

Estudo de Caso: Distribuição de Livros Didáticos no Quênia

Tópicos Especiais em Métodos Computacionais
Prof. Ricardo Ceneviva

Esta lista de exercícios foi elaborada com base no estudo de caso apresentado por Glewwe, Kremer e Moulin (2009), que avalia os impactos da distribuição de livros didáticos nas escolas públicas do Quênia sobre o desempenho dos estudantes em testes padronizados.

Os dados utilizados nos exercícios foram simulados com base nos resultados reportados no artigo. O objetivo desta lista é aplicar os conceitos discutidos no guia 'Modelos Lineares em R' em um contexto empírico realista, seguindo uma estrutura de estudo de caso.

Instruções:

- Todos os exercícios devem ser realizados utilizando o software R.
- Utilize os pacotes `'tidyverse'`, `'car'`, `'lmtest'`, `'sandwich'`, `'ggplot2'`, `'AER'`, `'glmnet'` quando necessário.
- Documente o seu código e interprete todos os resultados obtidos.

I. Introdução aos Modelos Lineares

- 1.1 Carregue a base de dados simulada 'kenya_books_sim.csv'.
- 1.2 Explore os dados: número de escolas, número de estudantes por escola, distribuição de notas.
- 1.3 Crie gráficos de dispersão entre o número de livros por aluno e a nota média na prova padronizada.
- 1.4 O que se espera da relação entre livros por aluno e desempenho médio? Justifique teoricamente.

II. Pressupostos Fundamentais da Regressão Linear

- 2.1 Estime uma regressão linear simples entre nota média e livros por aluno.
- 2.2 Verifique os resíduos do modelo: são homocedásticos? são normalmente distribuídos?
- 2.3 Teste a presença de autocorrelação nos resíduos.
- 2.4 Calcule o VIF para verificar multicolinearidade em um modelo com múltiplas variáveis.

III. Implementando Regressão Linear no R

9. 3.1 Estime um modelo de regressão múltipla com nota média como variável dependente e:
 10. - livros por aluno
 11. - nível médio de escolaridade dos pais
 12. - presença de alimentação escolar
 13. - número de professores por aluno
14. 3.2 Interprete os coeficientes estimados.
15. 3.3 Gere gráficos de diagnóstico usando a função `plot(modelo)`.

IV. Procedimentos Diagnósticos para Modelos Lineares

16. 4.1 Aplique o teste de Breusch-Pagan (`bptest`) e Durbin-Watson (`dwtest`).
17. 4.2 Identifique observações influentes com Cook's Distance.
18. 4.3 Reestime o modelo excluindo essas observações e compare os resultados.

V. Além da Regressão Linear Básica: Regularização e GLMs

19. 5.1 Estime um modelo de regressão Lasso usando o pacote `glmnet`.
20. 5.2 Compare os coeficientes com o modelo linear anterior.
21. 5.3 Estime um modelo GLM (família Poisson) com o número de estudantes aprovados como variável dependente.

VI. Referências e Continuidade

22. 6.1 Consulte os livros recomendados para aprofundar a discussão sobre regularização e GLMs.
23. 6.2 Proponha uma extensão do estudo com base em um novo desenho de política pública.