que consideravam o conteúdo desafiador;
- Estrutura e planejamento: o planejamento do curso foi elogiado, mesmo pelos cursistas sem conhecimento prévio dos conceitos e métodos abordados.

Mantivemos a menção direta ao professor Marcos Vital (um dos autores deste relato), pois o mérito e a importância de um professor que humaniza o ensino é inestimável, e houve nas respostas dos questionários uma unanimidade em elogios para o instrutor - trazendo um retorno quantitativo de um aspecto qualitativo difícil de ser capturado. Mas entendemos que este retorno específico pode ser compreendido a partir de um comportamento demonstrado, que parece ter gerado este efeito: em salas de aula com públicos heterogêneos, especialmente envolvendo assuntos desafiadores como programação, esse educador teve a sensibilidade de acalmar ansiedades e, mesmo não sendo possível reconhecer as necessidades individuais em turmas grandes, elencar exemplos coletivos que beneficiam a todos. Ao desmistificar a aplicação da teoria na prática, principalmente em momentos síncronos de aprendizado, o educador replica ambientes inclusivos que incentivam o questionamento abrindo a porta para o interesse e conexões, refletindo positivamente no resultado. Entendemos, portanto, que a sensibilidade e empatia em relação aos diferentes perfis e dificuldades dos cursistas teve grande efeito prático no processo de aprendizagem.

A relevância do conteúdo e suas aplicações no campo profissional, como relatado anteriormente a partir dos *feedbacks* durante encontros síncronos e outros momentos de comunicação, foi confirmada nos questionários. Entendemos que existe uma relação bastante simples e direta entre o interesse em se dominar um conteúdo e a sua aplicação prática de forma concreta. Além disso, as respostas dos questionários reforçam a importância de uma tutorial efi-

ciente e uma estrutura bem organizada para os conteúdos. A tutoria ágil permitiu que os participantes com maiores dificuldades (o que acontecia especialmente em relação à execução de métodos na linguagem R) pudessem ter suas dúvidas e problemas resolvidos rapidamente, de forma que se sentiram acolhidos e construíram confiança para dar continuidade ao aprendizado. A organização do curso como um todo, com tópicos iniciais básicos e focado em um aprendizado “passo a passo”, também contribuiu para isso. Nossa percepção é de que conteúdos considerados mais difíceis e desafiadores de uma forma geral (como conteúdos de matemática e de programação) estão ao alcance de estudantes e profissionais que não sentem facilidade com eles - mas para que isso de fato aconteça, um planejamento que tenha em mente especificamente as dificuldades típicas de aprendizado destes temas é essencial.

Por outro lado, como pontos principais de melhorias e dificuldades encontradas, nossa análise do conteúdo dos questionários nos indicou:

- Dificuldade em conciliar a rotina de trabalho com a carga horária das aulas e atividades. O tempo demandado pelo conteúdo do curso foi maior do que o previsto pelos participantes, e na nova edição revisamos o formato para gerar aulas ainda mais enxutas e diretas (reduzindo o tempo para no máximo 15 minutos por aula, e dispersando as atividades em formatos menores mas mais frequentes, no lugar de atividades mais complexas e menos frequentes).
- Foram relatados problemas práticos de acessibilidade e infraestrutura: dificuldades com acesso à internet e com o acesso a equipamentos no trabalho foram apontadas como pontos que comprometeram a participação. Estas são questões que fogem da nossa competência como instrutores, mas que foram levadas de volta ao instituto responsável pelo curso e aos órgãos de vínculo dos servidores participantes.

A partir das reflexões e da própria escrita deste relato (que nos fez revisitar e repensar os *feedbacks*, os questionários e a experiência completa), consideramos que esta primeira edição do curso foi um sucesso, mesmo considerando uma quantidade final pequena de concluintes em relação ao grupo inicialmente matriculado. A estrutura em módulos permitiu que parte dos participantes encerrassem seu aprendizado no momento desejado, uma vez que o conteúdo pode não ter tido o mesmo apelo a todos. Dos pontos negativos e dificuldades

apontadas, ao mesmo tempo escrevemos este relato e planejamos uma nova edição, que está se iniciando no momento do encerramento da redação deste texto. Alteramos, como já relatado, a duração das aulas para um máximo de 15 minutos, e muitas vezes apenas 12 minutos de duração. Esperamos que este formato reduzido diminua as dificuldades de conciliar o conteúdo com o trabalho. Também simplificamos o conteúdo como um todo, deixando temas de alta complexidade (como *Machine Learning*, por exemplo) apenas como algo a ser mencionado, com possível indicação de materiais para estudo. Esperamos, com isso, que os participantes que se interessam ou necessitam destes métodos possam utilizar o curso como um ponto de partida para começar a explorar o tema e seguir de forma autônoma (ou buscando cursos específicos). Também reduzimos a estrutura para apenas dois módulos, de forma a tentar aumentar a retenção dos participantes até o final do curso. Assim, esperamos atingir um resultado mais satisfatório com um maior número de pessoas, mesmo que com uma menor densidade de conteúdo adquirido. Esperamos, mesmo assim, que os concluintes desta nova etapa possam, como o da relatada aqui, aplicar os conhecimentos em situações práticas, e se sentirem estimulados e preparados para prosseguir com seu aprendizado.

CONSIDERAÇÕES FINAIS

A partir dos projetos, feedbacks e conversas com os concluintes, podemos concluir que cursos profissionalizantes como este podem ter grande sucesso e impacto nos participantes, desde que atentem para algumas premissas e necessidades básicas. A compreensão de que a turma era heterogênea e com participantes com diferentes inclinações, formações, interesses e competências - e conversar abertamente com eles a respeito disso - foi essencial na construção dos conteúdos, sua ordem e o ritmo de aprendizagem ideal. Entendemos que uma boa compreensão do público alvo seja essencial para se reproduzir este tipo de experiência, e que esta compreensão vai além do contato inicial, e deve ser expandida ao longo do trajeto (com chances para se rever o planejamento realizado). Também, ao longo do trajeto é essencial realizar momentos de escuta e empatia, permitindo ajustes e fazendo com que os participantes se sintam ouvidos e acolhidos em suas individualidades e dificuldades.

Foi essencial trabalharmos não apenas com temas considerados interessantes ou relevantes, mas que são também úteis de um ponto de vista prático. O

curso aqui executado apresentou e explorou o estado da arte no que tange às formas de realizar uma análise de dados com uso do software R. Abordou as tendências atuais para os profissionais da área de *Data Science*, focando no tema da saúde pública. O mesmo também mencionou levemente como as inovações em *Machine Learning* transformaram a análise de dados, sendo um ferramental essencial para a tomada de decisões informadas. Além disso, o curso proporcionou uma visão que abarcou as competências necessárias que os profissionais devem desenvolver para se destacar no campo da ciência de dados através de estudos de caso e exemplos práticos. Aqueles cursistas da SESA que avançaram até o final tiveram a oportunidade de atuar diretamente em seus problemas propostos, impactando diretamente suas atividades, e promovendo uma gestão mais eficiente.

Por fim, gostaríamos de ressaltar a grande relevância de uma proposta como esta, de trazer métodos e abordagens para um campo essencial de nossa sociedade: as políticas de saúde pública. Em tempos nos quais a desinformação tem impacto prático em nossas vidas e tomadas de decisão, é essencial ressaltar e projetar a importância do uso de métodos sólidos e eficazes para se analisar dados, compreender fenômenos e tomar decisões baseadas em evidências científicas, que possam nos conduzir a um cenário de melhoria da saúde coletiva e individual. Neste sentido, uma boa alfabetização estatística é extremamente positiva para nossa sociedade, e não apenas para aquelas pessoas envolvidas com a área da saúde.

REFERÊNCIAS

AUKER, L. A.; BARTHELMESS, E. L. Teaching R in the undergraduate ecology classroom: approaches, lessons learned, and recommendations. **Ecosphere**, v. 11, n. 4, p. e03060, 2020.

BARDIN, L. **Análise de Conteúdo**. 4281. ed. Portugal: Edições 70, 2020.

GIORGI, F. M.; CERAOLO, C.; MERCATELLI, D. The R Language: An Engine for Bioinformatics and Data Science. **Life**, v. 12, n. 5, p. 648, maio 2022.

JUAVINETT, A. L. The next generation of neuroscientists needs to learn how to code, and we need new ways to teach them. **Neuron**, v. 110, n. 4, p. 576–578, 16 fev. 2022.

MAIR, P. et al. Motivation, values, and work design as drivers of participation in the R open source project for statistical computing. **Proceedings of the National Academy of Sciences**, v. 112, n. 48, p. 14788–14792, dez. 2015.

NORDMANN, E. et al. Data Visualization Using R for Researchers Who Do Not Use R. **Advances in Methods and Practices in Psychological Science**, v. 5, n. 2, p. 1–36, 2022.

PROVOST, F.; FAWCETT, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. **Big Data**, v. 1, n. 1, p. 51–59, mar. 2013.

ROSA, S. S. DA; SILVA, F. C. C. DA; PAVÃO, C. M. G. Iniciativas de acesso aberto no combate à pandemia dados abertos e propriedade intelectual na disseminação da informação e conhecimento. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 19, p. e021025, 9 jun. 2023.

SHARMA, S. Definitions and models of statistical literacy: a literature review. **Open Review of Educational Research**, v. 4, n. 1, p. 118–133, 1 jan. 2017.

SILVA, A. R. Storytelling com dados: proposta de visualização de dados de pesquisa científica. **Ciência da Informação Express**, v. 3, p. 1–4, 15 set. 2022.

TONI, J. DE; DORNELES, R. **Ciência de dados em políticas públicas: uma experiência de formação**. 1. ed. [s.l.] Escola Nacional de Administração Pública (Enap), 2022.

TUCKER, M. C. et al. Teaching Statistics and Data Analysis with R. **Journal of Statistics and Data Science Education**, v. 31, n. 1, p. 18–32, 2 jan. 2023.