

Eficiência das Escolas Públicas Estaduais de Minas Gerais: considerações acerca da qualidade do ensino.

Victor Maia Senna Delgado

Cedeplar/UFMG

Ana Flávia Machado

Cedeplar/UFMG

RESUMO

O presente trabalho desenvolve o método semi-paramétrico denominado *Two-stage Data Envelopment Analysis* para detectar a fronteira de eficiência das escolas públicas estaduais de Minas Gerais nos níveis fundamental e médio. A primeira etapa desse processo consiste em calcular a eficiência pela DEA e a segunda em comparar os resultados de eficiência por meio de uma regressão com variáveis de *background* familiar, de infra-estrutura e dotação das escolas. As bases censitárias do Sistema Mineiro de Avaliação da Educação Pública (SIMAVE), que aplica os exames de proficiência de Matemática e Português para 4ª e 8ª séries do fundamental e 3ª série do ensino médio, e do Sistema Informacional de Custo Aluno (SICA), além do Censo Educacional MEC/INEP de 2005 permitiram aplicar uma das primeiras análises de eficiência por escola desenvolvida para o ensino básico do país. Os resultados encontrados sugerem que uma complementaridade dos insumos, dentro e fora da escola, possibilita o seu melhor desempenho. Escolas localizadas nas mesoregiões do estado onde há mais abundância de recursos educacionais possuem chance maior de serem mais eficientes e prestarem um ensino de maior qualidade. Porém, existem bons exemplos de desempenho em regiões mais carentes e, em termos gerais, os resultados do produto educacional do estado podem melhorar bastante se se conseguir um maior nível de eficiência para as escolas estaduais.

Palavras-Chave: Eficiência-Técnica, Ensino Básico, DEA-bootstrap, Proficiência e Infra-Estrutura escolar.

Classificação JEL: I21, C14, C24, D61

Área ANPEC: Área 11 – Economia Social e Demografia Econômica.

ABSTRACT

This paper develops the semi-parametrical two-stage Data Envelopment Analysis to detect the efficiency frontier of public Brazilian Minas Gerais state schools. The first stage computes the DEA-efficiency, the second one inputs the DEA results to a truncated-regression with background and endowment variables. The new data basis of school proficiency, SIMAVE, and cost-measurement, SICA, either the Educational Census of Education Ministry 2005, permitted this new application of efficiency measurement for schools. The results shows that input-complementarity **in** and **out** school contribute to greater performance. Best located schools, where the educational inputs are better, are more efficient in probability. However, there are some good examples of efficient schools emerging from poor regions and, in general terms, all stated educational results can improve considerably if the schools get to improve their efficiency.

Key-Words: Technical-Efficiency, Elementary School, DEA-bootstrap, Proficiency, Infrastructure.

JEL Classification: I21, C14, C24, D61

Eficiência das Escolas Públicas Estaduais de Minas Gerais: considerações acerca da qualidade do ensino.

1 – INTRODUÇÃO.

Garantir qualidade e eficiência é uma das preocupações recentes da política educacional do país. Nas últimas décadas, o Brasil conquistou algumas melhorias nos indicadores do seu quadro educacional como o aumento da escolaridade média da população, a diminuição da evasão escolar e do trabalho infantil e uma maior cobertura do ensino fundamental, hoje bastante próximo do objetivo da universalização.

Apesar da favorável evolução dos indicadores, os exames internacionais e a confrontação no mercado de trabalho mostram que a formação do nosso estudante está aquém do previsto quando o comparamos com os de outros países em desenvolvimento. As evidências apontam que a qualidade da educação brasileira é menor em relação aos países desenvolvidos e em desenvolvimento. O impacto de uma menor qualidade implica que um ano a mais no ensino fundamental no Brasil corresponde a um tempo menor de formação nos outros países. Outro motivo é a ineficiência do sistema educacional, já que se gasta cerca de 4,5% do PIB e os resultados, ainda assim, são decepcionantes.

No âmbito dessa constatação, emergem questões referentes à eficiência da provisão de serviços em educação. A alocação de recursos faz parte de um dos desafios com o qual uma sociedade deve lidar em seu dia a dia. Por ser tão importante a distribuição de recursos, é preciso que seu uso se dê de forma a promover o máximo de benefício social possível. E, no caso do sistema educacional, essas restrições são reforçadas, uma vez que educação formal não é um bem qualquer.

Esse artigo busca avaliar a eficiência nas escolas públicas estaduais mineiras, empregando o método não paramétrico da *Data Envelopment Analysis* (DEA). O método DEA permite identificar as escolas que se dempenham melhor em termos de custo-oportunidade identificando as que oferecem maior aprendizado aos alunos dados os recursos disponíveis. A escolha de Minas Gerais se justifica pelo fato do estado refletir o contexto nacional, retratando a diversidade do processo de desenvolvimento socioeconômico brasileiro. Uma outra razão é a disponibilidade de base de dados. Têm-se, em caráter censitário, tanto uma base de dados de custo-aluno de escolas públicas estaduais quanto uma base de avaliação do ensino, ambas desenvolvidas pela Secretaria Estadual da Educação.

Na seção a seguir, revisamos a literatura econômica sobre educação e o seu emprego na análise de eficiência por meio da análise envoltória. Na terceira parte, descrevemos o método DEA-clássico e o recente método *DEA-bootstrap*. Na 4ª seção, apresentamos as três bases de dados principais empregadas, SIMAVE, SICA e Censo Escolar, e a construção das variáveis do modelo. Na quinta parte, analisamos os resultados dos modelos de eficiência deste artigo. Na última seção, tecemos alguns comentários finais.

2 - DESEMPENHO EDUCACIONAL E EFICIÊNCIA NAS ESCOLAS.

A eficiência econômica pretende alcançar o produto máximo dado determinado volume de recursos. Ou então, estabelecida uma meta para o produto, como conseguí-la com um gasto mínimo. No âmbito da educação, a eficiência está associada à qualidade do ensino, uma vez que esse atributo permite às crianças e jovens, já adultos, serem mais produtivos e, quiçá, socialmente integrados. Nesse contexto, a eficiência da educação incorpora componente intergeracional ao contrário do estabelecido na definição estrita.

Na linguagem da eficiência econômica a construção teórica importante é a função de produção ou fronteira de eficiência na educação, formulada pela primeira vez por Coleman et ali (1966). Objetivando encontrar evidências de que o desempenho dos alunos negros aquém do esperado era

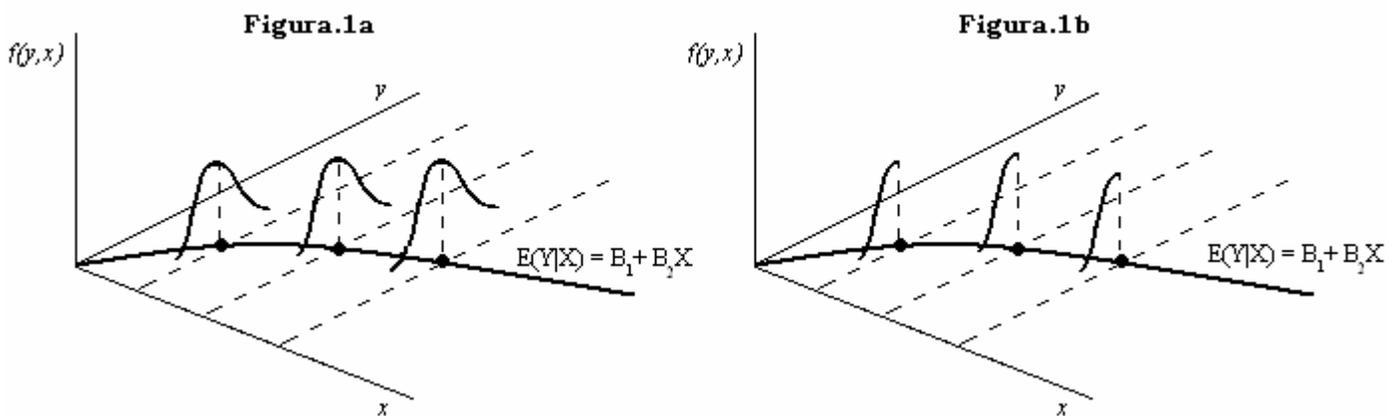
devido a uma insuficiência de insumos em suas escolas, Coleman et ali (1966) agrupam os dados de educação em uma função de produção para a escola:

$$y_i = f(x_i, z_i, d_i) \quad (1)$$

Onde y_i é o rendimento dos alunos da escola ‘ i ’; x_i são os insumos educacionais supervisionados pela direção da escola, z_i são características individuais dos alunos e das famílias dos alunos daquela escola (*background* familiar) e d_i são variáveis de ‘dotação’.

No entanto, usar a função de produção para medir eficiência ou ineficiência requer uma interpretação diferente sobre a teoria. Os trabalhos até aqui mencionados consideram a função de produção como um *fit* para os pontos médios da distribuição da variável dependente. A partir da equação (1), os modelos que tomam o “ y ” como produto, por exemplo, acreditam que o ajustamento correto da função passa pela média condicionada $E(Y|X)$. Ocorre, então, que o nível de produto observado pode estar tanto acima quanto abaixo da função, como se observa na figura 1a.

Figura 1. Estimação da Fronteira Condicionada com Erros Normais e Erros Normais-Truncados.



Implicitamente, o que os modelos postos dessa forma supõem é que, em média, as observações são eficientes, ou seja, situam-se sobre a fronteira. Considerando-se o longo prazo, não há nenhum problema em se fazer tal afirmação. No contexto macroeconômico onde há concorrência, é difícil se supor que uma firma permaneça ineficiente por anos a fio. Nessa situação a firma entraria em falência e não a teríamos como observação em uma série temporal.

Para análise microeconômica de firmas ou de unidades gestoras do governo (como é o caso das escolas), é razoável supor uma ineficiência que pode permanecer por um determinado período de tempo. Porém, para se incorporar ineficiência a uma função de produção, são necessárias algumas alterações nos pressupostos estatísticos aplicados à função a ser estimada.

É possível visualizar como a função descrita na equação (1) incorpora ineficiência econômica, basta apenas acrescentar um θ_i , índice de eficiência, multiplicado à função. Como o índice de eficiência varia de 0 a 1, temos a possibilidade de incorporar ineficiência a uma função de produção. O valor de θ_i inferior à unidade indica que a observação produz aquém do que é possível.

$$y_i = \theta_i f(x_i, z_i, d_i) \quad (2)$$

Uma maneira de obtenção da fronteira é a estimação não-paramétrica, que não utiliza erros aleatórios. A construção é feita por uma fronteira envoltória de dados, método conhecido como *Data Envelopment Analysis* (DEA), desenvolvido por Charnes, Cooper e Rhodes (1978). Os autores argumentam que a DEA pode ser um procedimento bastante útil para a análise da eficiência no setor público, uma vez que, *ao policy maker*, interessa a hierarquização de unidades de análise por determinado conjunto de insumos/produtos e não a análise dos determinantes dessa hierarquização.

A partir de então, seguiu-se uma vasta literatura aplicada a este setor, a começar com os próprios autores Charnes et al (1981), que fizeram uma das primeiras aplicações ao caso da educação.

Recentemente, dois trabalhos de aplicação da DEA à eficiência da educação se destacam: Wilson (2005) e Afonso e Aubyn (2005). Em comum, os dois estudos apresentam as novas incorporações do método DEA para tornar os índices de eficiência mais robustos. As aplicações utilizam dados de proficiência do PISA – “*Program for International Student Assessment*”¹ – e variáveis não-discricionárias como status sócio-econômico e escolaridade dos pais. Wilson incorpora todas as variáveis em uma só estimação da DEA, enquanto que Afonso e Aubyn desenvolvem o método de dois estágios da maneira antiga e com as novas incorporações.

No Brasil, são poucos os estudos sobre eficiência utilizando a DEA e, ainda em menor número, os que se aplicam à educação. Porém, podemos destacar o trabalho de Façanha e Marinho (1999) aplicado às instituições do Ensino Superior brasileiro, de Sampaio de Sousa e Ramos (1999), trabalho que se concentram na eficiência dos gastos públicos municipais em geral, e Faria e Januzzi (2006), sobre eficiência de gastos na área de educação e saúde dos municípios do Rio de Janeiro.

3 – MÉTODO DEA: DATA ENVELOPMENT ANALYSIS.

Os modelos de medição de eficiência utilizando a DEA ganharam novo fôlego a partir da segunda metade da década de 90 e, principalmente nos anos 2000, com as incorporações de Gibels et al. (1999) Kneip et al. (2003) e Simar e Wilson (1998, 2002 e 2007). No entanto, seu referencial teórico inicial surge em Debreu (1951). A partir desse marco teórico, Farrel (1957) constrói o novo método não paramétrico de mensuração da eficiência por programação linear.

A vantagem dos modelos DEA não-paramétricos é sua flexibilidade, tais modelos assumem poucas hipóteses sobre o comportamento dos dados e por isso não implicam nenhuma forma funcional *a priori* para a fronteira de educação. Captam a melhor prática existente da organização produtiva e fornecem um *benchmark* para as instituições analisadas. Estão bem fundamentados teoricamente, com base apenas nos axiomas mais fracos da teoria econômica, e podem aplicar mais de um produto ao mesmo tempo em uma estimação. As desvantagens do modelo advêm também de sua não-parametricidade: a convergência é lenta, o que para pequenas amostras pode ser um fator limitador, já que os indicadores que serão obtidos poderão ser inconsistentes; também por esse motivo os modelos possuem um limite do número de variáveis a ser incluído um pouco mais estreito.

A abordagem moderna da DEA define um **conjunto de possibilidades de produção** P :

$$P = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \text{ pode produzir } \mathbf{y}\}; P \subset \mathfrak{R}^{S+M} \quad (3)$$

Onde \mathbf{x} e \mathbf{y} fazem parte de dois conjuntos de vetores de variáveis observadas que são independentes e identicamente distribuídas (*iid*). Em nosso caso, \mathbf{x} é o vetor de *inputs* $\mathbf{x}_{iM} = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iM})$, o subscrito ‘i’ identifica a unidade de análise, são $i = 1, 2, \dots, N$ observações e o subscrito ‘M’, o número de *inputs* diferentes, $M = 1, 2, \dots, M$ *inputs*; \mathbf{y} é o vetor de *outputs*, $\mathbf{y}_{iS} = (\mathbf{y}_{i1}, \dots, \mathbf{y}_{iS})$ e ‘S’ identifica o número de *outputs*, $S = 1, 2, \dots, S$. Dessa forma, temos que P define a tecnologia educacional e a partir desta tecnologia obtemos a função de produção estabelecida na equação (1) da seção anterior: $\mathbf{y} = f(\mathbf{x}_i)$, porém, aqui representada em termos vetoriais e ainda sem as “*environment variables*”.

O conjunto de variáveis observadas é definido como $L_N = \{(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i, \mathbf{d}_i)\}$ onde além de ‘ \mathbf{x} ’ e ‘ \mathbf{y} ’, temos os vetores com as variáveis de *background* familiar, \mathbf{z}_i , e o vetor das variáveis de *dotação*, \mathbf{d}_i . De P surge um processo gerador de dados que delimita uma **fronteira de eficiência** a ser captada

¹ Exame aplicado entre os 28 países membros da OCDE mais quatorze países não-membros: Albânia, Argentina, Brasil, Bulgária, Chile, Hong Kong – China, Indonésia, Israel, Letônia, Liechtenstein, Macedônia, Peru, Rússia e Tailândia no ano 2000.

pela DEA ou pelo método de fronteiras estocásticas.² O método não paramétrico da DEA estabelece que a fronteira será construída somente com os pontos que atingiram o máximo de produto (*outputs*) dado determinado nível de insumos (*inputs*) ou com o mínimo de inputs para dado nível de outputs. Ou seja, a fronteira é delimitada pela melhor prática no esquema de produção em voga. A construção de tal fronteira exige um método de programação linear que trabalha com o seguinte processo:

$$\theta_i = \theta_i(\mathbf{x}_i, \mathbf{y}_i | P) \equiv \max \{ \theta_i | (\mathbf{x}_i, \theta \mathbf{y}_i) \in P, \theta_i > 0 \} \quad (4)$$

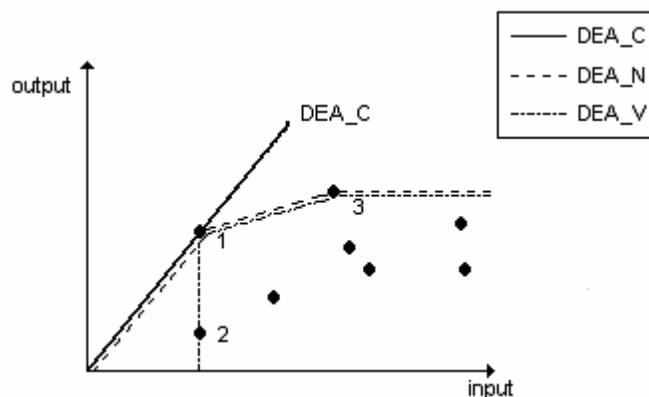
$$sa. P(\mathbf{x}_i, \mathbf{y}_i | C, D) = \left\{ (\mathbf{x}_i, \mathbf{y}_i) \mid \theta_i \mathbf{y}_i \leq \sum_{i=1}^n \lambda_i \mathbf{y}_i, \mathbf{x}_j \geq \sum_{i=1}^n \lambda_i \mathbf{x}_i, \lambda \in \mathfrak{R}^n, \mathbf{y} \in \mathfrak{R}^s, \mathbf{x}_i \in \mathfrak{R}^m \right\} \quad (5)$$

A equação (4) revela um procedimento de maximização onde θ_i é o **índice de eficiência-Farrel**, medido de 0 a 1, sendo a unidade o indicador de eficiência máxima, situação onde a observação está sobre a fronteira. A eficiência pode recair sobre os insumos no que é chamada de *input-oriented*, $\theta_i \mathbf{x}_i$, ou sobre os produtos, *output-oriented*, $\theta_i \mathbf{y}_i$, forma que foi exposta acima.

A equação (5) reúne as restrições para a maximização presente em (4). Há alguns novos elementos na equação (5), pois ao definirmos a fronteira de produção precisamos estabelecer duas propriedades microeconômicas. A primeira é a de retornos constantes de escala, indicada por ‘C’, e a segunda é a livre disponibilidade de insumos (*Free Disposal*), ‘D’. O λ_i é um vetor de intensidade, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) \in \mathfrak{R}_+^N$, os λ 's denotam os pesos que possibilitam a construção de uma fronteira convexa, outro pressuposto microeconômico, o da possibilidade de combinação convexa de fatores.

Assim, por meio de programação linear, é possível construir as fronteiras de eficiência representadas na figura 2 abaixo. Cada uma das fronteiras indicadas se assenta em um tipo de rendimentos escala e na livre disponibilidade de insumos e convexidade de fatores. A partir de alterações na equação de restrição (5) é possível obter dois outros índices além do de rendimentos constantes (DEA-C): o índice de rendimentos não-crescentes, DEA-N, e o índice de rendimentos variáveis, DEA-V. Observamos que o envelope da DEA-V é o mais maleável, envolve os dados a uma distância menor do que as outras duas medidas, isso se refletirá no índice-Farrel de eficiência que, por construção, possui média maior que a DEA-C. As três formas de medição são úteis para a obtenção dos rendimentos de escala das observações eficientes.³ As fronteiras estão representadas no \mathfrak{R}^2 (um *output*, um *input*) por simplificação.

Figura 2. Construção das Fronteiras de Eficiência DEA.



² Na linguagem estatística, todos os dados que observamos na realidade possuem por trás um invisível processo gerador de dados (Data Generating Process – DGP, do inglês). Portanto, o DGP é um processo gerador de dados conhecido.

³ Pode-se obter os rendimentos das observações não-eficientes a partir de sua projeção na fronteira. A escala é obtida comparando-se os índices das observações com o nível de produto necessário para deixarem de ser ineficientes.

Os pontos representam as observações (x_i, y_i) e no caso da DEA-C, a fronteira será sempre uma reta, um plano ou hiperplano. A hipótese de retornos constantes pode ser alterada para a possibilidade de retornos não-crescentes, ‘N’, onde uma nova restrição $\sum \lambda_i \leq 1$ não permite expansões radiais por um escalar ‘ α ’ maior do que 1, ou alterada para a possibilidade de rendimentos variáveis de escala, ‘V’, obtidos por meio da adição da restrição $\sum \lambda_i = 1$, a soma dos escalares está “presa” à unidade, indicando que a fronteira será construída aos pares de pontos eficientes. Na DEA-V os pontos se reportam sempre a um segmento de fronteira ($\lambda_i + \lambda_j = 1$, onde $i \neq j$).

A fronteira DEA-V possui ao menos um ponto em comum com a fronteira DEA-C, como o ponto 1 exemplificado na figura 2. Isso indica que quando $\theta_{\text{IDEA-V}} = \theta_{\text{IDEA-C}}$, tem-se rendimentos constantes de escala (RCE). Se não são iguais, cabe verificar duas outras possibilidades: se $\theta_{\text{IDEA-V}} = \theta_{\text{IDEA-N}}$, variável e não crescente, o que indica rendimentos decrescentes de escala (RDE); ou se $\theta_{\text{IDEA-V}} \neq \theta_{\text{IDEA-C}} \neq \theta_{\text{IDEA-N}}$, caso onde resta apenas a possibilidade de retornos crescentes (RCC).⁴

A construção dos índices é feita radialmente a partir da origem e em relação à fronteira, trata-se da distância radial entre o ponto observado e a fronteira de eficiência. Como observamos, duas formas de construção são possíveis: a eficiência *input-oriented* e a *output-oriented*. Neste trabalho, utilizaremos o índice de eficiência medido pelo critério de Sheppard (1970) orientado pelo produto (*output-oriented*) e denotado pelo δ_i . Não há muita diferença entre o **Sheppard-eficiência** e o **Farrel-eficiência**, apenas a construção, um caracteriza-se por ser a inversa do outro índice.⁵

A real técnica de produção P é observada por meio de uma realização estocástica denominada \hat{P} . Para obter uma distribuição dos dados realizados em \hat{P} que se aproxime do P -real é empregado o *bootstrap* proposto em Simar e Wilson (1998). A partir da técnica de *bootstrap* proposta, é possível obter uma série de estimativas mais confiáveis para os índices de eficiência, assim como construir intervalos de confiança, impossíveis de serem obtidos sem a nova técnica.

3.1. O novo método de dois estágios.

O método de dois estágios emprega duas etapas de estimação: a primeira é a mensuração não-paramétrica da DEA abordada na seção anterior; e a segunda é a estimação dos parâmetros das variáveis que podem influir nos índices de eficiência obtidos no primeiro estágio, com o uso de MQO ou outra técnica. A regressão mais simples dos modelos de dois-estágios envolve o índice de eficiência δ_i e as variáveis do ambiente, destacadas aqui como z_i :

$$\delta_i = \beta z_i + \varepsilon_i \quad (6)$$

O β_i é um vetor de parâmetros empregados para captar a influência de z_i sobre o índice estimado. Muitas variáveis externas ao controle gerencial da escola impactam em sua eficiência, é praticamente impossível incorporar todas ao modelo DEA, primeiramente porque a presença de muitas variáveis dificultaria a convergência dos estimadores e, segundo, porque optamos por deixar no primeiro modelo apenas variáveis gerenciais-pedagógicas que estão sob o controle das escolas ou das superintendências de ensino. Os ε_i 's são erros da regressão, onde $E(\varepsilon_i) = 0$ e $E(\varepsilon_i^2) = \sigma^2$.

Como não há erros aleatórios na obtenção da DEA, os índices de eficiência estão de alguma forma correlacionados entre si, impedindo estimarmos uma equação como a apresentada em (6). Para superar esse problema, recorreremos à aplicação dos modelos *bootstrap*. Assim, observamos o comportamento da fronteira não apenas com a base de dados inicial, mas capturamos, por meio

⁴ Usa-se também as siglas CCR e BCC para referir ao modelo de retornos constantes e variáveis de escala. É uma homenagem à Charnes Cooper e Rhodes (1978) e Banker Charnes e Cooper (1984), autores que propuseram cada uma das formas de medição.

⁵ A adoção do **Sheppard-eficiência** em *output-oriented* baseia-se no fato de que nessa forma de medição podemos comparar o índice de eficiência a uma medida de capacidade que varia de 0 a 100%. Dessa forma, uma observação eficiente ($\delta_i = 1$) está produzindo 100% da sua capacidade presumida. Observe que, por ser a sua inversa, para obtermos o mesmo efeito de interpretação com *input-oriented*, teríamos de usar o **Farrel-eficiência**.

desta, sua distribuição provável e conseguimos a consistência dos estimadores, os intervalos de confiança e a aplicação dos testes de hipóteses. Quando se obtêm uma mesma fronteira várias vezes por *bootstrap*, retira-se a correlação existente nos δ_i 's, exclui-se também a parte aleatória do erro que pode ocasionar uma ineficiência.⁶ Cada observação permanece, portanto, apenas com a parte que confere ineficiência devido ao mau uso dos fatores. A equação (6) é rescrita da seguinte maneira:

$$\delta_i^* = \beta \mathbf{z}_i + \varepsilon_i \quad (7)$$

Além das considerações acima, por ser determinístico, o método DEA está sujeito a alterações bruscas da eficiência devido a observações *outliers*, chamadas **super-eficientes**. A ocorrência destes casos desloca toda a fronteira e prejudica a análise das demais observações. Neste trabalho atentamos para os métodos de correção de *outliers* propostos por Sampaio de Sousa e Stosic (2005) e Cazals et al. (2002).

4 – FONTE DE DADOS E TRATAMENTO DAS VARIÁVEIS.

O emprego de dados educacionais sobre o desempenho e custo por aluno em Minas Gerais é um dos pontos relevantes deste artigo. Nesse estudo, estão combinadas três fontes de dados. Os dados do Sistema Mineiro de Avaliação da Educação Pública (SIMAVE) permitem a abordagem de aspectos ligados à qualidade do ensino. A base do Sistema Informacional Custo Aluno (SICA), por sua vez, permite tratar de aspectos associados à eficiência. O Censo Educacional do MEC/IBGE possibilita construir variáveis de infra-estrutura e oferta de serviços educacionais. Essas três bases visam compor as variáveis de “outputs”, “inputs”, “dotação” e “*background*” presentes nos modelos.

O SIMAVE é um sistema de avaliação que tem como base o PROEB – Programa de Avaliação da Rede Pública da Educação Básica. A cada ano, os alunos das turmas de 4^a e 8^a série da rede pública estadual, assim como os alunos do 3^o ano do ensino médio, realizam provas em uma das disciplinas básicas do currículo. Dessa forma, em 2002, os estudantes fizeram prova de Língua Portuguesa e, em 2003, de Matemática. A comparação das notas é possível pelo critério de Resposta ao Item, método que permite a identificação do desempenho dos alunos por item específico e possibilita a comparação das notas entre os diversos anos do ciclo educacional. Assim sendo, podemos comparar a evolução das médias por série avaliada.⁷

A utilização de duas provas ao invés de uma é justificada pela avaliação de que dois exames são mais representativos do que um. Credita-se isso ao fato de se evitar, com os dois exames, oscilações muito bruscas de média a que estão sujeitas as escolas com turmas muito pequenas.

Tabela 4.1. Resumo dos resultados das provas por anos de Ensino.

	Média	Desvio	Min	Max
<i>4^a. Série:</i>				
Matemática	183.73	25.22	106.72	284.93
Português	186.72	22.86	53.25	249.86
<i>8^a. série:</i>				
Matemática	237.49	20.24	164.00	315.56
Português	239.85	13.90	164.06	296.77
<i>3^o. ano:</i>				
Matemática	270.07	20.05	218.13	340.88
Português	269.92	12.12	208.84	331.81

Fonte: SIMAVE (2002 e 2003).

⁶ Dados sobre variáveis econômicas apresentam discrepância estatística (erro), nesses casos uma observação pode ser declarada eficiente ou ineficiente. O uso da DEA-*bootstrap* atenta para o problema, dada a distribuição do DGP descrita, os estimadores de eficiência-*bootstrap* passam a ter um componente aleatório e não tão determinístico para sua distribuição (SIMAR E WILSON, 2007).

⁷ Sobre o critério de Resposta ao Item ver Soares e Pereira, 2002.

Tão importantes quanto os exames de proficiência são as informações do questionário sócio-econômico das famílias aplicado pelo SIMAVE. A tabela 4.2 a seguir resume as variáveis utilizadas.

Tabela 4.2. Variáveis do Questionário do SIMAVE.

Variáveis para *background* (segundo estágio):

cor	1- brancos e amarelos, 0 pretos e pardos.
esc_res_fem	1- responsável do sexo feminino tem 8ª série ou mais, 0 – caso contrário.
livros	1- mais do que 20 livros em casa, 0 caso contrário.
computador	1- computador na residência, 0 caso contrário.
ler_liv_inf	1- se o aluno leu livros de histórias infantis, 0 caso contrário.
ler_jornais	1- se o aluno leu jornais durante o ano, 0 caso contrário.
ntrab_fora	1- não trabalha fora de casa, 0 caso trabalhe.
ntrab_doméstico	1- aluno não exerce tarefas domésticas, 0 caso gaste algum tempo.
devermat	1- se faz sempre o dever de matemática, 0 caso contrário.
reprovado	1- não reprovou de ano nenhuma vez, 0 caso tenha reprovado.
noite	1- aluno estuda à noite, 0 caso contrário.

Fonte: SIMAVE, 2003.

A segunda fonte de dados desta pesquisa é o SICA – Sistema Informacional Custo Aluno. O SICA fornece informações de gasto por aluno mensal através da coleta de informações sobre o financiamento dos vários níveis de ensino: Infantil, Fundamental e Médio. O sistema foi desenvolvido pela Superintendência de Planejamento da Secretaria do Estado de Minas Gerais em 1997 e faz parte do Sistema de Informações sobre Orçamentos Públicos em Educação – SIOPE.⁸

O SICA reúne o gasto por aluno dentro das divisões dos chamados custos “**dentro**” e custos “**fora**” da escola. Os custos “dentro” são compostos em **diretos**, custos com profissionais da educação e convênios, e **indiretos**, no qual entram os custos com o pessoal administrativo e as despesas gerais. Os custos “fora” são aqueles que não estão direcionados ao custeio do pessoal ativo e tampouco às despesas escolares. Neste trabalho utiliza-se apenas o custo dentro da escola, visto que se enquadra melhor à análise e se refere a gastos diretamente ligados à nossa unidade de observação.

Por fim, a terceira fonte de dados é o Censo Escolar de 2003, realizado desde 1998 pelo INEP/MEC (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) As variáveis do censo estão separadas, nesse artigo, em duas categorias: as agrupadas e as não agrupadas. A primeira categoria engloba os modelos DEA no primeiro estágio. A segunda categoria é a das variáveis não-agrupadas, utilizadas na maneira em que estão disponíveis no censo e empregadas na etapa do segundo estágio, com exceção das matrículas, que fazem parte dos *outputs* do primeiro estágio.⁹

⁸ Sistema recém implementado pelo INEP: <http://www.siope.inep.gov.br/apresentacao.do>

⁹ Os termos “agrupadas” e “não-agrupadas” a que nos referimos dizem respeito a maneira como os dados estavam disponíveis no Censo Escolar. “Agrupadas” se referem a dados que se constituíam mais de uma variável no censo e foram reunidas em uma só categoria, Não-agrupadas são apresentadas de acordo com a sua disposição no Censo e representam apenas uma variável.

Tabela 4.3. Variáveis agrupadas e não-agrupadas do Censo Escolar.

Variáveis Agrupadas (1º. Grupo):	
<i>qdocente</i>	Número de professores com 3º Grau Licenciatura completa para 4ª 8ª série e 3º ano respectivamente.
<i>salas</i>	Número de salas de aula de toda escola.
<i>Infra</i>	Índice de 0 a 5 que verifica presença de sanitário; energia pública; água, esgoto e coleta periódica do lixo.
<i>Vcm</i>	Variável obtida por <i>Principal Component Analysis</i> para captar concomitantemente presença de aparelhos de “vídeo”, “computadores” e recursos de “multimídia”.
Variáveis não-agrupadas (2º. Grupo):	
<i>urbana</i>	1- se escola se situa em área urbana, 0 caso contrário.
<i>matriculas</i>	numero de alunos matriculados separado por período: 1a4ª série, 5a8ª série e ensino médio.
<i>Meso</i>	<i>dummy</i> para mesoregião do estado
<i>Fnde</i>	1- escolas que recebem o Fundo Nacional de Desenvolvimento da Educação, 0 caso contrário.
<i>mere_esc</i>	1- escola oferece alimentação aos alunos.

Fonte: Censo Escolar MEC/INEP, 2003.

Os dados do SIMAVE nos mostram que, para o ensino fundamental a minoria é composta por alunos brancos, 41% na 4ª série, 45%, na 8ª série, alternando para uma pequena maioria, 51%, no terceiro ano. Há, também, no terceiro ano, a presença significativa de alunos estudando a noite, 69%. Quase todos os indicadores sócio-econômicos se destacam negativamente, com a notável exceção da proporção de leitura de livros (e a leitura de jornais para o caso específico do 3º ano). No ensino fundamental, os indicadores positivos são a grande parcela de alunos não trabalhando fora de casa, 84% e 64%, para 4ª e 8ª séries, respectivamente, além do número de alunos que faz o dever de matemática (71% e 55%, respectivamente, na 4ª e 8ª série, e que não são reprovados 73% na 4ª e 67% na 8ª série). A tabela 4.4 reúne informações das escolas públicas estaduais mineiras.

Tabela 4.4. Resumo das informações do Questionário do SIMAVE.

Variável	4ª. Série	8ª. Série	3ª. Ano
<i>cor</i>	0.41	0.45	0.51
<i>esc_resp_fem</i>	0.31	0.19	0.20
<i>livros</i>	0.22	0.24	0.26
<i>computador</i>	0.13	0.11	0.11
<i>ler_livro_inf</i>	0.78	0.70	0.71
<i>ler_jornais</i>	0.44	0.45	0.62
<i>ntrab_fora</i>	0.84	0.64	0.47
<i>ntrab_domestico</i>	0.28	0.26	0.25
<i>devermat</i>	0.71	0.55	0.42
<i>nreprovado</i>	0.73	0.67	0.63
<i>noite</i>	-	-	0.69

Fonte: SIMAVE 2003.

No que tange à variável de custo, após a retirada das escolas *outliers*, a variável *cdentro* apresenta ainda uma cauda direita bastante acentuada, entretanto, as análises demonstraram que não seria preciso um tratamento maior do que o já dado. Abaixo temos a apresentação na tabela 4.5 para estas variáveis.

Tabela 4.5. Tabela com resumo dos Custo-Aluno Dentro por Série.

	Média	Desvio	Min	Max
4ª. Série:	56.49	23.38	9.06	329.33
8ª. série:	52.33	17.72	4.11	180.60
3º. ano:	41.02	17.20	1.34	169.85

Fonte: SICA 2005.

No Censo Escolar, a variável “condições de infra-estrutura (Infra)” é a soma de cinco variáveis dummy: *sani_den*, observa se o sanitário está dentro da escola; *ener_pub*, capta se a escola é ligada à rede pública de energia elétrica; *agua_pub*, verifica se a escola possui sistema de água da rede pública, *esg_pub*, se a escola é ligada ao esgotamento da rede pública, *lixo_col*, se o destino do lixo é coleta periódica da prefeitura. Na tabela 4.6 nota-se que algumas escolas não possuem nenhuma condição de infra-estrutura, sendo que é crescente, ao longo de cada série, o número das que apresentam algumas ou todas, além disso, observa-se que as condições de infra-estrutura também evoluem com o avanço das séries.

Tabela 4.6. Tabela com resumo para variável de Infra-estrutura.

	Frequência (%)						Média
	0	1	2	3	4	5	
4ª. Série:	0.53	3.46	9.85	8.18	13.79	64.20	4.24
8ª. série:	0.08	1.22	5.29	6.40	14.93	72.07	4.51
3º. ano:	0.00	0.14	1.62	4.30	13.88	80.06	4.72

* O indicador 0 denota a completa ausência dos itens avaliados, 5 é o total.

Fonte: Censo Escolar MEC/IBGE 2003. Variável ‘Infra’ construída, ver tabela 4.3.

As variáveis *Apvideo*, *Comp&Imp* e *multimidia* foram construídas pelo método de PCA – *Principal Component Analysis*. “*Apvideo*” é o número de aparelhos relacionados à televisão, uma variável quantitativa obtida através da soma entre *vvideo*, *vtv* e *parabol*, quantidade de vídeo-cassetes, televisores e antenas parabólicas respectivamente. “*Comp&Imp*” é variável que indica o número de aparelhos de informática que a escola possui. “*Multimidia*” capta os recursos dentro da escola disponíveis aos alunos. Abarca a soma das variáveis originais: “*biblioteca*”, “*videoteca*”, *saltvvid* (Sala de Tv e Vídeo), *sala_lei* (Sala de Leitura) e *lab_info* (laboratório de informática).

Do grupo das variáveis desagregadas destaca-se *matriculas*, variável que é utilizada como indicador de *output*. Para o caso das matrículas é preciso um cuidado semelhante ao que foi abordado na parte de custos devido à dispersão considerável dos valores dessa variável, distribuição assimétrica à direita, com o mínimo de cinco alunos em uma pequena escola do primeiro ciclo até um máximo de 4078 na maior escola em termos de matrícula no ensino médio. As médias de matrículas se elevam ao longo das séries e o desvio padrão aumenta consideravelmente, o que pode ser observado na tabela 4.7. As informações estão apresentadas por ciclo, matrículas de 1ª a 4ª série foi utilizada para a DEA 4ª série, matrículas 5ª a 8ª série, para a DEA 8ª e *matric_médio* é o total das 1ª, 2ª e 3ª séries do Ensino Médio, empregada na análise da 3ª série.

Tabela 4.7. Resumo do número de matrículas por ciclos do Ensino.

	Média	Desvio	Min	Max
Matric_1a4 (4ª. Série):	247.64	193.38	5	1675
Matric_5a8 (8ª. Série):	419.98	252.27	8	1406
Matric_médio (3º. ano):	524.38	411.25	45	4078

Fonte: Censo Escolar MEC/IBGE 2003.

Tabela 4.8. Resumo das informações do Censo escolar, variáveis não agrupadas Todas as séries.

Variável	4ª. Série	8ª. Série	3º. Ano
urbana	0.79	0.89	0.99
fnde	0.81	0.79	0.76
merenda	0.96	0.95	0.94
TemTV	0.95	0.98	0.99
TemPC	0.58	0.68	0.74
TemMult	0.83	0.89	0.94
sani_den	0.94	0.97	0.98
ener_pub	0.98	1.00	1.00
agua_pub	0.83	0.91	0.96
esg_pub	0.71	0.78	0.85
lixo_col	0.77	0.86	0.93

Fonte: Censo Escolar MEC/IBGE 2003.

Nota-se a maior urbanização de acordo com o nível de ensino, a menor necessidade de auxílio do FNDE (Fundo Nacional de Desenvolvimento da Educação), ligeira queda na merenda escolar e melhoria geral nos indicadores de infra-estrutura e *VCM* (*Tem TV*, *Tem PC*, *TemMult*, que indicam apenas se a escola tem ou não os aparelhos e recursos que constam em *VCM*).

Para os acontecimentos e características que afetam o resultado, mas não estão sobre o controle nem das escolas nem da Secretaria, é que se compõe o método de dois estágios que conta com a presença da média das características dos alunos e condições a que as escolas estão sujeitas, algumas delas podem, inclusive, ser alteradas por meios de políticas públicas. O segundo estágio do modelo inclui as variáveis *background* que são: dummy urbana (controle), cor, série correta, escolaridade da mãe, livros em casa, internet em casa, ler livros ou jornais, não trabalhar fora de casa, não desempenhar trabalho doméstico, realizar o dever de casa, não ser reprovado e, para o caso do ensino médio, estudar à noite. As variáveis de *dotação* são: urbana (controle), merenda escolar, auxílio FNDE, presença de computadores e impressoras, recursos multimídia, sanitários, ligada à rede pública de energia, água e esgoto e se possui coleta de lixo urbana periódica. Cada uma das colunas, *background* e *dotação*, compõem um tipo diferente de regressão. O regressando será o índice de eficiência (invertido) obtido no primeiro estágio.

5 - RESULTADOS.

O modelo de dois estágios empregado nesta seção implementa todas as técnicas recentes enunciadas nas seções anteriores. No primeiro estágio, os resultados são descritos para as doze mesoregiões de Minas Gerais em cada uma das séries. No segundo estágio, estenderemos o alcance dos principais resultados encontrados relativos à eficiência da escola estadual mineira.

5.1. Primeiro Estágio.

Neste primeiro estágio, o índice de eficiência de cada escola é calculado considerando três variáveis de output: proficiência em Matemática, proficiência em Português e Matrículas. Além de mais cinco variáveis de insumos: custo-aluno, docentes com superior, número de salas, infra-estrutura e aparelhamento escolar. O índice DEA-eficiência resume as variáveis e revela aspectos que seriam difíceis de serem captados por uma análise exaustiva da relação insumo produto solucionados um a um. Assim, regiões que a princípio poderiam se sair bem no *ranking* de notas não se apresentam como eficientes quando levados em conta os seus insumos, enquanto que outras que não revelavam desempenho excepcional acerca da proficiência se saem bem, considerando-se o baixo nível de

insumos que possuem. A configuração de eficiência no estado de Minas Gerais é descrita por intermédio de suas doze mesoregiões.

Para entendermos melhor as propriedades da eficiência por escolas, e a construção de suas médias por mesoregiões, analisamos a eficiência para cada uma das séries. A tabela 5.1 fornece a média de produtos e insumos para cada uma das doze regiões do estado no nível da 4ª série.

Tabela 5.1. Média dos Produtos e Insumos por Mesoregião 4ª série.

Código	Mesoregião	Escolas	Produtos			Insumos				
			<i>Pmate</i>	<i>Pport</i>	Alunos	<i>Custo</i>	<i>Qdoc</i>	Salas	<i>Infra</i>	<i>Vcm</i>
310001	Noroeste de Minas	54	187.43	188.94	274.91	57.01	4.89	21.09	4.48	20.35
310002	Norte de Minas	411	165.86	167.49	195.89	43.47	2.15	15.74	3.20	19.74
310003	Jequitinhonha	191	171.89	175.24	218.36	49.80	2.34	16.23	3.72	19.45
310004	Vale do Mucuri	98	164.25	176.36	223.64	58.49	3.57	18.02	4.06	19.61
310005	Triângulo Mineiro	192	194.11	198.61	292.97	66.44	7.44	23.03	4.77	20.76
310006	Central Mineira	55	190.02	191.64	199.91	63.34	3.38	19.29	4.55	20.26
310007	Metropolitana	527	188.63	192.56	340.25	55.64	5.17	23.21	4.70	19.80
310008	Vale do Rio Doce	303	180.28	183.17	188.15	62.72	3.43	15.88	4.05	20.05
310009	Oeste de Minas	90	204.56	206.15	221.52	78.72	5.92	20.10	4.58	20.50
310010	Sul/Sudoeste de Minas	183	200.12	197.40	281.28	66.16	6.19	21.26	4.82	20.66
310011	Campo das Vertentes	61	201.63	199.92	223.57	75.29	5.44	18.85	4.31	20.00
310012	Zona da Mata	293	188.75	193.28	208.58	64.97	4.78	18.34	4.53	19.79
310013	Minas Gerais	2458	183.73	186.72	247.64	58.41	4.36	19.21	4.24	19.98

Fonte: Elaboração própria a partir das bases do SICA 2003, SIMAVE 2003 e Censo Escolar 2003.

Constatamos que, via de regra, em Minas Gerais, os custos por aluno aumentam à medida que se deslocam para o sul do estado. As regiões desenvolvidas do centro-sul possuem médias maiores tanto para o nível de insumos como o de produtos, enquanto que, nas mesoregiões pobres do Norte, o nível de insumos e produtos é precário. Percebemos, portanto, que os resultados dos *outputs* estão de alguma forma relacionados diretamente com o nível de *inputs*, mas a determinação de eficiência considera a **melhor prática** dentro de uma escala de rendimentos, ou seja, eficiência é o resultado de um bom balanceamento entre insumos e produtos.

Podemos dividir o *ranking* entre as seis primeiras e seis últimas mesoregiões. O critério para a classificação é a média das eficiências. Logicamente, o uso de médias esconde muitas nuances dentro de cada uma das mesoregiões. O *ranking* de eficiência é sensível ao tipo de índice adotado, no que se refere ao método de construção, pode variar de forma significativa.

O *ranking* aqui empregado é a DEA *bootstrap* (DEA_c*) proposto por Simar e Wilson (1998). A eficiência é corrigida por *smooth-bootstrap* de maneira que se obtêm o intervalo de confiança para o índice e uma distribuição mais confiável após a aplicação da técnica. Outro *ranking* foi obtido a partir do estimador “*jackstrap*” de Sampaio de Sousa e Stosic (2005), denominado por DEA_ss. O índice Sampaio-Stosic detecta *outliers* e confecciona novos índices após descartar as observações super-influentes. Após a exclusão dos *outliers*, a mesoregião que mais se beneficiou foi a Norte de Minas, passando para o primeiro grupo. As escolas *outliers* prejudicavam a comparação da região como um todo.¹⁰

O terceiro *ranking* construído foi o de número de escolas eficientes por total de escolas pertencentes à mesoregião. A Região Metropolitana é a que possui o maior número de escolas

¹⁰ Embora qualquer escola eficiente do estado possa servir de comparação para todas as outras, independente da região específica, a proximidade geográfica traz semelhança nos *inputs* e *outputs*. Com isso, a localização mais provável de influência de um *outlier* ocorre nas escolas do seu entorno.

eficientes, entretanto, é também a que possui maior fração da população e mais escolas, por isso o quociente **nº eficientes/total de escolas** apresenta um resultado relativo. Observa-se que a Zona da Mata é a que mais perde posições em relação ao *rank1*. O contrário ocorre com a mesoregião Norte de Minas, a região possui uma razão elevada de escolas eficientes sobre o total de escolas.

A última coluna da tabela 5.2 a seguir resume o número de escolas eficientes com retornos decrescentes de escala, útil para ver a incidência de rendimentos decrescentes em cada mesoregião. O número de eficientes varia de acordo com o uso da DEA-C, DEA-N ou DEA-V. Por construção, há sempre mais escolas eficientes na DEA-V do que na DEA-C, a primeira é mais flexível, aborda qualquer tipo de variação da escala. Neste trabalho, o DEA-C é rejeitado apenas para a 3ª série do Ensino Médio. No entanto, a construção dos três índices é importante para captar as escolas por grupo de retornos de escala. A tabela 5.3 posteriori apresenta essas informações em maiores detalhes.

Tabela 5.2. Rankings das mesoregiões para a 4ª série.

Código	Mesoregião	Escolas	DEA_c*	Rank1	DEA_ss	Rank2	quocient	Rank3	Decresc
310009	Oeste de Minas	90	0.654	1	0.757	4	0.033	6	2
310007	Metropolitana	527	0.650	2	0.767	1	0.047	4	6
310011	Campo das Vertentes	61	0.650	3	0.747	3	0.082	1	1
310012	Zona da Mata	293	0.647	4	0.747	6	0.017	11	1
310006	Central Mineira	55	0.646	5	0.744	7	0.055	2	3
310003	Jequitinhonha	191	0.641	6	0.755	5	0.016	12	0
310001	Noroeste de Minas	54	0.629	7	0.731	10	0.019	10	0
310010	Sul/Sudoeste de Minas	183	0.629	8	0.732	9	0.027	7	3
310008	Vale do Rio Doce	303	0.623	9	0.734	8	0.040	5	6
310002	Norte de Minas	411	0.623	10	0.760	2	0.054	3	4
310005	Triângulo Mineiro	192	0.621	11	0.724	11	0.021	8	2
310004	Vale do Mucuri	98	0.607	12	0.701	12	0.020	9	0
310013	Minas Gerais	2458	0.635		0.748		0.037		28

Fonte: Construção dos índices de eficiência por elaboração própria a partir de variáveis do 1º estágio.

Tabela 5.3. Distribuição das escalas de rendimento por Mesoregião:

Código	Mesoregião	Escolas	Constante	Decrescente	Crescente
310001	Noroeste de Minas	54	0.00%	98.15%	1.85%
310002	Norte de Minas	411	5.35%	89.54%	5.11%
310003	Jequitinhonha	191	5.26%	89.47%	5.26%
310004	Vale do Mucuri	98	5.10%	91.84%	3.06%
310005	Triângulo Mineiro	192	2.09%	97.38%	0.52%
310006	Central Mineira	55	3.57%	91.07%	5.36%
310007	Metropolitana	527	3.99%	94.87%	1.14%
310008	Vale do Rio Doce	303	3.63%	92.74%	3.63%
310009	Oeste de Minas	90	2.30%	97.70%	0.00%
310010	Sul/Sudoeste de Minas	183	2.69%	96.77%	0.54%
310011	Campo das Vertentes	61	5.00%	91.67%	3.33%
310012	Zona da Mata	293	4.41%	94.58%	1.02%
310013	Minas Gerais	2458	3.13%	94.34%	2.52%

Fonte: Construção dos retornos de escala por elaboração própria.

A tabela acima confirma a intuição inicial: as regiões que possuem melhores níveis de insumo, melhores condições de infra-estrutura e condições sócio-econômicas acima da média do

estado apresentam mais escolas referentes aos retornos **decrecentes de escala**, enquanto que as regiões mais carentes têm maior tendência a apresentar proporções mais elevadas de escolas com rendimentos **constantes** e **crecentes**. A Região Metropolitana e Oeste de Minas, por serem mais populosas e apresentarem as maiores escolas, possuem um dos maiores níveis de escolas com rendimentos decrecentes, ao passo que as regiões do Jequitinhonha e Norte de Minas, por serem o oposto das primeiras regiões, possuem muitas escolas com rendimentos crescentes.

Para a 8ª série temos poucas modificações em relação ao quadro de eficiência apresentado anteriormente. Cabe ressaltar que é a serie com maior número de escolas na esfera estadual e que o ganho nas médias das notas é significativo (mais de 50 pontos). Com exceção de algumas mudanças de posição um pouco mais drásticas para a mesoregião Sul/Sudoeste que subiu para segundo lugar no *rank1* e a Metropolitana que caiu para nono, o *ranking* da 8ª série corresponde ao da 4ª série. Para entender a mudança na posição relativa da Metropolitana é preciso observar o *rank2*. Nessa segunda classificação, a região volta ao grupo dos seis primeiros, porque, na oitava série, a Região Metropolitana possui relativamente muitas escolas eficientes (*rank3*): são 14 escolas na parte decrescente dos ganhos de escala (*decresc*).¹¹

Entre os primeiros lugares, a meso de Campo das Vertentes merece novamente destaque, continuou na terceira posição. O Oeste de Minas sai da primeira posição, mas fica em 4º, continuando no grupo dos seis primeiros. Sobre as variáveis de Campo das Vertentes e Oeste de Minas ocorre um efeito de “transbordamento”, pois os indicadores destas duas mesoregiões muitas vezes não são tão melhores quanto os da Metropolitana, mas são mais uniformes. Por isso, apesar de não terem os picos das notas, seus resultados são eficientes. Observamos que no *rank2* essas duas regiões não se saem tão bem, isso se deve ao fato de o *ranking* DEA_ss ser menos suave.

Por outro lado, o desempenho é mais fraco, em termos de eficiência, na mesoregião do Vale do Mucuri, novamente em 11º, no Noroeste de Minas e no Triângulo Mineiro. As duas primeiras apresentam resultados precários em termos de *outputs*, a última possui um nível de insumos elevado, mas resultados fracos, ainda não condizentes com sua quantidade de insumos.

Tabela 5.4. Ranking das Mesoregiões para a 8ª série.

Código	Mesoregião	Escolas	DEA_c*	Rank1	DEA_ss	Rank2	quocient	Rank3	Decresc
310003	Jequitinhonha	163	0.766	1	0.964	2	0.043	4	2
310010	Sul/Sudoeste de Minas	259	0.765	2	0.942	5	0.039	6	8
310011	Campo das Vertentes	70	0.762	3	0.903	8	0.043	5	1
310009	Oeste de Minas	117	0.759	4	0.902	10	0.026	10	1
310012	Zona da Mata	292	0.758	5	0.880	7	0.034	8	4
310006	Central Mineira	62	0.758	6	0.867	3	0.032	9	1
310002	Norte de Minas	347	0.756	7	0.866	1	0.069	1	3
310008	Vale do Rio Doce	307	0.755	8	0.864	6	0.046	3	5
310007	Metropolitana	557	0.753	9	0.858	4	0.050	2	14
310005	Triângulo Mineiro	226	0.743	10	0.853	9	0.035	7	4
310004	Vale do Mucuri	71	0.738	11	0.825	11	0.014	11	0
310001	Noroeste de Minas	60	0.737	12	0.819	12	0.000	12	0
310013	Minas Gerais	2531	0.755		0.844		0.043		43

Fonte: Construção dos índices de eficiência por elaboração própria a partir de variáveis do 1º estágio.

¹¹ A correção do *bootstrap* confere um intervalo maior para os eficientes trazendo sua correção para baixo. O modelo Sampaio-Stosic (DEA_ss) mostra, também, uma queda de posições para esta região mas não tão grande quanto a do *bootstrap*.

Na terceira série, há algumas mudanças mais importantes. A primeira delas diz respeito à mudança do índice DEA, principal a ser utilizado. Sob o teste de rendimento de escala de Banker (1993) e Simar e Wilson (2002), rejeitamos a hipótese de rendimentos constantes para a função de produção educacional. Indicando que a análise ideal é feita pelo índice de retornos variáveis DEA-V. Esse fenômeno está relacionado à segunda mudança principal para o 3º ano: o sistema de ensino médio é bem mais complexo que o das séries anteriores, uma vez que a média de eficiência DEA-C é extremamente baixa (0.534). Portanto, conjecturamos que se faz necessária uma abordagem mais ampla. O uso da DEA-V satisfaz em parte esse anseio. Por outro lado, o uso de mais indicadores de *output* e *inputs* seria bem vindo para tratar dessa maior complexidade, no entanto, reforçamos que a base do 3º ano possui menos escolas (1419) e é mais heterogênea do que as anteriores.

Quanto ao *ranking*, percebemos uma grande relação com o *ranking* anterior da 8ª série, em parte, isso acontece por conta de muitas escolas fornecerem simultaneamente o segundo ciclo do fundamental e o ensino médio, mas não necessariamente a classificação de uma mesma escola em diferentes séries é idêntica, ou mesmo próxima, depende do comportamento de todas as oito variáveis do modelo. A Central Mineira se destaca positivamente pulando para a primeira posição nos dois primeiros *rankings*. Sul/Sudoeste, Campo das Vertentes e Zona da Mata continuam se situando bem. O Triângulo Mineiro é a novidade, aparecendo pela primeira vez no grupo dos seis primeiros, e a Metropolitana e o Jequitinhonha não estão mais no primeiro grupo, mas ocupam a 5ª e 6ª posições no *rank2*, enquanto que o Triângulo Mineiro cai para 9º neste *ranking*.

As regiões de fraco desempenho se repetem: Vale do Rio Doce e Mucuri, Norte e Noroeste de Minas. Para o caso específico do Norte de Minas cabe uma interessante ressalva: em todas as tabelas de classificação, apesar de estar situada entre as últimas posições para o *rank1*, a mesoregião se sai bem nos *ranks* 2 e 3, esse fato ocorre porque há algumas escolas boas na região norte mas, apesar disso, a grande maioria é de escolas pobres de muito fraco desempenho.

Tabela 5.5. Ranking das Mesoregiões para a 3ª série.

Código	Mesoregião	Escolas	DEA_v*	Rank1	DEA_ss	Rank2	quocient	Rank3	Decresc
310006	Central Mineira	39	0.897	1	0.920	1	0.077	2	2
310010	Sul/Sudoeste de Minas	169	0.895	2	0.914	3	0.047	7	8
310009	Oeste de Minas	74	0.892	3	0.914	4	0.014	11	1
310011	Campo das Vertentes	53	0.889	4	0.906	8	0.038	8	2
310012	Zona da Mata	184	0.888	5	0.910	7	0.049	5	8
310005	Triângulo Mineiro	136	0.883	6	0.902	9	0.051	4	6
310007	Metropolitana	352	0.878	7	0.912	5	0.063	3	16
310003	Jequitinhonha	62	0.877	8	0.912	6	0.048	6	2
310008	Vale do Rio Doce	146	0.873	9	0.896	11	0.027	10	3
310004	Vale do Mucuri	32	0.868	10	0.898	10	0.031	9	1
310002	Norte de Minas	138	0.867	11	0.917	2	0.101	1	3
310001	Noroeste de Minas	34	0.864	12	0.885	12	0.000	12	0
310013	Minas Gerais	1419	0.903		0.903		0.052		52

Fonte: Construção dos índices de eficiência por elaboração própria a partir de variáveis do 1º estágio.

5.2. Segundo Estágio.

No segundo estágio é realizada uma regressão dos índices de eficiência obtidos para cada escola considerando-se as *environment variables* e as variáveis de ‘dotação’. Essa etapa permite identificar as variáveis que afetam o nível de eficiência, porém, não estão sobre o controle da gerência das escolas.

Foram realizadas seis regressões pelo modelo de dois estágios, o índice de eficiência empregado precisou ser alterado para o inverso do índice anterior ($1/\delta_i$). O intervalo para esta transformação se situa entre de 1 a ∞ . O valor 1 continua designando a eficiência, porém, a escala percorre sentido inverso da anterior, quanto maior o índice menor a eficiência. A interpretação para os coeficientes das regressões ocorre, portanto, da seguinte maneira: se o parâmetro é positivo o resultado se afasta da eficiência, se negativo, aproxima-se. As regressões (1) e (2) na tabela 5.6 a seguir são estimações das normais truncadas obtidas por máxima verossimilhança (MV), já que as variáveis regredidas possuem o limite inferior da na eficiência no valor igual a 1. As duas primeiras regressões seguem o algoritmo#1 proposto em Simar e Wilson (2007) e utilizado também em Afonso e Aubyn (2005).

A primeira regressão utiliza as informações do questionário SIMAVE como variáveis primárias. As informações se referem à turma de alunos por escolas e, na regressão, estão devidamente controladas para escolas urbanas ou rurais. Na segunda regressão, as variáveis são os indicadores de dotação presentes nos dados do censo escolar, onde também há correção urbano-rural. Estas duas primeiras regressões de cada série servem de “guia”, pois são mais rigorosas na construção dos intervalos dos testes de hipóteses e rejeitam a hipótese nula com menor frequência do que as regressões truncadas ou regressões tobit geralmente empregadas na DEA dois estágios. Na quarta série, com exceção de ‘*cor*’, ‘*internet*’ e ‘*ntrab_fora*’, todas as variáveis significativas apresentam a mesma direção das regressões (1) e (2).

Os sinais das variáveis significativas se apresentam na direção esperada. A surpresa é ‘*urbana*’, o coeficiente da regressão (1) de cada série é positivo, indicando que se a escola é urbana, essa tende a ser uma escola menos eficiente em probabilidade.¹² Esse coeficiente pode ser interpretado em termos de eficiência técnica no seguinte sentido: as escolas rurais possuem, em média, menos recursos do que as localizadas nas cidades, entretanto, muitas delas não são seriamente reprimidas em seus resultados por conta disso, ou seja, no quesito das notas, as rurais não perdem tanto quanto seria de se esperar quando se leva em conta que possuem relativamente menos insumos, ou, por outro lado, as urbanas não estão fazendo jus às melhores condições e não se saem relativamente melhor do que as rurais. A última interpretação ganha reforço, porque, ao controlarmos as regressões pelas condições de infra-estrutura, menores no meio rural (regressão 2), o coeficiente negativo passa indicar que as escolas urbanas apresentam melhor desempenho.

Outra variável de destaque é a escolaridade da responsável feminina (*esc_resp_fem*). É, em termos absolutos, o maior coeficiente e confirma toda a literatura que aponta que as condições familiares importam para o rendimento das crianças. Ter livros em casa (*livros*) aponta melhores condições socioeconômicas dos alunos. As eficientes possuem alunos com maior acervo em sua biblioteca particular. Embora não significativa em (1), o sinal dessa variável é esperado e assinala-se como significativa a 1% em regressões estimados na maneira clássica, como se verifica, também, para a 8ª série. Não podemos dizer o mesmo para o caso dos alunos que não trabalham fora (*ntrab_fora*), apesar de nas demais regressões o coeficiente também não se apresentar significativo a direção se altera e, para 4ª série, é contrária a que é esperada pela intuição.

Quanto a algumas variáveis de infra-estrutura, os resultados parecem corroborar interpretações de que muitas das práticas pedagógicas empregadas nas escolas não conseguem aproveitar muito bem os recursos disponíveis (Oliveira e Schwartzman 2002). O número de aparelhos de TV, “*apvideo*”, evidencia esta questão: várias das escolas ineficientes possuem televisão, em Minas Gerais, TV não é um recurso escasso para as escolas, tanto que um número

¹² Se nos deslocarmos do meio rural para o urbano, maior será a proporção de escolas ineficientes nesse meio em relação às rurais.

considerável de escolas da 4ª série possui 10 ou mais televisores (120 escolas).¹³ O resultado aponta é que as escolas eficientes não tendem a ter tantos televisores. O mesmo ocorre para “*multimid*”, ou seja, presença de biblioteca, videoteca, Sala de Tv e Vídeo, Sala de Leitura e laboratório de informática não contribuem para o desempenho escolar.

Os resultados da etapa de dois estágios com os mesmos modelos para as seis regressões são bastante similares para o caso da 8ª série. *Cor* se apresenta negativa, indicando que quanto maior proporção de alunos brancos, maior a eficiência, fator explicado também pela disparidade das condições familiares. A importância do computador é reforçada para 8ª série, sendo significativa a 1% nesse método de Simar e Wilson (2007) que é mais rigoroso.

No modelo de dois estágios da 3ª série o índice de eficiência empregado no *regressando* é o DEA-V. Poucas variáveis das duas primeiras regressões conseguiram atingir a convergência da Máxima Verossimilhança. No modelo (1) temos somente *urbana*, que deixou de ser significativa (quase totalidade das escolas do ensino médio são urbanas) e *esc_resp_fem* que é significativa a 1%, indicando que maior escolaridade dos pais colabora com a eficiência dos alunos e da escola. Computadores deixam de ser significativos em (2).

A nova variável acrescentada à 3ª série é *noite*, que indica a porcentagem de alunos que estudam à noite em determinada escola. Ao contrário das séries anteriores, para a 3ª série, essa variável é uma medição importante, já que grande parte dos alunos deste nível (70%) estuda neste turno. As escolas que ofertam aulas no horário noturno estão mais propensas à ineficiência, o que parece refletir muito as condições dos alunos que estudam à noite, com menores disponibilidades de horário, dupla jornada e precariedade das condições sociais, infelizmente, não atingiram a convergência para os modelos rigorosos do algoritmo Simar e Wilson (2007).

¹³ Maior até do que as 49 escolas sem nenhum aparelho e, de fato, 13 dessas escolas sem nenhum televisor foram definidas como eficientes, em parte, por se saírem bem mesmo com a completa ausência de um dos recursos que se revela em um baixo *VCM* que, na maioria dos casos, vêm acompanhado de baixa infra-estrutura. No entanto, sete daquelas 13, são, na verdade, *outliers*.

Tabela 5.6. Resultados das Regressões dois estágios todas as séries.

	<i>4ª Série</i>		<i>8ª Série</i>		<i>3ª Série</i>	
	Regressão (1)	Regressão (2)	Regressão (1)	Regressão (2)	Regressão (1)	Regressão (2)
constante	1.238*** (0.05)	0.253** (0.12)	0.444*** (0.08)	1.202*** (0.01)	0.829*** (0.13)	-2.147 0.01
urbana	0.136*** (0.05)	-0.216*** (0.01)	0.370*** (0.05)	0.019*** (0.00)	0.111 (0.12)	
fnde	0.071 (0.09)		0.034 (0.11)			
mere_esc	0.001 (0.10)		0.168 (0.14)			
cor	0.025 (0.13)		-0.164 (0.21)			
esc_resp_fem	-0.385** (0.16)		-0.414* (0.21)		-0.531*** (0.00)	
livros	-0.189 (0.18)		-0.191 (0.30)			
internet	-0.143 (0.12)		0.093 (0.15)			
ler_livro	-0.041 (0.10)		-0.135 (0.14)			
ler_jornais	-0.088 (0.13)					
ntrab_fora	0.092 (0.12)		-0.062 (0.14)			
ntrab_domestico	0.053 (0.11)					
devermat	-0.087 (0.11)		-0.014 (0.16)			
nreprovado	0.137*** (0.00)		0.121*** (0.00)			
apvideo		0.120*** (0.01)		0.011*** (0.00)		
comp_imp		-0.054 (0.04)		-0.008*** (0.00)		0.012 (0.10)
multimid		0.223 (0.15)		0.027*** (0.01)		0.202 (0.24)
sani_den		-0.098 (0.10)				0.069*** (0.01)
ener_pub						
agua_pub						
esg_pub		-0.037 (0.12)		0.032*** (0.01)		
lixo_col		0.052*** (0.00)		0.048*** (0.00)		
noite						
sigma_hat	0.599	0.899	0.525	0.104	0.233	0.633

* Significativa a 10%; ** Significativa a 5%; *** Significativa a 1%.

Fonte: Elaboração própria a partir das variáveis escolhidas para o 2º estágio.

6 – CONSIDERAÇÕES FINAIS.

Os resultados obtidos nesse trabalho mostram, de uma maneira geral, que as escolas selecionadas como eficientes provêm ensino de melhor qualidade a um custo relativamente menor do que as outras escolas. Alguns fatores, como a infra-estrutura, desempenham papel importante para propiciar tal resultado. Ademais, enquanto a presença de computador nas escolas contribui para incrementar o indicador de eficiência, equipamentos subtilizados como televisores e outros aparelhos correlatos (parabólica e vídeo-cassete), além de alguns outros recursos como bibliotecas, videotecas e laboratórios de ciências não ocasionam o mesmo resultado.

Outros indicadores importantes de eficiência são as variáveis referentes às condições familiares, citam-se, entre elas, a escolaridade da mãe, o estímulo à leitura, ter livros em casa, não exercer trabalho remunerado e cultivar o hábito do estudo contribuem para o melhor desempenho nas provas de matemática e português. Na verdade, tanto a família quanto a escola parecem ser responsáveis pelo bom resultado nas provas. A escola possui um papel importante no estímulo à leitura e no hábito de estudo, assim como o de fornecer o ambiente mínimo para que os alunos se desempenhem bem na tarefa de aprender. A família complementa a escola, já que as mães de maior escolaridade e as famílias com maiores recursos educacionais são, em geral, mais seletivas na escolha da instituição. No sistema público, a escolha não é totalmente livre, está atrelada à proximidade da residência familiar. Dessa forma, as escolas eficientes captam, também, um efeito da vizinhança, as melhores condições dos alunos que ingressam facilitam a trajetória do aproveitamento dos recursos.

A análise por mesoregiões destaca que as melhores áreas de eficiência são as regiões centrais de Minas: Metropolitana, Campo das Vertentes, Oeste de Minas, Central e Zona da Mata. Dentre as regiões pobres, Jequitinhonha é um exemplo positivo e, em algumas análises, o Norte de Minas, por obterem bons resultados mesmo que com poucos recursos.

Por fim, cabe ressaltar que esse trabalho pioneiro sobre o tema eficiência na provisão de serviço público em educação no Brasil é mais uma contribuição para formulação de políticas públicas. A experiência mineira, refletindo as disparidades socioeconômicas do país, evidencia que se faz mister conjugar maior destinação de assessoria técnica (em alguns casos também financeira) aos municípios localizados em regiões mais pobres que apresentam a vantagem de operar com rendimentos constantes ou crescentes e realocação de recursos naqueles onde a ineficiência não é fruto da escassez de insumos, mas sim de seu mau aproveitamento, por exemplo, nos municípios mais desenvolvidos. Para o primeiro conjunto, o acréscimo marginal nos insumos representará ganhos proporcionais (ou mais que proporcionais) nos resultados ao passo que, no segundo conjunto, a readequação pode torná-los mais eficientes.

7 – REFERÊNCIAS:

AFONSO, A.; AUBYN, M. ST. **Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs.** Frankfurt: European Central Bank, 2005. 39 p. (Working paper; 494).

BANKER, R. Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. **Management Science**, v. 39, n. 10, p. 1265-1273, Oct. 1993.

CAZALS, C.; FLORENS, J.; SIMAR, L. Nonparametric frontier estimation: a robust approach. **Journal of Econometrics**, v. 106, n. 1, p. 1-25, 2002.

CHARNES A; COOPER, W.W.; RHODES, E. Measuring the efficiency of decision making units. **European Journal of Operational Research**, v. 2, n. 6, p. 429-444, 1978.

CHARNES, A.; COOPER, W.W.; RHODES, E. Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. **Management Science**, v. 27, n. 6, p. 668-697, Jun. 1981.

- COELLI, T.; RAO, D. S. P.; BATTESE, G. E. **An introduction to efficiency and productivity analysis**, Boston: Kluwer Academic, 1998. 275 p.
- COLEMAN, J. S.; CAMPBELL, E. Q.; HOBSON, C. J.; MCPARTLAND, J.; MOOD, A. M.; WEINFELD, F. D.; YORK, R. L. **Equality of educational opportunity**. Washington, D.C.: U.S. Government Printing Office, 1966.
- DEBREU, G. The coefficient of resource utilization. **Econometrica**, v. 19, n. 3, p. 273-292, July 1951.
- FAÇANHA, L.; MARINHO, A. Instituições federais de ensino superior: modelos de financiamento e o incentivo à eficiência. **Revista Brasileira de Economia**, v. 53, n. 3, p. 357-386, jul./set. 1999.
- FARRELL, M. J. The measurement of productive efficiency. **Journal of the Royal Statistical Society, Series A**, v. 120, n. 3, p. 253-290, 1957.
- FARIA, A. F.; JANUZZI, P. M. Eficiência dos gastos municipais em saúde e educação: uma investigação através da análise envoltória no Estado do Rio de Janeiro. Rio de Janeiro: IPEA, 2006. (Prêmio IPEA-CAIXA 2006; Concurso de Monografias. Tema 1: Eficiência e Efetividade do Estado no Brasil).
- GIBELS, I.; MAMMEN, E.; PARK, B. U.; SIMAR, L. On estimation of monotone and concave frontier functions. **Journal of the American Statistical Association**, v. 94, n. 445, p. 220-228, Mar. 1999.
- KNEIP, A.; SIMAR, L.; WILSON, P. W. **Asymptotics for DEA estimates in nonparametric frontier models**. 2003. 37 p. (Technical report; 0323).
- OLIVEIRA, J.B. A.; SCHWARTZMAN, S. **A escola vista por dentro**. Belo Horizonte: Alfa Educativa, 2002.
- SAMPAIO DE SOUSA, M. C.; STOSIC, B. Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers. **Journal of Productivity Analysis**, v. 24, n. 2, p. 157-181, 2005.
- SAMPAIO DE SOUSA, M. C.; RAMOS, F.. S. Eficiência técnica e retornos de escala na produção de serviços públicos municipais: o caso do nordeste e do sudeste brasileiros. **Revista Brasileira de Economia**, Rio de Janeiro, v. 53, n. 4, p. 433-461, 1999.
- SHEPPARD, R. W. **Theory of cost and production function**. Princeton, NJ: Princeton University, 1970. 308p.
- SIMAR, L. Detecting outliers in frontier models: a simple approach. **Journal of Productivity Analysis**, v. 20, n. 3, p. 391-424, 2003.
- SIMAR, L. WILSON, P. W. Estimation and inference in two-stage, semi-parametric models of production process. **Journal of Econometrics**, vol. 136, January 2007: 31-64.
- SIMAR, L. WILSON, P. W. Non-parametric tests of returns to scale. **European Journal of Operational Research**, v. 139, n. 1, p. 115-132, 2002.
- SIMAR, L.; WILSON, P. W. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models **Management Science**, v. 44, n. 1, Jan. p. 46-61, 1998.
- SOARES, T.; PEREIRA, D. Estudo de critérios de adequação para modelos da teoria da resposta ao item (TRI) aplicado ao caso do ensino fundamental da micro-região de Juiz de Fora em 1999. **Educação em Foco**, v. 6, n.2, p. 91-108, 2002.
- WILSON, P. W. **Efficiency in education production among PISA countries with emphasis on transitioning economies**. Texas: University of Texas. 2005. 40 p. Disponível em: <http://upeg.eerc.kiev.ua/conf/papers/Wilson_Paul.pdf> Acesso em: 23 jan. 2007