



Universidade de Brasília  
Departamento de Estatística

**A eficiência dos gastos públicos educacionais no DF e determinantes,  
usando análise envoltória de dados com bootstrap.**

**Carlo Aleksandr Rosano de Almeida**

Projeto apresentado para o Departamento  
de Estatística da Universidade de Brasília  
como parte dos requisitos necessários para  
obtenção do grau de Bacharel em Es-  
tatística.

**Brasília  
2022**

**Carlo Aleksandr Rosano de Almeida**

**A eficiência dos gastos públicos educacionais no DF e determinantes,  
usando análise envoltória de dados com bootstrap.**

Orientador: Prof. Dr. Eduardo Monteiro de Castro Gomes

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2022**

## Sumário

<b>1</b>	<b>Introdução . . . . .</b>	<b>4</b>
<b>2</b>	<b>Objetivos. . . . .</b>	<b>6</b>
<b>3</b>	<b>Justificativa . . . . .</b>	<b>7</b>
<b>4</b>	<b>Metodologia . . . . .</b>	<b>8</b>
<b>5</b>	<b>Cronograma . . . . .</b>	<b>9</b>
<b>6</b>	<b>Referências . . . . .</b>	<b>11</b>

# 1 Introdução

O desenvolvimento socioeconômico de um país dificilmente pode ser sustentado sem uma população instruída. A partir dos inovadores trabalhos de Theodore Schultz e Gary Becker (laureados com o prêmio nobel em economia) na década de 1950, várias vertentes da teoria econômica consideram o investimento em capital humano como um fator essencial no crescimento econômico e na resiliência em situações de crises Uzawa [1] e Lucas [2]. Para estes economistas, o capital humano é tão importante quanto o capital financeiro no crescimento do PIB (Produto Interno Bruto) e define-se como o conjunto de conhecimentos, habilidades e atitudes que as pessoas possuem e as tornam aptas para desenvolver atividades específicas. Além disso, para Louchner [3] está comprovado que níveis mais altos de escolaridade em um país estão associados a uma menor incidência criminal e corruptiva, maior salubridade da população, a níveis mais alto de produtividade e empregabilidade nacional, entre outros .

Diante dessas constatações, nas ultimas décadas o Brasil vem realizando altos investimentos na educação. Segundo dados do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) o investimento público em relação ao PIB na educação básica cresceu de 3.7% em 2000 para 4.8% em 2018 [4]. Este valor representa um percentual maior do que a media dos países desenvolvidos, segundo o relatório Education at a Glance (EaG), divulgado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) [5].

No entanto, o investimento na educação, em geral, mostra resultados insatisfatórios no que se refere a qualidade do ensino básico, tornando o gasto publico ineficiente. Um dos maiores estudos sobre educação do mundo, o Programa Internacional de Avaliação de Estudantes (Pisa), em sua edição 2018, indica que o Brasil possui baixa proficiência em Leitura, Matemática e Ciências, quando comparado com outros 78 países que participaram da avaliação. Revela índices de desempenho estagnados desde 2009, indicando que 68,1% dos estudantes brasileiros, com 15 anos de idade, não possuem nível básico de Matemática, considerado como o mínimo para o exercício pleno da cidadania. Em Ciências, o número chega 55% e em Leitura, 50%. Quando comparado com os países da América do Sul analisados pelo Pisa, o Brasil é o pior país em Matemática e Ciências empatado estatisticamente com a Argentina. Já em Leitura o país é o segundo pior, atrás da Argentina e Peru que se encontram empatados [6].

Para superar este desafio, os tomadores de decisão e gestores escolares precisam de convincentes respostas às seguintes questões:

1. É possível aumentar a qualidade e o desempenho escolar com os escassos recursos disponíveis?

2. Em que medida a educação brasileira pode ser mais eficiente?
3. Quais são os fatores determinantes da eficiência dos gastos públicos em educação?

Em torno destas questões está o problema central de pesquisa desse trabalho.

Para abordar este problema, a Estatística e a Pesquisa Operacional apresentaram grandes contribuições nas últimas décadas, desenvolvendo técnicas para análise da eficiência e de seus fatores determinantes. Dentre elas, dois métodos de benchmark e fronteira eficiente mostraram-se relativamente úteis:

- Stochastic Frontier Analysis (SFA).
- Data Envelopment Analysis (DEA).

O primeiro é chamado de paramétrico e usa técnicas econométricas a partir de uma relação funcional entre insumos e produtos. O segundo é denominado de não-paramétrico e determinístico e utiliza técnicas de programação matemática. Diferenciam-se em função da forma de construir a fronteira eficiente e como são interpretados os desvios da fronteira. Ambos os métodos apresentam vantagens e desvantagens, mas nenhuma é claramente superior à outra.

Para superar o determinismo do modelo DEA, uma nova abordagem tem ganhando importância na literatura na última década. Trata-se dos trabalhos de Simar e Wilson [7] que estimam índices robustos de eficiência com bootstrap em dois estágios. No primeiro, calcula-se índices de eficiência DEA somente com as variáveis controláveis do processo produtivo. No segundo, utiliza-se o modelo econométrico de regressão truncada (ou outros modelos equivalentes) para regredir os escores de eficiência obtidos contra as variáveis ambientais, não discricionárias. Esse novo método enquadra-se no denominado DEA estocástico (ou semi-paramétrico), que busca integrar o termo aleatório contido nos métodos do tipo SFA à fronteira calculada por métodos do tipo DEA, sem precisar da especificação, a priori, da relação funcional entre os insumos e produtos.

## 2 Objetivos

Neste contexto, o presente estudo tem como objetivo estimar escores robustos de eficiência DEA com bootstrap em dois estágios para o ensino médio das escolas do DF. Nesse sentido, trabalha-se com dados do Censo Escolar 2020 e do Enem 2019 (Exame Nacional do Ensino Médio), ambos elaborados pelo INEP. No primeiro estágio, utilizaram-se os insumos e produtos clássicos do setor. Neste estágio, técnicas bootstrap serão utilizadas para detectar outliers, estimar o viés, corrigir os escores e testar o tipo de retornos de escala. No segundo estágio, variáveis contextuais relacionadas a indicadores de desenvolvimento socioeconômico da família serão inseridas para explicar a eficiência escolar com o uso de modelos de regressão truncada tobit.

### **3 Justificativa**

Acredita-se que os resultados aqui encontrados fornecerão novas contribuições práticas para subsidiar a definição de estratégias e ações que melhorem o retorno dos investimentos educacionais. O presente estudo se justifica ao indicar melhorias para uma gestão escolar mais eficiente e ao determinar fatores contextuais que induzem a adequação escolar almejada.

## 4 Metodologia

O método usado será a Data Envelopment Analysis (DEA) com bootstrap em dois estágios que terá como objeto de estudo as escolas do DF do ensino médio. No primeiro estágio, utilizaram-se os insumos e produtos clássicos cotejados na literatura para o estudo do setor. Neste estágio, técnicas bootstrap serão utilizadas, com ajuda da função R `boot.sw98` do package FEAR [8], para detectar outliers, estimar o viés, corrigir os escores e testar o tipo de retornos de escala. No segundo estágio, variáveis contextuais relacionadas a indicadores de desenvolvimento socioeconômico da família do aluno serão inseridas para explicar a eficiência escolar com o uso de modelos de regressão truncada tobit. Desta forma, testa-se de forma consistente quais são as variáveis não controláveis pelos gestores influentes e significativas, seu sinal (positivo ou negativo) e permite ainda corrigir os índices de eficiência com base nos parâmetros estimados.

Simar e Wilson [7] sugerem dois algoritmos que incorporam o bootstrap em um modelo de regressão truncado no segundo estágio. Usando um experimento de Monte Carlo, os autores examinam e comparam o desempenho desses dois algoritmos e demonstram que ambos superam os métodos de regressão convencionais (tobit e regressões truncadas sem bootstrap). Para amostras com menos de 400 unidades o Algoritmo 1 proposto se ajusta melhor do que o Algoritmo 2, que é mais eficiente a partir de amostras que excedem 800 unidades. Como o tamanho de nossa amostra não supera os 400, se adaptará melhor o Algoritmo 1.

Além disso, será elaborada uma plataforma interativa, através do pacote Shiny do software R [9], que permita aos gestores escolares verificar seu desempenho, identificar seus bechmarking e fazer simulações de melhorias.

As unidades de análise são as escolas de nível médio do DF. Para cada uma dessas unidades escolares serão utilizadas os tipos de variáveis do modelo: entradas, saídas e contextuais. Estas serão extraídas do Censo Escolar 2020 com base em 31 de maio e do Enem (Exame Nacional do Ensino Médio) realizado em 3 e 10 de novembro de 2019.



## 5 Cronograma

As atividades a serem desenvolvidas durante o Trabalho de Conclusão de Curso são:

1. Escolha do tema a ser abordado.
2. Desenvolvimento da proposta de projeto.
3. Entrega da proposta de projeto.
4. Revisão de literatura.
5. Elaboração da apresentação da proposta.
6. Apresentação oral da proposta.
7. Manipulação do banco de dados.
8. Análise exploratória do banco de dados.
9. Elaboração do relatório parcial.
10. Entrega do relatório parcial ao Professor Orientador(a).
11. Correção do do relatório parcial.
12. Entrega do relatório parcial para a banca.
13. Desenvolvimento do modelo.
14. Elaboração do relatório final.
15. Entrega do relatório final ao Professor Orientador(a).
16. Correção do do relatório final.
17. Entrega do relatório final para a banca.

Tabela 1: Cronograma

	1/2022				2/2022			
	Fevereiro	Março	Abril	Maiο	Junho	Julho	Agosto	Setembro
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								

## 6 Referências

- [1] Hirofumi Uzawa. “Optimum technical change in an aggregative model of economic growth”. Em: *International economic review* 6.1 (1965), pp. 18–31.
- [2] Robert E Lucas Jr. “On the mechanics of economic development”. Em: *Journal of monetary economics* 22.1 (1988), pp. 3–42.
- [3] Lance Lochner. “Non-production benefits of education: Crime, health, and good citizenship”. Em: (2011).
- [4] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Percentual do investimento total em relação ao Produto Interno Bruto (PIB), por nível de ensino – 2000 a 2018*. Mar. de 2022. URL: [https://download.inep.gov.br/informacoes\\_estatisticas/investimentos\\_publicos\\_em\\_educacao/indicadores\\_financeiros\\_educacionais/Investimento\\_pib\\_total.zip](https://download.inep.gov.br/informacoes_estatisticas/investimentos_publicos_em_educacao/indicadores_financeiros_educacionais/Investimento_pib_total.zip).
- [5] “Panorama da Educação : destaques do Education at a Glance 2019”. Em: *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (2019).
- [6] OECD. *PISA 2018 Results*. Mar. de 2022. URL: [https://www.oecd.org/pisa/publications/PISA2018\\_CN\\_BRA.pdf](https://www.oecd.org/pisa/publications/PISA2018_CN_BRA.pdf).
- [7] Leopold Simar e Paul W Wilson. “Estimation and inference in two-stage, semi-parametric models of production processes”. Em: *Journal of econometrics* 136.1 (2007), pp. 31–64.
- [8] Paul W. Wilson. *FEAR: A Software Package for Frontier Efficiency Analysis with R*. R package version 3.0.0. 2006. URL: <https://pww.people.clemson.edu/Software/FEAR/fear.html>.
- [9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: <https://www.R-project.org/>.