

Quarto com vista para o campo

Autores: Renata Florentino de Faria Santos, Carolina Musso, Leandro de Almeida Salles, Rafael de Acypreste Monteiro Rocha, Rogério Vidal de Siqueira, Aline da Nóbrega Oliveira, Luiz Antônio Gouveia de Oliveira.

Palavras chave: Quarto Book, survey, Reprodutibilidade, Serviço Público

Resumo

A Pesquisa Distrital por Amostra de Domicílios (PDAD) Rural, realizada em 2021, foi essencial para a compreensão das condições socioeconômicas das comunidades rurais do Distrito Federal. Este trabalho apresenta uma abordagem integrada, utilizando o R para todas as etapas da PDAD, desde a amostragem até a elaboração do relatório final. O relatório foi compilado utilizando o *Quarto Book*, que permitiu uma documentação reproduzível, colaborativa e bem estruturada. A criação de tabelas com *flextable* e gráficos de alta qualidade com *ggplot2* integrados por *patchwork* permitem um relatório final claro e informativo, e acima de tudo reproduzível. Dessa forma, ele pode ser atualizado e replicado por equipes futuras, assegurando a continuidade das políticas públicas locais.

Introdução

A Pesquisa Distrital por Amostra de Domicílios (PDAD) Rural, realizada em 2021, foi a primeira e única de seu tipo, sendo posteriormente substituída pela PDAD Ampliada, marcando a primeira vez que a população rural foi diretamente acessada para análise. Esta pesquisa desempenhou um papel crucial na compreensão das condições socioeconômicas das comunidades rurais do Distrito Federal. Durante sua execução, diversos estatísticos participaram do projeto, refletindo a rotatividade comum de profissionais e gestores no serviço público. Por esse motivo, a reproduzibilidade dos processos é fundamental para continuidade de pesquisas que impactam políticas públicas e o desenvolvimento do Distrito Federal. Este trabalho tem o objetivo de apresentar a importância R para conduzir todas as etapas da pesquisa, (desde a amostragem até a elaboração do relatório final) e a utilidade do formato *Quarto Book* para elaboração de um relatório final construído a várias mãos.

Metodologia

A metodologia foi majoritariamente conduzida no R, cobrindo todas as etapas do projeto. O processo começou com o sorteio de amostras de domicílios, utilizando funções do R base. Após a aplicação do questionário padrão nos domicílios amostrados, a base de dados foi cuidadosamente explorada e processada para remover duplicidades e inconsistências, utilizando pacotes como o *validate*. Os pacotes *survey* e *srvyr* foram utilizados para expandir a amostra e realizar a inferência populacional. Os pacotes

QUARTO COM VISTA PARA O CAMPO

ggplot2 e *flextable* foram utilizados para a confecção dos gráficos e tabelas respectivamente. Por fim, a ferramenta *Quarto Book* foi escolhida para a criação do relatório no formato *.docx* editável. Os resultados referentes às características do domicílio foram organizados em torno dos principais usos da terra: agricultura empresarial, agricultura familiar, lazer/moradia e assentamento/agrovila.

Resultados

O trabalho desenvolvido pelos estatísticos que integraram o projeto desde 2021 pôde ser reproduzido e complementado, graças à natureza reprodutível possibilitada pela linguagem de programação utilizada. Além disso, as estimativas populacionais foram obtidas de forma estatisticamente correta, respeitando a complexidade da amostra, por meio das implementações oferecidas dos pacotes *survey* e *srvyr*.

Destacamos aqui a inovação trazida pela utilização do *Quarto Book* para a elaboração do relatório final, que foi produzido de maneira reprodutível e colaborativa. A estrutura do *Quarto Book* facilita a criação de documentos independentes no formato *.qmd* dentro do projeto, promovendo uma colaboração mais fluida entre os membros da equipe. Assim, a divisão do trabalho em capítulos permitiu que diferentes membros contribuissem simultaneamente em partes distintas do relatório, minimizando os conflitos durante o versionamento no GitHub. Essa fragmentação também trouxe maior estabilidade ao documento, facilitando a identificação de erros e possibilitando o acesso paralelo às informações.

O arquivo da *Introdução* oferece uma contextualização histórica da população e das atividades rurais no Distrito Federal. O da *Metodologia* descreve o plano amostral em detalhes, os estimadores utilizados e o tratamento dos dados. A *Conclusão* retoma os principais resultados, relevantes para a administração pública. Essas partes puderam ser elaboradas por membros que não são tão fluentes para manipulação dos dados com R.

Os capítulos de *Desenvolvimento* focaram nos grupos de variáveis que foram estabelecidos no questionário aplicado: Grupo A, B, C, ..., I; sendo que os dois primeiros abrangem a porção domiciliar da pesquisa, que representa a etapa finalizada até setembro de 2024. O grupo **B** aborda as características da estrutura dos domicílios e o grupo **C** aborda as características dos arredores dos domicílios. Essa estruturação abriu caminho para a futura análise das variáveis relativas aos moradores (grupo **D** em diante), uma etapa consideravelmente mais complexa e que demandará maior esforço da equipe.

Conclusão

Este trabalho demonstra a flexibilidade e o poder do R, seja utilizando o R base, seus pacotes, ou integrando as ferramentas mais recentes desenvolvidas pela Posit, como o *Quarto Book*. A aplicação dessas ferramentas para a análise da PDAD Rural demonstrou a eficácia dessas ferramentas na realização de projetos complexos com reprodutibilidade, essencial para a continuidade das políticas públicas no serviço público.